

INSTITUTO DE ONCOLOGIA Y RADIOBIOLOGIA

## **Análisis multivariado de datos de laboratorio clínico en el diagnóstico de neoplasias malignas. II. Cáncer broncopulmonar**

Por:

Lic. EMILIO GRUEIRO AZCANO\*, Dra. SUSANA ESOUENAZI MITRANI\*\* y Dr. AGUSTIN LAGE DAVILA\*\*\*

Grueiro Azcano, E. y otros. *Análisis multivariado de datos de laboratorio clínico en el diagnóstico de neoplasias malignas. II. Cáncer broncopulmonar*. Rev Cub Med 23: 2, 1984.

Las técnicas de análisis multivariado de datos, ofrecen la posibilidad de evaluar la asociación simultánea de un conjunto de pruebas de laboratorio con determinadas enfermedades y producir funciones discriminantes compuestas de varios resultados de laboratorio, que pueden tener una asociación con la enfermedad en cuestión. Esto es superior a cada una de las pruebas de laboratorio por separado. La característica de la función discriminante consiste en que es capaz de tomar en cuenta la tendencia del fenómeno, ya que el valor de una prueba, aunque esté dentro del rango normal, si se mueve en el sentido de la alteración, influirá en el resultado que tendrá Y (la variable dependiente). Los estudios de laboratorio fueron hechos a más de 3 000 pacientes en el momento de su Inscripción en el IOR. En este artículo se presentan los resultados correspondientes a un grupo de 138 pacientes que presenten o no cáncer broncopulmonar y 40 sujetos "sanos", a los cuales se les determinó: ceruloplasmina, eritrosedimentación, cobre, zinc, calcio, magnesio, proteína C reactiva, deshidrogenasa láctica y catalasa. Las muestras estudiadas fueron divididas al azar en dos grupos. A y B, que se emplearon para obtener dos funciones discriminantes separadamente, y luego la función obtenida con los datos del grupo A se evaluó contra la función obtenida con los datos del grupo B y viceversa. Con este estudio se obtuvo una función discriminante con los resultados de la eritrosedimentación, y las determinaciones séricas de deshidrogenasa láctica, catalasa y proteína C reactiva, que proporcionó un diagnóstico correcto en el 88% de los casos estudiados, cuando la Y tomaba un valor mayor de 0,4.

\* Investigador. Laboratorio de análisis bioquímicos clínicos. IOR.

\*\* Doctora en Ciencias Físico Químicas. Jefa del Laboratorio de análisis bioquímicos clínicos. IOR.

\*\*\* Médico Especialista en Bioquímica. Jefe del departamento de bioquímica. IOR.

## INTRODUCCION

La búsqueda de pruebas de laboratorio para el diagnóstico, pronóstico y seguimiento de los pacientes que presentan neoplasias malignas ha sido una de las principales direcciones de trabajo de la bioquímica en Oncología.<sup>1</sup> Sin embargo, hasta la fecha, ninguna prueba de laboratorio ha mostrado una asociación suficientemente fuerte con la presencia de neoplasias malignas que permita que ésta sea utilizada por sí sola como criterio de diagnóstico, pronóstico, o ambas. Con excepción de algunos tumores infrecuentes que pueden ser diagnosticados y seguidos con dosificaciones de gonadotropina coriónica u otras secreciones ectópicas de hormonas, la mayor parte de las neoplasias carece de marcadores séricos útiles.

Independientemente de las investigaciones en curso para la detección de nuevos marcadores bioquímicos de las neoplasias, las técnicas de análisis multivariados de datos ofrecen la posibilidad de evaluar la asociación simultánea de un conjunto de pruebas de laboratorio con determinada enfermedad, e intentar producir funciones discriminantes compuestas de varios resultados de laboratorio que puedan tener una asociación con la entidad en cuestión, mejor que cada una de las pruebas de laboratorio por separado. La característica de la función discriminante consiste en que es capaz de tomar en cuenta la tendencia del fenómeno, ya que el valor de una prueba, aunque esté dentro del rango normal, si se mueve en el sentido de la alteración en estudio, éste influirá en el resultado que tendrá la Y (variable independiente).

