A comparative method for processing immunological parameters: developing an "Immunogram"

Riccardo Ortolani[1], Paolo Bellavite[2], Fiorenza Paiola[1], Morena Martini[1], Martina Marchesini, Dino Veneri, Massimo Franchini[4], Salvatore Chirumbolo, Giuseppe Tridente[1], Antonio Vella[1]

[1] Sezione di Immunologia, Dipartimento di Patologia Generale, Azienda Ospedaliero-Universitaria di Verona; [2] Dipartimento di Scienze Morfologiche e Biomediche, Università di Verona; [3]Sezione di Ematologia, Dipartimento di Medicina Clinica e Sperimentale, Azienda Ospedaliero-Universitaria di Verona; [4] Servizio di Immunoematotologia e Medicina Trasfusionale, Dipartimento di Patologia e Medicina di Laboratorio, Azienda Ospedaliero-Universitaria di Parma, Italy

Background. The immune system is a network of numerous cells that communicate both directly and indirectly with each other. The system is very sensitive to antigenic stimuli, which are memorised, and is closely connected with the endocrine and nervous systems. Therefore, in order to study the immune system correctly, it must be considered in all its complexity by analysing its components with multiparametric tools that take its dynamic characteristic into account.

Methods. We analysed lymphocyte subpopulations by using monoclonal antibodies with six different fluorochromes; the monoclonal panel employed included CD45, CD3, CD4, CD8, CD16, CD56, CD57, CD19, CD23, CD27, CD5, and HLA-DR. This panel has enabled us to measure many lymphocyte subsets in different states and with different functions: helper, suppressor, activated, effector, naïve, memory, and regulatory. A database was created to collect the values of immunological parameters of approximately 8,000 subjects who have undergone testing since 2000. When the distributions of the values for these parameters were compared with the medians of reference values published in the literature, we found that most of the values from the subjects included in the database were close to the medians in the literature.

To process the data we used a comparative method that calculates the percentile rank of the values of a subject by comparing them with the values for others subjects of the same age.

Results. From this data processing we obtained a set of percentile ranks that represent the positions of the various parameters with regard to the data for other age-matched subjects included in the database. These positions, relative to both the absolute values and percentages, are plotted in a graph. We have called the final plot, which can be likened to that subject's immunological fingerprint, an "Immunogram". In order to perform the necessary calculations automatically, we developed dedicated software (Immunogramma) which provides at least two different "pictures" for each subject: the first is based on a comparison of the individual's data with those from all age-related subjects, while the second provides a comparison with only age and disease-related subjects. In addition, we can superimpose two fingerprints from the same subject, calculated at different times, in order to produce a dynamic picture, for instance before and after treatment. Finally, with the aim of interpreting the clinical and diagnostic meaning of a set of positions for the values of the measured parameters, we can also search the database to determine whether it contains other subjects who have a similar pattern for some selected immune parameters.

Conclusions. This method helps to study and follow-up immune parameters over time. The software enables automation of the process and data sharing with other departments and laboratories, so the database can grow rapidly, thus expanding its informational capacity.

Keywords: immune monitoring, immunodiagnostics, systems biology, lymphocyte subsets, multiparametric analysis.

© SIMTI Servizi Srl

Introduction

The immune system is a complex network of numerous cellular components that communicate with each other both directly and through soluble mediators. It is a dynamic, 'intelligent' system, in that it responds to stimuli, adapts according to them and memorises them; furthermore, it is integrated with the neuroendocrine system. Given that it has features of "recognition" and "memory", it is extremely sensitive to various types of external stimuli, which it remembers for a variable period, sometimes for the whole life. These characteristics give the system marked dynamics, especially in the first decades of life when the antigenic experiences (stimuli) are numerous and, in particular, are "new". The complexity of the immune system makes continuous updating of analytic and diagnostic methods necessary.

