A New Approach for Clinical Biological Assay Comparison and Standardization: Application of Principal Component Analysis to a Multicenter Study of Twenty-One Carcinoembryonic Antigen Immunoassay Kits

Jean-Claude Rymer,1 Robert Sabatier,2 Alain Daver,3 Jacques Bourleaud,4 Marcel Assicot,5 Jacqueline Bremond,6 Jacqueline Rapin,7 Sharon Lynn Salhi,8 Bruno Thirion,9 Anne Vassault,10 Jacques Ingrand,11 and Bernard Pau,2,12* for the joint group convened by the Société Française de Biologie Clinique (SFBC), the Commission de Radioanalyse et Techniques Associées (CORATA), the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC), and the Syndicat de L’Industrie du Diagnostic In Vitro (SIDV)

Background: Principal component analysis (PCA) is a powerful mathematical method able to analyze data sets containing a large number of variables. To our knowledge, this method is applied here for the first time in the field of medical laboratory analysis.

Methods: PCA was used to evaluate the results of a blind comparative study of 21 carcinoembryonic antigen (CEA) reagent kits used to determine CEA concentration in a panel of sera from 80 patients.

Results: The mathematical technique first eliminated the variations attributable to the use of different calibrators. The PCA representation then gave a global view of the dispersion of the kits and allowed the identification of a main homogeneous group and of some discrepant kits.

Conclusions: PCA applied to the in vitro diagnostic reagent field could contribute to the standardization process and improve the quality of medical laboratory analyses. A standardization method using a panel of patient sera is proposed.

The serum concentrations of tumor markers, measured by immunoassay, are considered by clinicians to be major indicators of cancer disease status. To be of clinical interest, these immunoassays must be reproducible and reliable. In addition, standardization of the entire analytical procedure should lead to homologous results when the same marker is assayed using kits from different commercial sources. Theoretically, the results obtained for the same patients tested in different places using different

1 Laboratoire de Biochimie, CHU Henri Mondor, Avenue De Lattre de Tassigny, 94000 Créteil, France.
2 CNRS UMR 9921, Faculté de Pharmacie, Avenue Charles Flahault, 34060 Montpellier Cedex 2, France.
3 Centre Paul Papin, Laboratoire de Radio-Immunologie, 2 Rue Moll, 49036 Angers Cedex, France.
4 Agence Française de Sécurité Sanitaire des Produits de Santé, Site de Montpellier-Vendargues, 13 Rue de la Garenne, 34740 Vendargues, France.
5 Département de Biologie Clinique, Institut Gustave Roussy, 94805 Villejuif, France.
6 Laboratoires Abbott, 12 Rue de la Couture Silic 203, 94518 Rungis Cedex, France.
7 Syndicat de l’Industrie du Diagnostic en Vitro, 6 Rue de la Trémoille, 75008 Paris, France.
8 Laboratoire d’Immunologie et Biotechnologie, Faculté de Pharmacie, Avenue Charles Flahault, 34060 Montpellier Cedex 2, France.
9 Cis Bio International, BP 21, 91192 Gif sur Yvette Cedex, France.
10 Biochimie A, Hôpital Necker, 149 Rue de Sèvres, 75743 Paris Cedex 15, France.
11 Laboratoire de Radioanalyse, Hôpital Cochin, 27 Rue du Faubourg St. Jacques, 75674 Paris Cedex 14, France.
12 Comité de Immunoanalyse IFCC, Faculté de Pharmacie, Avenue Charles Flahault, 34060 Montpellier Cedex 2, France.
*Author for correspondence. Fax 33 (0)4 67 54 86 10; e-mail sumr9921@pharma.univ-montp1.fr.
Received December 17, 1998; accepted March 23, 1999.
kits should be equivalent. Nevertheless, numerous comparative studies (1–4) indicate that this ideal goal is rarely reached, and major discrepancies appear between reagents used to assay the same marker, even when the assays are performed in parallel with approved methods (5).