Se ha publicado, recientemente, un estudio de varias pruebas de laboratorio en cáncer mamario,<sup>2</sup> evaluándolas simultáneamente mediante un programa de regresión multilínea, que produjo una función discriminante capaz de diagnosticar correctamente el carácter maligno de un nódulo de mama en el 86% de los casos. Esta función está actualmente en uso en la clínica. En el presente trabajo mostramos los resultados de un estudio similar en pacientes de cáncer broncopulmonar. Se muestra la función discriminante obtenida con los resultados de la eritrosedimentación y las determinaciones séricas de deshidrogenasa láctica, catalasa y proteína C reactiva la cual proporcionó un diagnóstico adecuado en el 88% de los casos.

## MATERIAL Y METODO

**Muestra:** Los estudios de laboratorio fueron realizados en más de 3 000 pacientes en el momento de su inscripción en el Instituto de Oncología y Radiobiología. Para este trabajo se seleccionaron los pacientes que presentaban cáncer broncopulmonar, que fueron 102, y 36 pacientes negativos, que habían sido inscritos por sospechosos de esta enfermedad, y, además; 40 donantes "sanos" del Banco Provincial de Sangre de Ciudad de La Habana, todos adultos y de ambos sexos.

**Exámenes de laboratorio:** Las muestras de suero proveniente de cada individuo fueron empleadas para determinar ceruloplasmina, eritrosedimentación, cobre, deshidrogenasa láctica, zinc, catalasa, calcio, magnesio y proteína C reactiva.

Para la determinación de la catalasa se empleó el método descrito por uno de los autores. La actividad de ceruloplasmina se midió a través de la oxidación de la p-fenildiamina a pH 5,0. La proteína C reactiva se evaluó por el método de doble inmunodifusión (Ouchterlony), y se empleó un monoantisuero específico (Difeo).

Las técnicas estándares utilizadas fueron objeto de control de calidad, según lo establecido por el MINSAP.<sup>8</sup> Las técnicas modificadas (ceruloplasmina), o creadas (catalasa), fueron objeto de estudios estadísticos de precisión y obtención de valores normales, que han sido presentados en eventos científicos y serán objeto de otras publicaciones.

**Análisis de los datos:** De cada individuo se dispuso de los nueve datos de laboratorio antes mencionados (variables predictoras) y se evaluó su capacidad para clasificar a cada individuo en grupos que presentaron o no una neoplasia maligna. Este problema se abordó por el método de análisis de discriminante.<sup>11</sup> Se definió una variable (Y) que toma valor cero si el individuo no presenta una neoplasia maligna, y toma valor cercano a uno si el individuo presenta cáncer broncopulmonar.

Sea entonces:

$$Y = B_1X_1 + B_2X_2 \dots \dots B_pX_p + C$$

donde  $X_1 \dots \dots X_p$  son los valores de las variables predictoras (análisis de laboratorio),  $B_1 \dots \dots B_p$  son coeficientes y C es un término independiente.

Los valores de los coeficientes B y el término independiente C fueron calculados con un programa de regresión multilíneal, paso a paso<sup>10</sup>, escrito en lenguaje Fortran IV para la computadora CID-201-B. Este programa calcula la recta de regresión entre la variable dependiente y la variable predictora mejor asociada con Y; luego, introduce la segunda variable predictora mejor asociada con Y y calcula una nueva recta de regresión al eliminar el efecto de la asociación entre la nueva variable introducida y las variables que se introdujeron en los pasos precedentes, y después introduce la variable siguiente en orden de asociación con Y, y así sucesivamente.

