Original

Prediction of the concentration of CD4 T lymphocytes based on set theory applied to the monitoring of patients with HIV

Javier Rodríguez-Velásquez[1,2], Signed Prieto-Bohórquez[1], Martha Melo-de Alonso[3], Carlos Pérez-Díaz[4], Darío Domínguez-Cajeli[3], Juan Bravo-Ojeda[4], Nancy Olarte-López[5], Aura Wilches-Betancourt[1], Laura Méndez-Pino[2] y Laura Valero-Morales[2]

Abstract

Introduction: A CD4 predictive methodology from the white cell blood count and lymphocyte blood counts was developed in patients with HIV/AIDS, seeking an alternative measure to flow cytometry.

Methods: Membership to four sets: A, B, C and D, of triplets of: cells/mm³, leukocytes/mm³, and CD4 cells/μL, was assessed in samples taken from 33 patients, collecting 3 to 5 samples per patient, for a total of 144 samples. The assessment of (A∪C) , (B∪D) y (A∪C)∩(B∪ D) allows for predictions based on the percentage of belonging to these groups. The results were arranged in descending order in nine ranges of 1000 leukocytes. The number of patients with accurate predictions and the ranges of greater effectiveness in prediction were established.

Results: The intersection (A∪C)∩(B∪D) showed effectiveness of 85.71% in predicting CD4 in the range of 4999-4000 leukocytes, 83.33% for 3999-3000, and 100% in the range lesser than 3,000.

Conclusion: The predictive ability and clinical usefulness of the methodology developed were confirmed for the prediction of T CD4 lymphocites, allowing to lower costs compared to flow cytometry in monitoring patients with HIV/AIDS over time.

Keywords: Flow cytometry, CD4, blood count, Prediction, set theory, HIV / AIDS.

Date received: April 13, 2015   *Date Approved: February 04, 2016*

Introduction

In mathematics, set theory[1] consists of certain basic concepts such as the notion of membership as well as basic operations such as union, intersection, difference as well as symmetrical difference[2], all of which define the relationships among the elements of a given set. Thanks to this structure, it becomes possible to discover certain sets on the basis of information implied or contained in more general or larger sets.

Applications of set theory are not restricted to topics related to the exact sciences; they have been applied to subject matter such as that of infectious diseases, particularly in illnesses like infection with the human immuno-deficiency virus (HIV). In 2011 alone, close to 34 million people were carriers of this infection and nearly 60% of those living with HIV were not aware of their serological status, a fact that represents a barrier for timely treatment. The most alarming statistics, though, come from Sub-Saharan and Southern Africa since they are the world´s most affected regions[3].

This illness results in a progressive immune suppression which is caused by the retrovirus (HIV) and it affects in this manner, not only the cells called T lymphocytes which are responsible for co-ordinating the cellular immune response but also the lymph nodes and the macrophages, the latter being the cells of the immune system which express CD4 cells. It affects follicular dendritic cells (FDC´s) as well.

Based on the facts mentioned above, it becomes important to evaluate the progression of the illness across time and to determine the most adequate anti-retroviral treatment. For this, it is necessary to analyze the total amount of white blood cells as well as total and also CD4 lymphocytes present in the patient´s blood. The first two measures are easily and inexpensively obtained from the peripheral blood count whereas to determine the CD4 cell count, it is necessary to carry out a specific test called flow cytometry which has disadvantages such as its high cost and low availability in developing countries[4]. Flow cytometry also has a low rate of cell transmission which is a measure of the individual behavior of these cells and it also requires trained personnel[5].

Given the concepts of set theory as well as that of HIV/AIDS, so as to establish a relationship between them, some parameters were determined based on studies of the hematological profile of patients over time, using as a starting point certain ranges of laboratory values. Previous observations had determined that patients having less than 1,500 total lymphocytes, usually have less than 500 CD4/microliter while those that have more than 1,500 total lymphocytes, usually have 200 CD4 lymphocytes per microliter[6].

With the purpose of finding more precise values in the CD4 count, several predictive models have been developed, among them, certain models of an automated nature which have an 83% degree of precision. One of the latter models was based on the viral load as well as the number of weeks since the CD4 count had begun to be determined[7]. Others models of an epidemiological nature which were used in Zambia and South Africa and which had degrees of precision between 76% and 82%, respectively, showed patterns of CD4 count variation, not only for individuals but also for populations with and without HIV/AIDS[8] and these results were confirmed in three U.S. cohorts[9].