The measurement of carcinoembryonic antigen (CEA) constitutes a major model for pointing out discrepancies and defects in standardization. This marker, described in 1965 by Gold and Freedman (11), is a glycoprotein whose functions have not been elucidated (12). Its increase is associated with the progression of gastrointestinal tumors and other cancers (13, 14). Used worldwide for the diagnosis and follow-up of cancer patients (15), CEA is one of the most frequently measured tumor markers in France. The importance of the risk of error in CEA measurement must not be ignored by clinicians (16), particularly for results close to the commonly used cutoff value of 5 μg/L.

Taking into account these considerations, a joint group composed of representatives of several scientific societies (the Société Française de Biologie Clinique, the Commission de Radioanalyse et Techniques Associées, and the Fédération Nationale des Centres de Recherches et de Lutte contre le Cancer) and a federation of manufacturers, the Syndicat de L'Industrie du Diagnostic in Vitro (SIDV), decided to evaluate the actual degree of heterogeneity of the results given by the different CEA immunoassay kits distributed in France. For the first time, nearly all of the commercial kits available on a national market for the assay of one given marker were tested simultaneously on a large panel of serum samples: 21 kits (distributed by 14 companies) were evaluated by the measurement of CEA in sera from 80 patients.

Statistical analysis of so many data points raises a problem. Classical regression analysis used for evaluating results obtained by one method vs those of a reference method (or at least a commonly used one), and which only compares kits two-by-two, was inadequate, even if improvements in this approach have been proposed (17, 18). The two-by-two kit comparison is particularly unsuitable in the case of CEA immunoassays for which there are no methods that measure the absolute quantity of CEA and no recognized standard methods to compare the results. A “mathematical visualization” technique was thus necessary to provide an easy-to-understand representation of our large data set. Consequently, we used principal component analysis (PCA), which has been shown to be a very powerful tool for displaying relationships between different factors (particularly in social and agronomic sciences), but which has never been applied, to our knowledge, in the field of medical analysis evaluation.

A brief explanation of the principle of PCA (19, 20), as applied to our study, follows. PCA is a method that reduces the number of variables to a small number of principal components. These components summarize the information in the original variables and are linear combinations of them. Suppose we had tested the 21 kits by assaying only two sera. On a graph with two axes (on a plane, i.e., a two-dimensional space), we could represent each kit by a point whose coordinates are the CEA concentration of the first serum on one axis and the CEA concentration of the second serum on the second axis. (Each kit can also be represented by an associated vector whose origin is the intercept of the axes and whose extremity is one of the above-stated points). Kits giving quite similar results would be represented by points clustered into the same group, and discrepant kits would give outlying points. If we had three sera, we could use a third axis, orthogonal to the first two axes, in a three-dimensional space. However, two or three sera are not a statistically representative sample and the addition of more sera implies a space with more than two or three dimensions. Because we have a panel of 80 sera, which definitely constitutes a representative sample, we need an 80-dimensional space, mathematically possible but impossible to display. In this multidimensional space, we can still calculate, for each kit, an associated vector whose coordinates are, on 80 axes, the 80 initial components, i.e., the 80 results of the assay of the 80 sera with each kit. We then need to reduce the number of dimensions to obtain a displayable representation on a plane.

To this aim, in this multidimensional space, a two-dimensional subspace can be determined. This plane is defined by two axes chosen to conserve the maximum of the initial data (these axes must be independent, i.e., orthogonal). In this way, the two axes are determined in the two orthogonal directions representing the maximum dispersion of the data. The projections of each kit-associated vector on this plane are then calculated. The projected vectors will conserve the maximum of the distances between the initial vectors.

On these new axes, the coordinates of the projected vectors are its principal components. Clearly, in this representation, the projected points (i.e., the extremities of the projected vectors) will be clustered for kits giving similar results and separated for discrepant kits. The dispersion of the results for the sera, after elimination of the influence of their absolute CEA concentrations (explained below), can be analyzed in a similar manner, thus making it possible to identify outlying sera as well.

In both cases, we could also calculate a third axis, orthogonal to the first two, and obtain a third principal component. We could then calculate the subsequent axes (until n – 1 axes, where n is the number of dimensions) and components that obviously cannot be displayed. In

---

13 Nonstandard abbreviations: CEA, carcinoembryonic antigen; SIDV, Syndicat de L’Industrie du Diagnostic in Vitro; PCA, principal component analysis; and AMMI, additive main effects and multiplicative interaction.