The approach to immunological diagnostics usually consists of measuring a series of individual parameters, related mostly to cell types and their state of activity in the peripheral blood, which can be quantified predominantly by flow cytometric methods. In this context, the study of the lymphocyte populations, and even more so, the subpopulations, since these are more closely related to different functions, is a crucial instrument for measuring the dynamics of the immune system and its disorders. Furthermore, the possibility of simultaneously studying different lineage markers together with functional markers enables a large amount of important cell function information to be collected, although this body of data is difficult to evaluate.

The development of sophisticated technology, able to supply a large amount of data simultaneously, does allow greater understanding of the biology of complex systems. However, the amount and the analysis of single parameters are not, alone, sufficient to reveal the "dynamic" state of systems which does, in fact, require highly complex multiparametric processing and operative simulations; this is the field of so-called "systems biology".

In order to make a simple contribution that can be used in diagnostics, we developed a method of processing multiple data from individual patients and comparing these with analogous data from cohorts of patients in a database created ad hoc. Using this approach it is possible to evaluate a series of parameters, determining their nearness to median values (balance) or distance from the medians (imbalance) and giving a comparative representation in a single graph.

In order to evaluate the data collected, we developed a "Method for clinical analyses of the comparative type" (International Application No.:PCT/EP2006/003384; Priority Data: VR2005A000045 14.04.2005 IT), which facilitates the reading and clinical interpretation of complex measurements of this type. Other applications in immunological diagnostics will be presented in the future.

Materials and methods

In our Service we have adopted a multiparametric diagnostic system using lineage markers (CD45, CD3, CD4, CD8, CD19, CD16 and CD56) and functional markers able to detect different functional states of various subpopulations of lymphocytes. In particular, markers of the state of activation (HLA-DR, CD38, CD23), of naive/memory cells (CD5, CD45RA, CD28, CD27), and of effector function (CD57) were used to collection information on the behaviour of the immune system in apparently normal conditions or in the presence of disorders related directly or indirectly to immune function. In order to carry out these analyses, about 2 mL of whole blood were collected from each person into test-tubes containing EDTA; the automated BD FACS Sample Prep Assistant II (Becton Dickinson) distributes 50 mL of blood into each test-tube, distributes the mixtures formed of various monoclonal antibodies conjugated with six different fluorochromes (FITC, PE, PerCP, PE-Cy7, APC, APC-Cy7) (BD Biosciences), lyses the red blood cells and fixes the cells (BD FACS Lysing Solution). The white blood cells, in particular the lymphocytes, are then analysed automatically by flow cytometry (BD FACSCanto, Becton Dickinson) with BD FACSDiva software.

The numerous data collected during the flow cytometric analysis show considerable quantitative and qualitative variability for the cellular components, particularly those of the lymphocyte subpopulations. This variability, which represents a measure of the dynamicity of the system, is particularly dependent on the "antigenic experiences" that accumulate over time; for this reason the age of the subject is a factor

of paramount importance. The consequence of this is that it is essential to know the age and divide the subjects being studied into age groups in order for any comparative interpretation to be valid. The most frequent division is that of distinguishing subjects of paediatric age (in various groups) from adults. As far as concerns adults, the reference values used are commonly derived from blood donors, who are considered "healthy" subjects, or non-donors believed to be in good health[1-5]. As far as concerns paediatric subjects, the data in the literature are used[6-11]; these data are from children considered "healthy", although the age groups differ (in some cases substantially) and the numbers of subjects analysed per age group vary widely. Consequently, the reported paediatric reference values are not always comparable for diagnostic and scientific purposes and, in some cases, some to be discordant.

As far as concerns the paediatric age group, the variations within the wide range of data are substantial and it is, therefore, important to used very narrow age intervals for comparative analyses. We tackled this problem by using a weighted average of the medians reported in the literature, in this way obtaining a single median reference value per age and for each single marker that can be analysed (*unpublished data*). A further handicap to comparative multiparametric analyses is the limited availability of reference values for most of the lymphocyte subpopulations. Indeed, for many markers there are no reference values at all, particularly for children.

In conclusion, in order to carry out an adequate comparison for the multiparametric analysis and to obtain reliable reference values, the evaluations must be made for sufficiently restricted age groups, given the notable variability of the immunological markers assessed, even in conditions of apparent normality.