En cada paso el programa calcula e imprime el valor del coeficiente de correlación múltiple, el porcentaje de la varianza de Y explicado por función lineal obtenida y el valor de la "S" de Tukey.<sup>11</sup> Se retuvieron como las funciones más adecuadas aquellas que minimizan la "S" de Tukey.

Las muestras estudiadas fueron divididas al azar en dos grupos A y B, que se emplearon para obtener, dos funciones discriminantes, separadamente, y luego la función

obtenida con los datos del grupo A fue evaluada contra los individuos del grupo B, y viceversa. Esto evita que el valor predictivo de una función discriminante se evalúe con datos de la misma muestra de la que la función se obtuvo, lo cual usualmente sobrestima su utilidad.

El valor predictivo de las funciones obtenidas se evaluó mediante el programa DISCRIM \*

Este programa, dado un conjunto de individuos de los que se conoce su diagnóstico y los valores de las pruebas de laboratorio incluidas en la función discriminante a evaluar, compara el valor Y real de cada individuo (0 a 1) con el valor calculado de la media y la desviación estándar del valor Y calculado en el grupo de individuos que presentaban o no una neoplasia maligna y, luego, calcula los porcentajes de clasificación correcta e incorrecta, falsos positivos y falsos negativos que se obtendrían si se emplearan como criterio de la discriminación el tener un valor de Y mayor que 0,1, 0,2, 1,0 es decir, evalúa la eficacia de la función discriminante para diferentes fronteras de discriminación.

## RESULTADOS

Los aspectos esenciales de las técnicas empleados se resumen en el cuadro I.<sup>3-7</sup>

La asociación de las pruebas de laboratorio incluidas en este estudio con la presencia de cáncer broncopulmonar, puede evaluarse a través de los coeficientes de correlación de cada una de ellas con la variable Y definida arbitrariamente como valor 1,0 o próximo a 1,0 en presencia de cáncer broncopulmonar, y 0,0 o próximo a 0,0 en los sujetos sanos o afectados de enfermedades broncopulmonares no malignas [cuadro II].

Tanto los 40 sujetos sanos, como los 36 pacientes sospechosos de cáncer broncopulmonar que resultaron negativos, así como los 102 pacientes de cáncer broncopulmonar, fueron distribuidos aleatoriamente en dos grupos, A y B en ambos, se realizó un procesamiento independiente por el programa de regresión multilínea. El cuadro I' recoge las variables que fueron retenidas en ambos grupos en la ecuación de la función discriminante que mostró el valor mínimo de la "S" de Tukey.<sup>11</sup> Después de estas variables, en el grupo A se incluyeron, en orden sucesivo, el zinc y la ce- ruloplasmina, y en el grupo B, la ceruloplasmina, la catalasa y el calcio. Nótese, sin embargo, que las tres variables más asociadas en ambos casos fueron las mismas: la eritrosedimentación, la actividad de deshidrogenasa láctica y la proteína C reactiva, a lo cual se añade la catalasa en uno de los grupos. Este conjunto de cuatro variables de laboratorio explica el 50% de la varianza total de la muestra.

---

\* El programa DISCRIM fue desarrollado en el Centro de Cálculo del IOR y se encuentra disponible.

**CUADRO I**  
**DATOS ESENCIALES DE LAS TECNICAS EMPLEADAS**

Determinación	Base	Referencia
Eritrosedimentación	Velocidad de sedimentación de eritrocitos en tubos de Westergreen	3
Cobre	Espectrometría de absorción atómica	4
Zinc	Espectrometría de absorción atómica	4
Calcio	Reacción colorimétrica con alizarina	5
Magnesio	Reacción colorimétrica con titán amarillo	6
Deshidrogenasa láctica	Velocidad de oxidación del NADH, absorción a 340 nm	7

**CUADRO II**  
**COEFICIENTES DE CORRELACION**

Prueba	Coefficiente de correlación
Eritrosedimentación	0,630
Zinc	0,292
Cobre	0,531
Magnesio	0,030
Calcio	0,089
Proteína C Reactiva	0,584
Ceruloplasmina	0,574
Catalasa	0,387
Deshidrogenasa láctica	0,440

Cuadro de los coeficientes de correlación entre los valores de cada una de las determinaciones de laboratorio y la variable dependiente Y que se define con valor 0,0 para los controles y 1 para los pacientes de cáncer broncopulmonar.