Several studies have developed algorithms based on machine-learning with neural networks which relate the genotype with information on the treatment of this illness. These studies have obtained percentages of accuracy in predicting the viral load of 75%[10] and 69%[11]. Other algorithms which use a predictive approach to determine the success or failure of the treatment, have also included the dichotomous response of the virus[12], with an accuracy of 80%. These have also included the behavior at a genotypic level of HIV as a result of combined anti-retroviral therapy.[13] Even though the approach toward a specific prediction of the CD4 count was not a predominant factor, these studies have shown that it is possible to establish mathematical models or orders regarding different phenomena associated to the development of HIV/AIDS.

On the other hand, by means of physical and mathematical theories, diagnostic tools have been developed which are objective and reproducible. Specifically, Rodríguez *et al* have developed a method to predict CD4 lymphocyte count on the basis of individual values of total white blood cells in general as well as lymphocytes in the blood count by means of set theory. This was done by determining their membership in three different sets which define the behavior of leukocytes/cubic milliliter in regards to lymphocytes per cubic milliliter and also to CD4 cells per cubic milliliter.

To determine the effectiveness of prediction [6, 14], the registries were organized on the basis of size from larger to smaller in accordance with the number of leukocytes, defining ranges of 1,000 and evaluating the membership in the sets evaluated of the samples in each range. An upper range or limit was defined for values larger than 10,000 per cubic milliliter and a lower range or limit for values lower than 4,000 per cubic milliliter.

The clinical usefulness of this methodology especially in the lower ranges of leukocytes has been confirmed in two subsequent studies with 500[15] and 800 samples[14]. Both of these studies had percentages of effectiveness larger than 81.44% in 5 of the 9 ranges studied as well as an effectiveness larger than 91.89% in leukocyte ranges lower than 4,000 and, finally, an effectiveness of 100% in the prediction of the range lower than 3,000 [14,15].

From the above, one can see that the purpose of this paper is not only to apply the method developed based on set theory but also the mathematical prediction of CD4 T lymphocytes starting from triplets of different blood counts as well as to establish the usefulness of the method in carrying out follow-up of the evolution of patients across time.

Methods

Definitions

Ranges of leukocytes: ranges of 1,000 leukocytes, starting at 3,000 and until 10,000; values greater than 10,000 correspond to a single range and values less or equal to 2,999 make up another range.[15]

Sets:

A: $((x,y,z) \mid x \geq 6\ 800 \wedge y \geq 1\ 800)$

B: $((x,y,z) \mid x \geq 6\ 800 \wedge z \geq 300)$

C: $((x,y,z) \mid x < 6\ 800 \wedge y \leq 2\ 600)$

D: $((x,y,z) \mid x < 6\ 800 \wedge z \leq 570)$

Where (x,y,z) is defined as a triplet of values whereby x is the number of leukocytes, y is number of lymphocytes and z is the CD4-T lymphocyte count.

Procedure

The study was developed based on information from the blood count and flow cytometry of 33 patients, 21 of which had 4 samples and 12 of which had 5 samples, all taken at different times. The information was gathered from a database of previous research and was evaluated by an experienced, infectious disease physician. All the patients were on anti-retroviral treatment and their ages ranged between 27 and 46 years; 18 of the patients were men and 15 were women.

Based on previous research, the triplets of values were established for leukocytes per cubic millimeter, lymphocytes per cubic millimeter and CD4 cells per microliter in each sample. In addition, their membership in one of four sets was defined; A, B, C and D (see definitions above). Set A was defined by a number of leukocytes (x), a number of lymphocytes (y) and a number of CD-4 T lymphocytes (z), such that the number of leukocytes (x) was greater than or equal to 6,800 and the number of lymphocytes (y) was greater than or equal to 1,800. In the same manner, the other sets were defined as may be seen above.

Afterwards, we determined membership in the set (A U C), which allowed the evaluation of the specific relationships among the values of the leukocytes with regard to the lymphocytes. By the same token, we evaluated the set (B U D), which evaluates the specific relationships between leukocytes and CD4 cells. Finally, we carried out the prediction that, in mathematical terms, corresponds to the assumption of belonging to the set (A U C) intersection (B U D), in such a manner that the said intersection, establishes the mathematical relationship among the three values. In other words, with the union, all values are taken into account that belong to the set A with B and B with D; in contrast, with the concept of intersection, the only associations are with the values that simultaneously belong to the sets (A U C) and (B U D).