The database

The new approach to interpreting the multiparametric immunological data which we adopted in our laboratory is to exploit the set of data from patients attending our Service; these data have been collected over the years and entered into a specific database. There are some other descriptions of similar approaches, although partial and with different features, in the literature[12,13]. In detail, from 2000 all results of diagnostic measurements carried out to study lymphocyte populations and subpopulations in different pathological conditions have been collected and recorded. These data are integrated, when possible, with other immunological information (immunoglobulins, autoantibodies, etc.) as well as non-immunological, clinical, pharmacological, therapeutic and physiological (smoking, alcohol, diet, pregnancy, physical activity, occupation) information.

From among the comparative analyses made possible by this mass of data, we gave priority to comparisons between the value of a single immunological parameter and the set of values of the same parameter for the whole group or discrete subgroups of subjects with some of the same immunological characteristics or, conversely, with the same disorder, irrespectively of the immunological variations found. Since, as mentioned above, both the absolute values and the percentage values of each lymphocyte population vary considerably with age, comparison with data from homogeneous age groups is crucial for the detection of abnormalities. In this case, the comparative evaluation of a subject's data with the overall relevant data in the database requires that the data used for the comparison are drawn from subjects closely matched for age (for example, with the subjects ranked according to age, the 100 subjects in the database preceding the subject under evaluation and the subsequent 100). The comparison (calculation of the percentile rank) yields a value (from 0 to 100) which indicates the position of that particular parameter in the given subject with respect to the group of values of the same parameter in subjects in the groups used for the comparison. The large number and broad spectrum of diseases represented in the database enables the behaviour of a series of immunological parameters to be appreciated in various conditions.

Data from HIV-positive patients, including those with manifest disease (AIDS), are processed separately in a specific HIV-positive database both because this infection is known to cause profound immunological changes and because of the high number of HIV-positive individuals who attend our Service.

Software

Initially the database was managed using ordinary software for the management of archives.

Subsequently, both because of the increase in the data and, above all, because of the need to automate their processing, dedicated software was developed that fulfilled the specific requirements of the multiparametric comparative analysis. Currently the "Immunogramma" software (REAL-T srl, Verona, Italy) manages the database, enables the transfer of data directly from the instrument that generates them (the flow cytometer) or from the electronic spreadsheets, processes them, represents them in a graph and, finally, enables their export for further processing. "Immunogramma" is structured as a web application to allow wide access and data sharing, while respecting current legislation on privacy.

Results

At present, the database holds about 8,000 records. Each of these contains from a minimum of 30 to a maximum of about 200 parameters. These records include those from a subgroup of about 1,000 patients for whom there are also serological data on the common herpes virus infections (EBV, CMV, HSV, VZV) which have the greatest effect on immunological parameters. The registration of data in the database is, of course, continuing, with various functions available regarding the type of information that can be collected for each patient. For about one quarter of the subjects, the database contains more that one immunological record and, in some cases, data from up to 15 successive follow-up controls are available.

With the aim of verifying the overall behaviour of the immunological data in the database compared with the reference values reported in the literature and to compare these, albeit from different and heterogeneous sources, the mean of the medians in the literature (reported above) was calculated and this was overlaid on the analogous curve for the overall data in the database (with the exclusion of data from the HIV-positive group). This analysis showed that, despite the considerable and expected variability in the distribution, the medians in the literature and those of the database had similar behaviours, and that their partial discordance lessened progressively as the overall amount of data registered in the database increased and was compared (Figure 1). Some records