La obtención de dos funciones discriminantes con grupos independientes de pacientes permitió evaluar cada ecuación con un grupo de pacientes diferentes al grupo del cual se obtuvo, como se muestra en el cuadro IV. Nótese que las funciones discriminantes obtenidas producen un 86% de diagnóstico correcto e igualmente, la separación de los valores medios de la función discriminante Y en los pacientes y en los controles.

Los resultados obtenidos con las funciones discriminantes de los grupos A y B justifican el agrupamiento de todos los casos en una muestra única, con la cual se obtiene la función discriminante final que se presenta en el cuadro V.

El cuadro VI nos muestra tres funciones discriminantes en las cuales, además de la eritrosedimentación, la proteína C reactiva, la deshidrogenasa láctica (LDH) y la catalasa, se incorporaron, sucesivamente, la ceruloplasmina, el calcio y el zinc.

Las funciones discriminantes, tales como las halladas en este trabajo, permiten obtener para cada nuevo paciente un valor Y que, si es cercano a 1,0, indica alta probabilidad de que el paciente sea portador de una neoplasia maligna, y si es cercano a 0,0 indica poca probabilidad de este evento. En estos datos, como se ve en el gráfico 1, a medida que aumenta el valor de Y aumenta el porcentaje de pacientes que resultaron tener positivamente una neoplasia maligna. Los individuos cuya combinación de resultados de laboratorio produjo un valor de Y superior a 0,5 presentaban una neoplasia maligna en más del 90%.

Este límite de 0,5, que se emplea para discriminar si el sujeto pertenece probablemente a uno u otro grupo, es un límite arbitrario. El gráfico 2 muestra cómo se comporta el índice de diagnóstico correcto, los falsos positivos y los falsos negativos, a medida que vamos corriendo la frontera de discriminación desde  $Y = 0,1$  hasta  $Y = 1,0$ . Nótese que el máximo nivel de diagnóstico correcto (86%) se obtiene cuando clasificamos como positivos a los individuos con valores de Y, superiores a 0,4.

Tiene interés analizar cómo se comportan los sujetos que fueron estudiados en el Instituto por sospecha de cáncer broncopulmonar y que resultaron al final no presentar esta enfermedad. Como se muestra en el gráfico 3, los individuos de este grupo de "sospechosos" negativos mostraron un valor medio de Y ligeramente superior al de voluntarios saludables, pero inferior al de los pacientes de cáncer broncopulmonar. Si se toma en cuenta solamente el grupo de sospechosos y el de pacientes (es decir, al eliminar los 40 voluntarios normales), el índice de diagnóstico correcto de la función discriminante obtenida baja a 82%.

CUADRO III

FUNCIONES DISCRIMINANTES OBTENIDAS CON DOS GRUPOS DE MUESTRAS

Prueba	Coficiente de regresión	Varianza explicada	Coficiente de correlación	Coficiente de regresión	Varianza explicada	Coficiente de correlación
Eritrosedimentación	0,00502 ± (0,00155)	40,0%	0,633	0,00556 ± (0,00154)	39,5%	0,628
Deshidrogenasa láctica	0,00094 ± (0,00050)	45,1%	0,672	0,00113 ± (0,00063)	42,8%	0,654
Proteína C reactiva	2,25643 ± (0,4500 )	48,5%	0,646	0,20394 ± (0,12334)	44,6%	0,668
Catalasa	0,08773 ± (0,04838)	50,4%	0,710	—	—	—
Término independiente: 0,08389				Término independiente: 0,01045		