14 In our case, the rough results are first transformed into the residuals of the analysis of variance (see below).
fact, two or three axes bearing the first two or three components would summarize the data and should be easily interpretable. Additional components are mostly "noise". The ability to reduce the data to a very small number of vectors is known (21).

It is not guaranteed that the principal components will correlate with some experimental factor such as incubation time or temperature, antibody affinity, antibody heterogeneity, patient status, or degree of glycosylation, but such is possible and illustrates the exceptional analytical power of PCA.

Materials and Methods

SERA
The panel of 80 sera consisted of 77 samples obtained from patients suffering from breast (n = 51), colorectal (n = 16), pancreas (n = 3), liver (n = 2), stomach (n = 2), lung (n = 1), thyroid (n = 1), or ovary (n = 1) cancer. Three samples were obtained from non-cancer patients. To verify whether the between-assay variability could be decreased by the use of a common calibrator, each laboratory assayed a purified and concentrated 5000 µg/L CEA solution (referred to here as the purified common calibrator), kindly provided by Immunotech (Marseilles, France), which initially had been prepared for the Bureau Communautaire de Référence.

BLIND ASSAY PROCEDURE
The manufacturers involved in the study assayed the panel of sera using their own reagent kits. All tubes of serum were first coded to ensure that the sera could not be identified by the manufacturers. The results obtained with each kit were then coded by the SIDV to ensure that, after gathering the data, no one (except the SIDV) could identify the kits, the condition imposed by the manufacturers to participate in this study.

To test the 21 kits, 21 identical panels of the 80 serum samples were formed. Each serum, identified by its serum code, was divided into 21 aliquot parts in one of our laboratories, the sampling site. The encoding software was created and kept at a second site, the coding site. This software was used to print out 80 boards of 21 labels. Each board bore one of the serum codes. Each label had a letter and a unique randomized number. The 80 boards were then sent to the sampling site.

At the sampling site, the labels from each board were placed at random on the tubes of the corresponding serum sample. The tubes of this sample were then distributed equally among the 21 panels, so that all tubes bearing the same letter were assembled in the same panel. Each tube was thus identified by its unique randomized number code and one letter code common to the 80 tubes of each panel. It was therefore impossible, before the final decoding, to establish a connection between the serum samples in the different panels.

The 21 panels of 80 tubes were frozen and sent to the SIDV, where letter codes were assigned to the kits (the SIDV kept these letter codes secret and destroyed them at the end of the study). The panels were then dispatched to the manufacturers of the designated kits. Each manufacturer assayed the serum samples once or in duplicate (for 14 kits). The purified common calibrator was assayed once for 5 kits, in duplicate for 12 kits, and in triplicate for 2 kits; no results were given for 2 kits. Each manufacturer was free to choose the dilutions of the purified common calibrator so that they fell within the concentration range for the kit in question.

The results were returned, via the SIDV, to the coding site, where the number codes of the sera were decoded and the results tabulated for the corresponding serum-kit couples. The kits remained identified by their letter codes only.

ANALYSIS OF THE DATA
The data (i.e., single results or the mean of duplicates) were collected in an 80 × 21 array with the individuals (sera, n = 80) in rows and the variables (kits, p = 21) in columns. Ten missing values were replaced by the mean result for the serum in question.

As a first approach, we examined the raw data by calculating the general mean value of each kit and the mean value, SD, and CV of each serum. Results given by the purified common calibrator were also analyzed and regression lines (measured vs theoretical values) calculated.

The PCA study was then performed using the ADE 3.6 software (a gift from D. Chessel and S. Dolédec, Program Library for the Analysis of Environmental Data, URA CNRS 1451, Université Lyon 1, Villeurbanne, France).