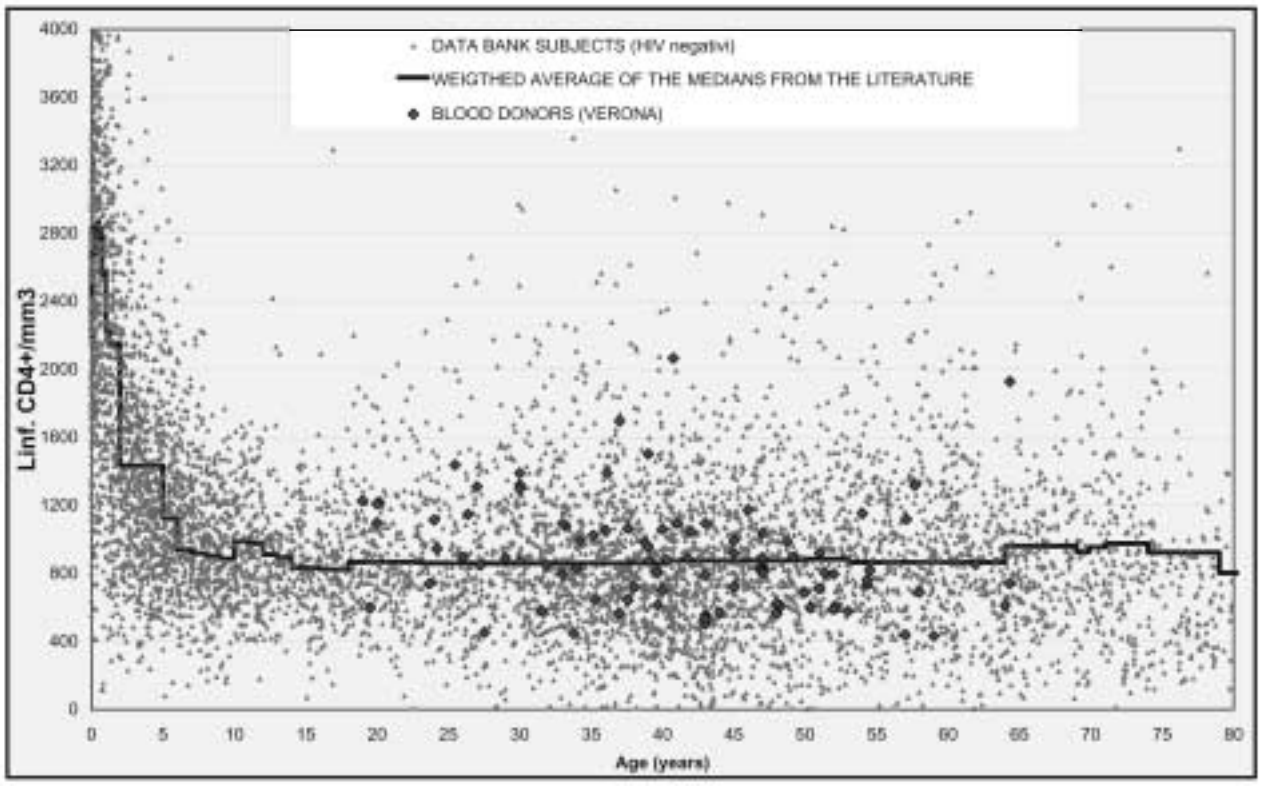

Figure 1 - Distribution of absolute values of $CD4^+$ T lymphocytes for subjects included in the database, blood donors and weighted average of medians from the literature for all ages

of very low levels of $CD4^+$ lymphocytes were due, in part, to data from various subjects with marked leucopaenia or lymphopaenia, in part to subjects with congenital immunodeficiency syndromes, and, possibly, in part due to subjects who could not be excluded from having HIV infection whose data, in the absence of serological or virological information, were included in the database of HIV-negative subjects.

Figure 2 shows a typical 'immunogram'. Each value of the percentile rank is located on a multiparametric graph which is defined as the "immunogram" and which immediately illustrates which parameters are distant from the median and are, therefore, different in that subject from those in other subjects in the cohort. The determination of 20^{th} and 80^{th} percentiles facilitates the identification of the most substantial differences.

In order to compare our findings with the "normal" reference values reported in the literature, the same figure includes the percentile ranks of the weighted means from the literature, calculated for the age of that particular subject under investigation. In this way it is possible to appreciate the extent of the difference of values for single parameters, with respect to both subjects of the same age in the database and, when they exist, with populations of so-called healthy subjects.