Next, the triplets´ sets were organized from higher to lower, according to the amount of leukocytes per cubic millimeter and they were grouped according to the defined ranges, with the goal of determining the number of triplets belonging to each set that was evaluated, according to each range. Lastly, the ability of the applied predictive methodology was established so as to predict accurately the evolution of the patients at different time periods. For the latter, the conditions seen among the patients whose evolution in time showed the more precise predictive percentages were studied as well as those where the prediction percentages were lower.

At a normative level, it must be pointed out that this research fulfills the conditions mentioned in Resolution Number 008430 dated 1993, particularly its paragraph number 11 which is related to research in human beings, since our study falls under the category of research without risks as it was based on information obtained in previous work. The mathematical calculations do not affect the patients and the handling of the information respects their integrity as well as their anonymity.

Results

When the 144 samples corresponding to the 33 patients were evaluated, the values of leukocytes in peripheral blood were found to be between 1,900 and 17,850 per cubic millimeter, the total lymphocytes were found to be between 70 and 5,910 per cubic millimeter and the CD4 T-cell count was found to be 21 and 1,366 per microliter. Due to number of cases used in the present research, a table was constructed so as to serve as an example and in it, the triplets of four specific patients were shown. In Table 1, it may be seen that all the values fall within the described ranges.

The set (A U C) showed prediction percentages that varied between 50% and 100%; the set (B U D) similarly, showed values between 50% and 100% and, finally, the set comprised of (A U C) intersection (B U D), reported values between 50% and 100%, also (Table 2).

N°	Total leukocytes (mm³)	Total lymphocytes (mm³)	CD4 T- cell count(µL)
Table 1. Complete sample of the triplets of the hematologic indicators obtained for four patients; the first two have four samples and the last two have five samples; study data are defined in leukocytes/ mm³, lymphocytes/mm³ and CD4 T lymphocytes /µL.			
10-1,1	4200	1667	65
10-1,2	4740	1600	128
10-1,3	2880	1150	54
10-1,4	7670	3810	554
101-1,1	5200	1700	569
101-1,2	4980	2520	964
101-1,3	6170	2720	1366
101-1,4	5680	2300	736
158- 1,1	5600	2161	293
158- 1,2	9080	2160	409
158- 1,3	6680	2080	266
158- 1,4	4980	1090	172
158- 1,5	6450	1750	234
160- 1,1	7820	3980	378
160- 1,2	5500	3210	285
160- 1,3	6490	2870	285
160- 1,4	5920	2500	218
160- 1,5	5830	2960	237

We found prediction percentages equal to or greater than 75% among 6 of the 9 ranges studied; of the latter, the ranges of leukocytes lower than 5,000 showed values belonging to the set

(A U C) intersection (B U D) which were progressively greater; in this way, for ranges that were lower than 5000, 4000 and 3000, a prediction was established of CD4 cell count under 570 which resulted in effectiveness percentages of 85.71%, 83.33% and 100%, respectively. We found, besides the above, that the range between 8999 and 8000, showed one of the highest percentages of effectiveness which was 88, 89% (Table 2).

As a result of the observation of the predictive capacity of this methodology in the same patient at periods across time, it was determined that patients whose illness was in more advanced stages, a fact that may be observed from the low CD4 values reported in addition to low leukocyte values (lower than 5,000 per cubic milliliter) turned out to be the patients who could be followed in a more reliable fashion at different time periods using our methodology. This is due to the fact that the latter set of patients show the highest predictive values. Thus, it was the case that patients with leukocyte measurements lower than 5,000 had a predictive value

between 83.33% and 100%. In contrast, those carriers of the virus who hadn´t developed advanced illness, showed lower predictive values since they were associated with leukocyte counts all above 5,000 (Tables 1 and 2).

Discussion

This is the first study to apply a prediction based on set theory and that establishes a number of CD4 per microliter, on the basis of leukocyte and lymphocyte count. This was done in 4 or 5 consecutive samples across different time periods, thus confirming the clinical applicability for the follow-up of the evolution of patients with HIV/AIDS. With this methodology, it becomes possible to contribute to a timely therapeutic intervention which is easily available to patients with HIV, specifically when their CD4 count is at a minimal level.

We found prediction percentages equal to or higher than 75% among 6 of the 9 ranges evaluated; the latter ranges correspond mainly to those where the leukocyte count was lower than 5,000 so that we were able to establish predictions of 85%, of 83% and of 100% for the ranges of 5000, 4000 and 3000 leukocytes, respectively.