To reduce and homogenize the variances, the raw data were first log transformed. The array was then bicentered. Bicentering, which leads to a mean value equal to zero in the columns and in the rows, is very useful when the data table is homogeneous (only one measurable dimension), which was the case for our data. This calculation consisted of subtracting from each logarithmic value the general mean value of the array (µ), the isolated effect of the serum (α_i in the rows), and the isolated effect of the kit (β_j in the columns). Each data point was thus reduced to the residual (ε_ij) of the analysis of variance with two factors without interaction. This residual represents the composed effect of each serum-kit couple.

This mathematical treatment, i.e., logarithmic transformation and bicentering, is a crucial point and has a very important consequence: it eliminates (a) the effect of the calibration on the results, (i.e., the differences because of the use of different calibrators in different kits); and (b) the absolute CEA concentration of each serum. In other words, if we set X_ij as the logarithm of the measured CEA concentration in serum “i” with kit “j”, then: ε_ij = X_ij - (µ + α_i + β_j).

15 If we set X_ij as the logarithm of the measured CEA concentration in serum “i” with kit “j”, then: ε_ij = X_ij - (µ + α_i + β_j).
words, it extracts the only remaining heterogeneity attributable to the differences in the reactivity between the 80 × 21 couples of CEA samples and kits.

To analyze the term $e_{ij}$, we chose to use the model with “additive main effects and multiplicative interaction (AMMI)”, also called “factor analysis of variance” (24–26) proposed by Mandel (24) and Gollob (25):

$$e_{ij} = \lambda \gamma_i \delta_j + \epsilon_{ij}^{***}.$$

In this model, the term $\lambda \gamma_i \delta_j$ represents the multiplicative interaction between serum “$i$” and kit “$j$” and is given by the PCA of the set of $e_{ij}$.

The variability (called “inertia”) of the entire set of 21 points included in the 80-dimensional space was evaluated, and the axes (bearing the principal components) were determined using matrix calculations. The first axis

The calculations are as follows. In the preceding term of the multiplicative interaction, $e_{ij} = \lambda \gamma_i \delta_j + \epsilon_{ij}^{***}$. $\lambda$ is a coefficient and $\gamma_i$ and $\delta_j$ are the coordinates of one component. In an $r$-dimensional space, with $|k| = 1$ to $r$ and $1 \leq r \leq \min (p-1, n-1)$: $e_{ij} = \lambda^{(1)} \gamma_i^{(1)} \delta_j^{(1)} + \lambda^{(2)} \gamma_i^{(2)} \delta_j^{(2)} + \ldots + \lambda^{(r)} \gamma_i^{(r)} \delta_j^{(r)} + \epsilon_{ij}^{***}$. Finally, $e_{ij}^{*} = e_{ij} - \lambda^{(k)} \gamma_i^{(k)} \delta_j^{(k)} + \epsilon_{ij}^{***}$. The terms $\lambda^{(k)}$, $\gamma_i^{(k)}$, and $\delta_j^{(k)}$ are found by matrix calculations on the $e_{ij}$ array.

The total inertia of a set of points is given by: $I = \sum m_i d_i^2$, where $m_i$ is the “weight” of each point and $d_i$ its distance from the origin $O$. In our case, the homogeneity of data led us to choose the weights equal to one, and then the total inertia is equal to the sum of the squares of the previous residuals $e_{ij}$: $I = \sum_i e_{ij}^2$. 

---

Fig. 1. CEA concentrations in all 80 serum samples, as measured with 6 of the 21 immunoassay kits tested.

The vertical columns represent the CEA concentration (rough data, expressed in µg/L) of each serum. The horizontal bars indicate the mean CEA concentration of each serum for the 21 kits (in µg/L). The sera are arranged in ascending order as a function of their mean CEA concentration determined by using the 21 kits.
was determined in the direction that allowed the representation of the maximum rate of the total inertia, i.e., the one for which the inertia of the values projected on this axis were the highest possible; the second axis was determined in a direction orthogonal to the first axis, giving access to the maximum of the remaining inertia. Additional axes were determined in the same manner. For each axis, the percentage of projected inertia was calculated vs the total inertia of the set and indicated the importance of the axis in the total variability of the set.