The possibility of comparing immunological and non-immunological parameters recorded during successive follow-ups enables the evolution of the parameters to be evaluated in the context of the clinical history or intercurrent events in a given interval of time (Figures 3 and 4).

Finally, the possibility of searching for subjects with substantial similarities in immunological profile, even among those with apparently different disorders, may direct further diagnostic investigations towards multiple or complex diseases and lead to the identification of "subgroups" of the same disease.

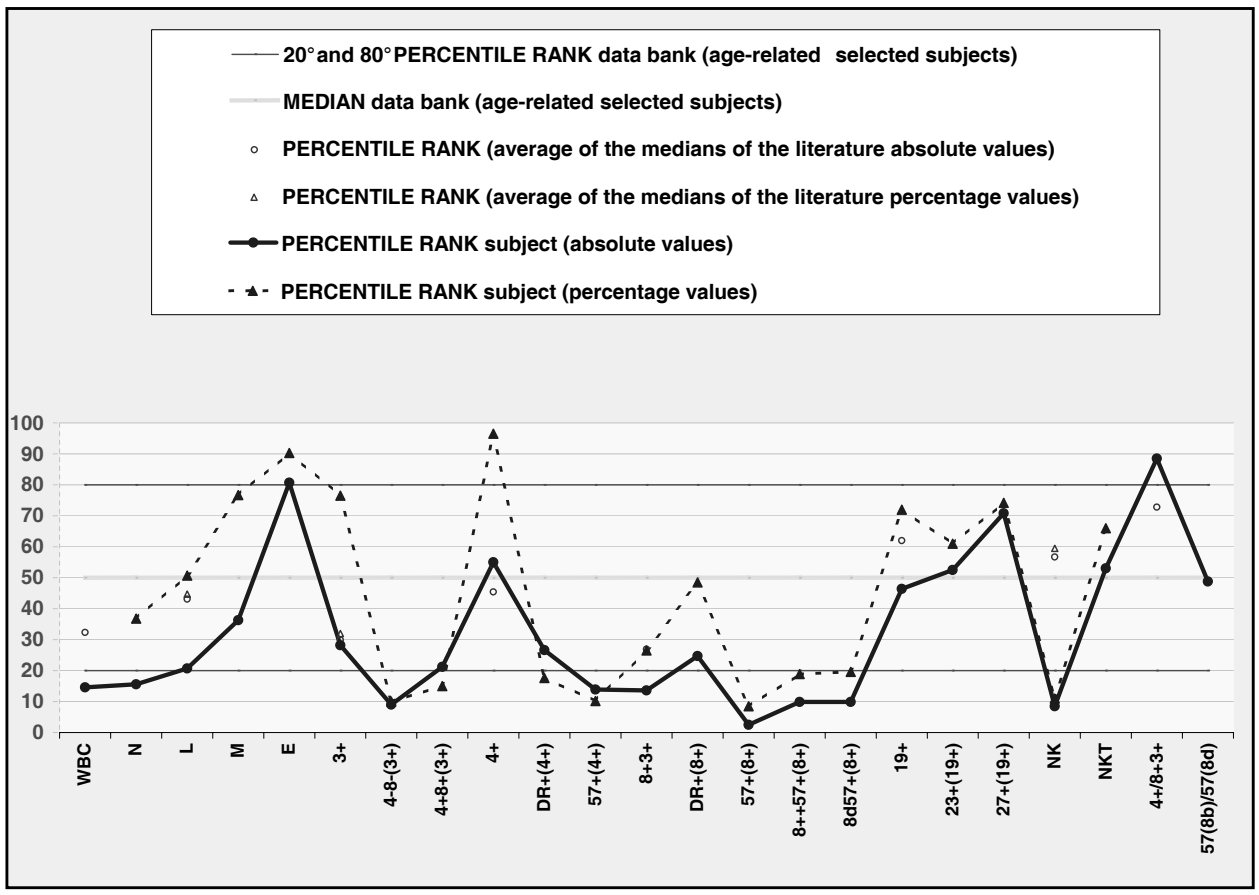

Figure 2 - Example of an immunogram, including percentile ranks of weighted medians from the literature, by age. The graph is for a female blood donor, aged 26 years old