Table 2. Asignation of elements to the operations of union and intersection of sets among ranges of leukocytes taken from patients							
Leukocyte Range	**Number of cases per range**	**(A∪C)**		**(B∪D)**		**(A∪C)∩(B∪D)**	
		No.	**%**	**No.**	**%**	**No.**	**%**
10,000 or more	4	4	100.00	3	75.00	3	75.00
9,999 to 9,000	9	7	77.78	8	88.89	7	77.78
8,999 to 8,000	9	8	88.89	8	88.89	8	88.89
7,999 to 7,000	4	4	100.00	2	50.00	2	50.00
6,999 to 6,000	22	11	50.00	20	90.91	11	50.00
5,999 to 5,000	27	19	70.37	22	81.48	17	62.96
4,999 to 4,000	35	34	97.14	30	85.71	30	85.71
3,999 to 3,000	6	6	100.00	5	83.33	5	83.33
2,999 or less	5	5	100.00	5	100.00	5	100.00
Total	**121**	**98**	**80.99**	**103**	**85.12**	**88**	**72.73**

The mathematical approach that we used allowed the simplification of the problem and its application to a clinical issue and this remained true, without regard to age, type of treatment, gender, viral load or hemoglobin. Our results have high clinical impact more so considering that developing countries do not have access to flow cytometry and similar types of tests. The analysis of patients´ evolution across time showed that lower prediction percentages were obtained in those who had already developed AIDS and who, because of thissituation, have lower CD4 values as well as lower values of leukocyte count. This observation has clinical importance because the latter group requires a more meticulous follow-up over time so as to establish the efficiency over a short time period of anti-retroviral treatment. In contrast, patients with laboratory results of leukocytes above 5,000 show lower predictive values, a situation which indicates the need to develop an improvement in the methodology posed, so that the prediction percentages in these ranges will increase.

At present, developing countries have problems with clinical follow-up of patients as well as with therapeutic decision-making with regard to the administration of anti-retroviral treatment due primarily to the high cost of flow cytometry for CD4 cell count. The scant coverage of this analysis in countries which need it the most reduces the possibility of anti-retroviral treatment[5], in spite of efforts promoted by the WHO regarding access to therapy[10]. Due to the above, it becomes of great practical utility to have easily accessible methods and methods that will be more available at a global level; from this, it can be seen that the prediction of CD4 count based on the number of leukocytes and lymphocytes from a regular blood count, constitutes a practical as well as an inexpensive option for these issues.

Set theory allowed us to understand the mathematical behavior of the phenomenon that we are describing; because of their characteristics, the samples used were always more likely to belong to sets A and C since these are the ones which establish the behavior of the lymphocytes; on the other hand, the values diminish when sets B and D were evaluated since the latter have to do with the behavior of CD4 cells. As a result, the behavior of the intersection between both sets, shows percentages equal to or lower than the set (B U D). Due to simultaneous analysis of leukocyte and lymphocyte behavior associated with evaluation of CD4 cells in the different ranges, we showed that set theory allows us to establish a comparison of the behavior of CD4 cells by indicating that the latter will be less than 570 CD4/microliter with a confidence of 100% whenever the leukocyte count is lower than 3000 per cubic millimeter.

During the last decade, models have been developed which have had a statistical and empirical approach and these have based their prediction of CD4 on epidemiological methods or those of neural networks[10-13] and thus, the latter have implied a restriction on their applicability to the population samples that are used in such studies. In contrast, the proposed methodology was developed from inductive reasoning based on physico-mathematical models and have obtained on the basis of only 7 cases, predictions applied to each one of these cases6 which were, subsequently, confirmed by

increasingly larger population studies so as to guarantee their clinical applicability[14,15]. An additional study which was conducted based also on set theory as well as on the theory of probability, allowed us by means of calculating the quadratic standard deviation, to show that the behavior of the sets was not equal or equi-probable but that, instead, there was a "loaded" behavior which allowed the establishment of predictions whose effectiveness will depend of the range that is to be evaluated[17].

References

1. Garcia C. Teoria de Conjuntos. Universidad Autonoma del Estado de Hidalgo. 12 paginas. Recuperado el 5 de mayo de 2015. En: http://repository.uaeh.edu.mx/bitstream/bitstream/handle/123456789/14886/LECT107.pdf?sequence=1

2. Rodriguez J. Diferenciacion matematica de peptidos de alta union de MSP-1 mediante la aplicacion de la teoria de conjuntos. Inmunologia 2008;27:63-68.