In the AMMI model (24–26), it is usually possible to statistically verify the requisite number of components (i.e., the number of dimensions). In our study, this was not possible for two essential reasons: (a) the interaction by the use of the residual variance ($e_{ij}$) could not be tested, and (b) other procedures (23) could not be used because of the excessively large number of kits and sera (exceeding the dimensions of the proposed tables).

The PCA representation of the log-transformed and bicentered data underscored the similarities and the differences between kits (and between sera, which were treated in exactly the same manner).

**Results**

**EVIDENCE FOR OVERALL DISPERSION OF THE RESULTS**

Preliminary examination of the results confirmed their suspected heterogeneity. The general mean values for the kits varied from 21.8 µg/L (kit X) to 34.1 µg/L (kit S). The discrepancies between six kits taken as examples are illustrated in Fig. 1. The results for kit B are close to the mean values, but the results for kits I, X, and D increasingly differ from those for kit B. The results for kit D are very far from the mean results. On the other hand, the results for kits K and R (general mean value, 33.7 and 34.1 µg/L, respectively) are very close to each other and seem, perhaps, to differ from those for kit B (general mean value, 26.6 µg/L) only by a proportionality factor.

Inspection of the results for individual sera clearly confirmed the discrepancies. The results for four sera are represented in Fig. 2. The interkit reproducibility, expressed as the CV, ranged from 16% for serum 26 to 49% for serum 10 (Fig. 2A). For serum 10, with a mean value of ~19 µg/L, the results with kits F and X, under or near the cutoff value of 5 µg/L, would have led to an error in diagnosis. The results of the assays of serum 20 and serum...
58, whose mean results were close to the cutoff value, underscore how serious this problem is (Fig. 2B). For serum 20, even with a rather low CV (19%), the results varied from 3.5 mg/L to twice that concentration, and indicated for the patient either a favorable or an unfavorable prognosis, depending on the kit used. In the case of serum 58, the discrepancies were more obvious: the mean value of 6.1 mg/L, which was above the cutoff value, apparently indicated a poor prognosis for the patient; nine kits, however, gave reassuring results because they were below the cutoff value. In addition, the same kits did not yield the largest discrepancies for all samples: for example, the largest discrepancies for serum 58 were in kits F and D, and the largest discrepancies for serum 10 were in kits L and X. It must be noted that, even if this kind of representation highlights some degree of heterogeneity, it cannot be used to give an entire and global view of the assays of the 80 sera by the 21 kits.

In most cases, correlation between theoretical values and measured values of the purified common calibrator showed good linearity of the regression lines, but their slopes differed, ranging from 0.48 to 1.04, depending on the kit. An attempt to equalize the results of the assays by dividing the values by the slopes failed to reduce the heterogeneity.

DATA TRANSFORMATION AND REDUCTION
The differences between the 21 kits are also seen in the histograms and gaussian curves of the log-transformed results in Fig. 3, which shows particularly marked differ-
Fig. 4. An array of the log-transformed and bicentered data. The raw data were transformed into the residuals of the analysis of variance ($e_i$). The circles correspond to positive values, and the squares correspond to negative values. The areas of the circles and squares are proportional to the residuals.
ences between kit D and the other kits. These differences are even more pronounced in Fig. 4, which is a view of the data after they were bicentered, the areas of the circles (positive values) and the squares (negative values) being proportional to the residuals $e_{ij}$. Kit D again appears to be markedly different from the others; kit E can also be singularized.

PCA REPRESENTATION OF THE DATA

PCA allows a finer analysis of the heterogeneity. The closer the points were, the more similar the corresponding kits were. In contrast, a point separated from the others represented a kit giving discrepant results.

In the first representation obtained for the kits (Fig. 5), the first two axes determining the plane account for 61.1% of the total inertia. A third axis would add 15% more inertia, but this representation is more difficult to display. The two-dimensional scheme confirms that kit D differs strongly from the others. Its point is so distant from the others that all the data processing (including the PCA of the sera) was performed again without the data from this kit, which is responsible by itself for a large part of the variability of this first representation.