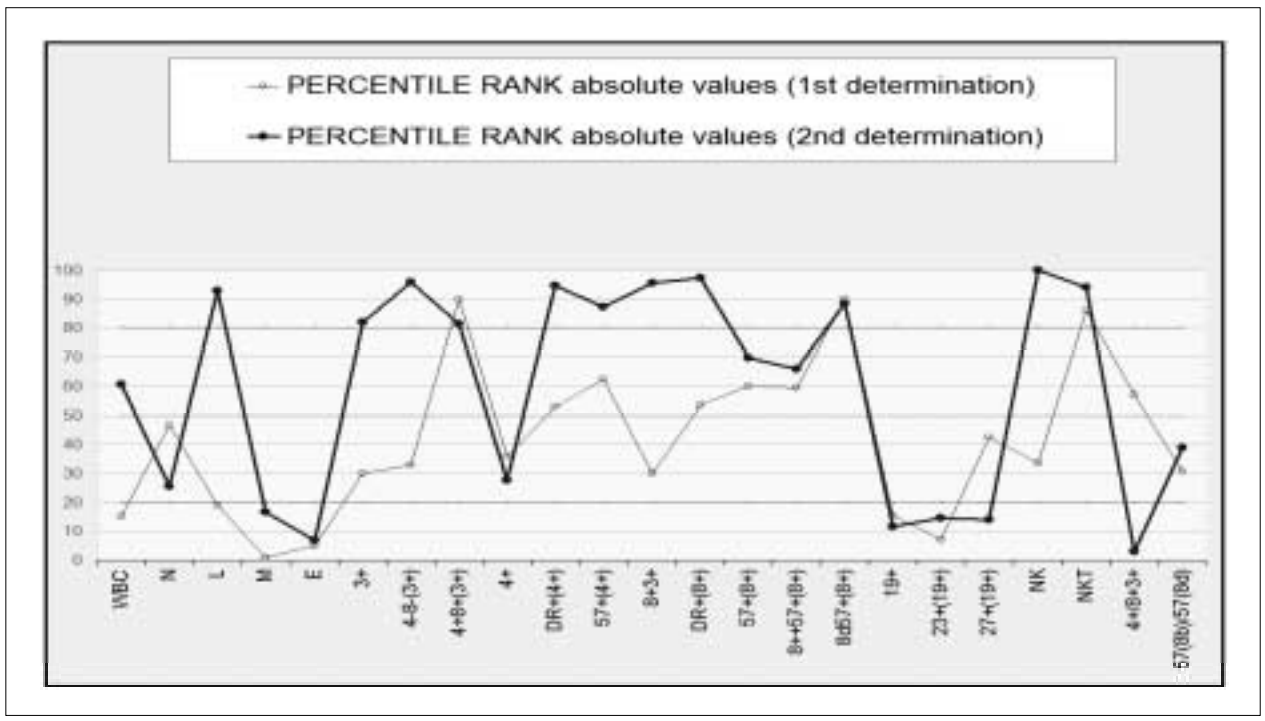

Figure 3 - Comparison between percentile ranks of absolute values related to two determinations of values in the same subject carried out at different times. Determinations are for a female child with Down's syndrome: the first was done when the girl was 2 years old, the second at 3 years old with CMV infection