3. Informe de ONUSIDA para el dia mundial del SIDA /2011. Como llegar a cero mas rapido. Mas inteligente. Mejor. Programa conjunto de las Naciones Unidas sobre VIH/SIDA. 52 paginas. Recuperado el 5 de mayo de 2015. En: http:// www.unaids.org/sites/default/files/media_asset/JC2216_WorldAIDSday_report_2011_es_0.pdf

4. Zijenah L, Kadzirange G, Madzime S, Borok M, Mudiwa Ch, Tobaiwa O, et al. Affordable flow cytometry for enumeration of absolute CD4+ T-lymphocytes to identify subtype C VIH-1 infected adults requiring antiretroviral therapy (ART) and monitoring response to ART in a resource-limited setting. J Transl Med 2006,4:33.

5. Castillo N, Barriga G, Solis M, Arumir C. Carga viral en el en el sindrome de inmunodeficiencia adquirida. Estudio comparativo de tres metodos. Rev Mex Patol Clin. 1998; 45: 155-156.

6. Rodriguez J, Prieto S, Bernal P, Perez C, Correa C, Vitery S. Teoria de conjuntos aplicada a poblaciones de leucocitos, linfocitos y CD4 de pacientes con VIH. Prediccion de linfocitos T CD4, de aplicacion clinica. Rev Fac Med 2011;19:148-156.

7. Singh Y, Mars M. Support vector machines to forecast changes in CD4 count of VIH-1 positive patients. Sci Res Essays 2010; 5: 2384-2390.

8. Williams BG, Korenromp EL, Gouws E, Schmid GP, Auvert B and Dye C. VIH Infection, Antiretroviral Therapy, and CD4+ Cell Count Distributions in African Populations. J Infect Dis 2006;194:1450-1458.

9. Williams BG, Korenromp EL, Gouws E, Dye C. The rate of decline of CD4 T-cells in people infected with VIH. Cornell university Library. Recuperado el 5 de mayo de 2015. En: http://arxiv.org/ftp/arxiv/papers/0908/0908.1556

10. Wang D, De Gruttola V, Hammer S, Harrigan R, Larder B, Wegner S, et al. A Collaborative VIH Resistance Response Database Initiative: Predicting Virological Response Using Neural Network Models. Poster presentation at: The XI International VIH Drug Resistance Workshop, Seville. Recuperado el 12 de enero de 2014. En: https://www.hivrdi.org/abstract_3.htm

11. Larder B, Wang D, Revell A, Montaner J, Harrigan R, De Wolf F, et al. The development of artificial neural networks to predict virological response to combination VIH therapy. Antivir Ther 2007;12: 15-24.

12. Altmann A, Rosen-Zvi M, Prosperi M, Aharoni E, Neuvirth H, Schulter E, et al. Comparison of Classifier Fusion Methods for Predicting Response to Anti HIV-1 Therapy. PLoS ONE 2008; 3: e3470.

13. Altman A, Daumer M, Beerenwinkel N, Peres Y, Schulter E, Buch J, et al. Predicting the Response to Combination Antiretroviral Therapy: Retrospective Validation of geno2pheno-THEO on a Large Clinical Database. J Infect Dis 2009; 199: 999-1006.

14. Rodriguez J, Prieto S, Correa C, Perez C, Mora J, Bravo J, et al. Predictions of CD4 lymphocytes' count in HIV patients from complete blood count. BMC Medical Physics 2013; 13:3.

15. Rodriguez J, Prieto S, Correa C, Forero M, Perez C, Soracipa Y, et al. Teoria de conjuntos aplicada al recuento de linfocitos y leucocitos: prediccion de linfocitos T CD4 de pacientes con VIH/SIDA. Inmunologia 2013; 32: 50-56.

16. Progress on global access to antiretroviral therapy: an update on "3 by 5" and beyond, World Health Organization, 2006. 84 paginas. Recuperado el 5 de mayo de 2015. En: http://www.who.int/hiv/fullreport_en_highres.pdf

17. Rodriguez J, Prieto S, Bernal P, Perez C, Correa C, Alvarez L, et al. Prediccion de la concentracion de linfocitos T CD4 en sangre periferica con base en la teoria de la probabilidad. Aplicacion clinica en poblaciones de leucocitos, linfocitos y CD4 de pacientes con VIH. Infection 2012; 16: 15-22.