In the new representation (Fig. 6) without kit D, the first two axes account for 61.8% of the new total inertia, and the array is rearranged. Four groups of reagents can now be clearly identified: a main group of 13 (A, B, C, H, J, K, L, N, Q, R, S, V, and, somewhat separated, Y), a group of 4 (G, T, M, and I), a group of 2 (F and X), and finally 1 isolated kit (E).

We pointed out above (Fig. 1) that the results for kit B differed from those for kits K and R, perhaps only by a proportionality factor. Interestingly, it can be noted that, on the PCA plot (Fig. 6), kits K, R, and B fall into the same group (the main group). This strongly suggests that the differences between the results for kit B vs those for kits K and R are essentially caused by a difference in calibration and that this difference has been eliminated by the log transformation and bicentering of the data.

The PCA biplot for the sera is presented in Fig. 7. The differences attributable to their absolute CEA concentration have also been eliminated by the log transformation and the bicentering of the data. The majority of the sera belong to the same cluster of points, suggesting that these sera react similarly; however, some of them, sera 19, 10, 67, 11, 21, and 71, are distinct from this main cloud and can be identified as outliers. These six sera appear to react differently with the set of kits than the main group of sera. Interestingly, the results for serum 10 already appeared as highly variable in the preliminary examination of the results (Fig. 2).

The analyses of variance and the PCA representations of the 20 kits (without kit D) and the 80 sera are summarized in Fig. 8. Each result (each serum-kit couple, represented by a circle or a square, as in Fig. 4) is now positioned by the projection of the kit on the first axis of Fig. 6 and of the serum on the first axis of Fig. 7. A major group, composed of the most homogeneous kits and sera, is easily distinguishable. Kit Y is slightly isolated from this group. The second group of kits (G, T, M, and I), the isolated kits (E, X, and F), and the most distant sera (19, 10, 67, and 71) appear clearly separated.

Discussion

The preliminary examination of the data showed their high degree of variability: differences between the mean value of the results for each kit, differences between the profiles of the kits with respect to all the sera, very large differences in the CEA concentrations of the samples when measured by the different kits, and differences in
the calibration curves when a common calibrator was used. Nevertheless, using this rather classical approach, we were not able to determine exactly which kits and sera were responsible for the dispersion.

After logarithmic transformation and bicentering, differences between kits (and also between sera) were more striking. The first PCA biplot then showed that the kits could be distributed into a main group and one isolated kit. This analysis confirmed that this kit, kit D, was very different from the others, and its data had to be excluded from further calculations of the PCA representations. Among the remaining 20 kits, 13 kits were found to form a main group, meaning that their results were practically equivalent or proportional. Another kit, kit E, now appeared to be very different from the others.

PCA can be a very useful tool to identify outliers caused by human mistakes (data transcription errors) or instruments out of control; however, although the hypothesis of human blunder cannot be completely eliminated, precautions were taken so that we can be reasonably confident in the conclusions of this work.

The PCA of the sera revealed that 74 samples were similar and 6 were different. For these outlying sera, different results were obtained with the majority of the kits used. No relationship between the results and the nature of the pathology of the patients could be identified, neither with respect to localization of the tumor nor with respect to disease progression, metastatic status, or therapy. These clinical variables are independent data and cannot explain the discrepancies.

When the results of the study were shown to all the manufacturers, each received a copy of the entire set of transcribed rough data for the assay manufactured by that company. Each kit was still identified only by its letter code and each serum only by its number codes. Each manufacturer knew the letter code for its kit and was asked to verify the results. None of the manufacturers, including the manufacturers of kits D and E, ever contested the results of the study.
It can be concluded that 8 kits and 6 samples were the main causes of most of the heterogeneity we observed in our assay of 80 sera, using 21 kits. In addition, the conclusions of the mathematical analysis of the kits did not vary when the outlying samples were removed from the data set.

With regard to the calibrators, Börmer (6) showed that neither cross-testing of the CEA calibrators of various kits nor the use of the International Reference Preparation OMS 73/601 reduced the discrepancies. In our study, we were not able to equalize the results of the set of kits by the use of our purified common calibrator. In addition, among the 10 kits calibrated against the OMS 73/601 preparation, only 6 kits belonged to the main group of 13 determined by the PCA analysis (data obtained under the control of the SIDV), thus confirming that a common calibrator alone is not sufficient to improve the results.