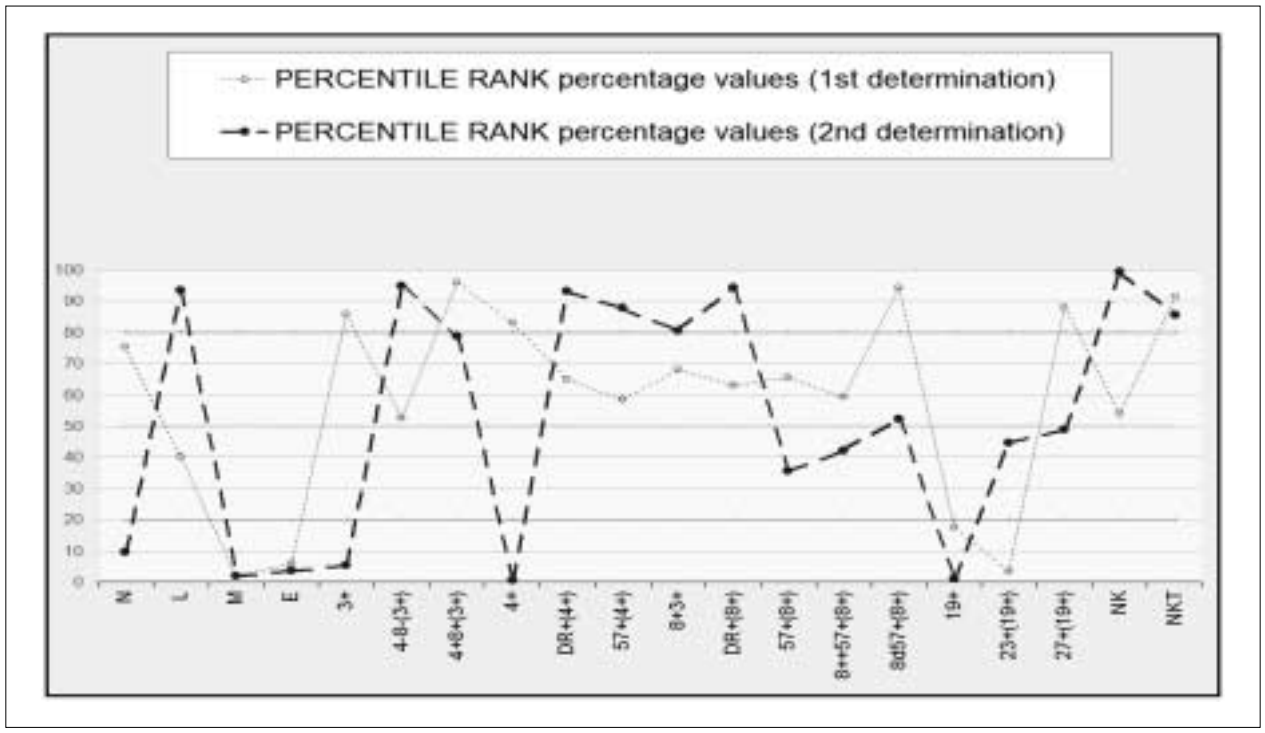

Figure 4 - Comparison of percentile ranks of percent values related to the same determinations as those illustrated in figure 3

Discussion

In 1997, Van den Hove *et al.* published a very interesting study in Cytometry[14], in which a multiparametric approach to the study of lymphocytes in the peripheral blood was reported. A technique of multivariable analysis was used which allowed a certain number of observations (subjects) and numerous variables (parameters) to be represented in a two-dimensional graph. Thus, already several years ago, there was a perceived need to analyse various parameters and represent them graphically in order to study their relations better and to have a view of them as a whole.

The approach described in the current report also represents an attempt to highlight relations between parameters but, above all, to demonstrate differences for each single parameter from the medians for the general population against which the individual subject is compared. Through the graphical representation of data, it is immediately possible to detect those parameters that differ significantly from the medians, consider and analyse them individually or as a whole. Furthermore, it is possible to search for associations between different groups of immunological parameters (e.g., the relationship between cellular parameters, phenotypes or activation status) and humoural parameters (e.g. the quantity and quality of immunoglobulin classes, presence or absence of autoantibodies); associations between immunological parameters and positivity or negativity for viral infections (EBV, CMV, HSV, VZV, HCV, etc.) can be revealed; and associations can be searched for between immunological parameters, life styles, current and previous therapies, vaccinations, etc. The identification of the position in which the subject's values lie with respect to those of various different cohorts of other subjects would not be possible if the data were analysed only with respect to reference ranges.