Moreover, with regard to this main group of 13 kits, we could have expected that the proportional differences in their results would have been reduced by the use of the purified common calibrator. Surprisingly, even in this group, correcting the results according to the slopes of this calibrator failed to equalize the assays. This could be explained by the choice of our somewhat artificial calibrator, and a more natural calibrator obtained by mixing sera from patients might be more efficient. Along these lines, one solution was to multiply the individual results by the ratio of the mean results for the whole group of 13 kits to the mean result of each of the kits used; this allowed a rather good homogenization of the results (data not shown). In this way, the whole serum panel would in fact become the calibrator and perhaps the best one we can use.

No particular link with any of the other known characteristic of the kits (information obtained under the control of the SIDV) was found to explain the differences between the groups. In the main group, five kits use at least one polyclonal antibody and five use a monoclonal system. Antibodies directed against three epitopes on the CEA molecule—gold I, IV, and V (27, 28)—are used most often, but seem to be distributed randomly among the kits. The fact remains that CEA immunoassays need to be studied in more detail. A clear identification of the CEA epitopes recognized (28, 29) is a key point, given the existence of mutants of the CEA molecule (30) and variations in glycosylation (31). Other characteristics of the assay must be taken into consideration: for example, the number of monoclonal antibodies used to capture the antigen, the kinetic constants of interaction, the buffer, the coating procedure, and the nature of the solid phase and of the detection system. An attempt to select antibodies...
Fig. 8. Summary of the entire study: analysis of variance and PCA representations for 20 kits (kit D was excluded) and the sera. As in Fig. 4, each point (circle or square) represents a result: circles are positive values and squares are negative values. The areas of circles and squares are proportional to the residuals $e_i$ given by the analysis of variance. Localization of each point is given by the distance of the corresponding kit on the first axis of Fig. 6 and by the distance of the serum on the first axis in Fig. 7 (these axes account for the main inertia).
directed against particular antigenic sites has been successful for the immunoassay of thyroglobulin (32) and luteinizing hormone (33, 34).

In conclusion, although we were unable to explain the differences between the results obtained using different kits and to attribute these differences to a particular characteristic of the kits, PCA representation allowed us to compare simultaneously, in a large serum panel, almost all of the reagents available on a national market proposed for the measurement of an analyte. To our knowledge, this is the first attempt to apply PCA in the field of medical laboratory analysis. The coding procedure and the collaboration between scientists and manufacturers avoided any biases and disagreements.

PCA, which has generally been used to identify and represent the differences and similarities between populations, can also be used to compare reagent kits (and serum samples). It is a powerful method, able to analyze the results of a large comparative assay, to provide a comprehensive and clear presentation of a large data array, and to graphically display homologies and differences between reagent kits. Nevertheless, it does not allow the determination of whether the results given by a kit are the true values, but only that these results differ from those given by another kit. A common calibrator should no longer be seen as the primary solution to reduce the dispersion of the results, nor should a search for common calibrators be the main focus for manufacturers and researchers. What is essential is to identify those kits giving similar results; only in this situation might a common calibrator significantly reduce the dispersion. In fact, we found that, instead of our artificial common calibrator, the whole panel of patient sera itself was the best calibrator. For any study on standardization of reagents kits, we conclude that a panel of sera from all of the reagents available on a national market proposed for the measurement of an analyte. To our knowledge, this is the first attempt to apply PCA in the field of medical laboratory analysis. The coding procedure and the collaboration between scientists and manufacturers avoided any biases and disagreements.

We thank Francis Paolucci and Gerard Derzko for their help in the design of the study and Andrew Kramar for critical comments. The following companies kindly collaborated to this project: Abbott, Amersham Kodak, BioMérieux, Bio-Rad, Boehringer-Mannheim, Chiron, Cis Bio International, Hoechst-Behring, Merck-Clèvenot, EG&G, Realef, Produits Roche, Serono Diagnostics, and Sorin Biomedica.

References


17. Bland JM, Altman DG. Statistical methods for assessing agree-