In the future, the data collected in the database could be increased by sharing the software with general practitioners, paediatricians, laboratory staff, hospital ward personnel, specialist physicians and yet other groups. The subsequent processing of the data, using the above described method, generates a series of graphs and a report that is available to all users, primarily for diagnostic, clinical, therapeutic and research purposes. More sharing would increase the size and potential of the database, providing greater reliability and stronger significance of the analyses. In diagnostic and clinical practice, this support would enable prompt comparison of a particular "case" with numerous, well-selected groups homogeneous with the particular subject under analysis.

The potential of ever more powerful instruments (increased number of fluorochromes), together with the increasing and greater availability of reagents and more sophisticated methods of analysis, such as the one described here, will make a "systematic" study of the immune system possible, also for diagnostic purposes. Such studies could not even have been dreamt of until recently.

References

1) Santagostino A, Garbaccio G, Pistorio A, et al. An Italian national multicenter study for the definition of a reference range for normal values of peripheral blood lymphocyte subsets in healthy adults. Haematologica 1999; **84**: 499-504.
2) Huppert FA, Solomou W, O'Connor S, et al. Aging and lymphocyte subpopulations: whole-blood analysis of immune markers in a large population sample of healthy elderly individuals. Exp Gerontology 1998; **33**: 593-600.
3) Bisset LR, Lung TL, Kaelin M, et al. Reference values for peripheral blood lymphocyte phenotypes applicable to the healthy adult population in Switzerland. Eur J Haematol 2004; **72**: 203-12.
4) Deneys V ; Mazzon AM, Marques JL, et al. Reference values for peripheral blood B-lymphocyte subpopulations : a basis for multiparametric immunophenotyping of abnormal lymphocytes. J Immunol Methods 2001; **253**: 23-36.
5) Jentsch-Ullrich K, Koenigsmann M, Mohren M, Franke A. Lymphocyte subsets' reference ranges in an age- and gender-balanced population of 100 healthy adults—a monocentric German study. Clin Immunol 2005; **116:** 192-7.
6) Denny T, Yogev R, Gelman R, et al. Lymphocyte subsets in healthy children during the first 5 years of life. JAMA 1992; **267:** 1484-8.
7) Shearer WT, Rosenblatt HM, Gelman RS, et al. Pediatric AIDS Clinical Trials Group. Lymphocyte subsets in healthy children from birth through 18 years of age: the Pediatric AIDS Clinical Trials Group P1009 study. J Allergy Clin Immunol 2003; **112:** 973-80.
8) Berrington JE, Barge D, Fenton AC, et al. Lymphocyte subsets in term and significantly preterm UK infants in the first year of life analysed by single platform flow cytometry. Clin Exp Immunol 2005; **140:** 289-92.
9) Robinson M, O'Donohoe J, Dadian G, et al. An analysis of the normal ranges of lymphocyte subpopulations in children aged 5-13 years. Eur J Pediatr 1996; **155:** 535-9.

10) Comans-Bitter WM, de Groot R, van den Beemd R, et al. Immunophenotyping of blood lymphocytes in childhood. Reference values for lymphocyte subpopulations. J Pediatr 1997; **130:** 388-93.

11) Erkeller-Yuksel FM, Deneys V, Yuksel B, et al. Age-related changes in human blood lymphocyte subpopulations. J Pediatr 1992; **120:** 216-22.

12) Loria A, Salas R. Patient data for evaluating reference values. Rev Invest Clin 1989; **41:** 381-5.

13) Grossi E, Colombo R, Cavuto S, Franzini C. The REALAB Project: a new method for the formulation of reference intervals based on current data. Clin Chem 2005; **51:** 1232-40.

14) Van den Hove LE, Symons F, Vandenberghe P, et al. Multivariate reconstruction of lymphocyte profiles in a two-dimensional graphical model as a tool for the investigation of lymphocyte subset distribution in health and disease. Cytometry 1997; **28**: 220-7.

Received: 26 March 2009 - Revision accepted: 25 June 2009
Correspondence: Riccardo Ortolani
Servizio di Immunologia Clinica, Dipartimento di Patologia Generale
Azienda Ospedaliero-Universitaria di Verona
P.le L.A. Scuro, 10
37134 Verona, Italy
e-mail: riccardo.ortolani@univr.it