Los pacientes y los individuos controles fueron distribuidos al azar en dos grupos, A y B, compuestos cada uno de 51 pacientes y 38 controles, y los datos de las 9 determinaciones de laboratorio de los individuos de cada grupo fueron empleados para obtener de cada uno los coeficientes de la función discriminante  $Y = f$  (eritrosedimentación, LDH, PCR, ... etc.) que mejor se ajustara a los datos. En el cuadro se muestran los coeficientes de la ecuación que minimizan la "S" de Tukey y que se considera el mejor compromiso entre un buen ajuste de datos y un número reducido de determinaciones de Laboratorio. El cuadro indica, igualmente, el porcentaje de las varianzas de Y que se explica por la función obtenida en cada inclusión una nueva variable y el coeficiente de correlación múltiple obtenido.

CUADRO IV

	Controles			Pacientes			Cáncer	Pulmón	
	N	Y media	Desviación estándar	N	Y media	Desviación estándar	Diag. correcto*	Falso-negativo	Falso-positivo
Ecuación grupo A vs. datos grupo B (S-Tukey mínima)	38	0,298	0,135	51	0,773	0,331	86,5	7,9	5,6
Ecuación grupo B vs. datos grupo A (S-Tukey mínima)	38	0,299	0,114	51	0,742	0,286	86,5	7,9	5,6
Ecuación grupo A vs. datos grupo B (final)	38	0,278	0,114	51	0,668	0,277	86,5	7,9	5,6
Ecuación grupo B vs. datos grupo A (final)	38	0,296	0,118	51	0,755	0,284	85,4	6,7	7,9

\* El máximo nivel de diagnóstico correcto se obtiene tomando  $Y = 0,4$  como frontera de discriminación.

**EVALUACION DE LAS FUNCIONES DISCRIMINANTES**

El cuadro de los valores promedio de las Y obtenidas para cada individuo del grupo B cuando se le aplica la ecuación obtenida con los datos del grupo A y viceversa. N es el número de casos en cada evaluación. Al final se indica el porcentaje de casos que se clasifica correctamente como positivo o negativo si se consideran positivos los casos con Y mayor de 0,4, y los porcentajes de falso-positivo, o negativos si se consideran positivos los casos con Y mayor de 0,4, así como porcentajes de falso-positivos y falsonegativos en ese caso. Las cuatro ecuaciones evaluadas fueron las siguientes:

*Ecuación grupo A (S-Tukey mínima)*

$$Y = 0,00502 \text{ (eritrosed.)} + 0,00094 \text{ (LDH)} + 0,25643 \text{ (PCR)} + 0,08773 \text{ (catalasa)} + 0,08389$$

*Ecuación grupo A (final)*

$$Y = 0,00426 \text{ (eritrosed.)} + 0,00082 \text{ (LDH)} + 0,21000 \text{ (PCR)} + 0,08963 \text{ (catalasa)} - 0,00126 \text{ (zinc)} + 0,00221 \text{ (ceruloplasmina)} - 0,01796$$

*Ecuación grupo B (S-Tukey mínima)*

$$Y = 0,00556 \text{ (eritrosed.)} + 0,00113 \text{ (LDH)} + 0,20394 \text{ (PCR)} + 0,01050$$

*Ecuación grupo B (final)*

$$Y = 0,00464 \text{ (eritrosed.)} + 0,00053 \text{ (LDH)} + 0,19993 \text{ (PCR)} + 0,00381 \text{ (ceruloplasmina)} + 0,05828 \text{ (catalasa)} + 0,00124 \text{ (calcio)} - 0,07963.$$

**CUADRO V**  
**ECUACION FINAL**

X	(1)	Eritrosedimentación: 0,005062 (B1)
X	(2)	Deshidrogenasa láctica: 0,000852 (B2)
X	(3)	Proteína C reactiva: 0,23950 (B3)
X	(4)	Catalasa: 0,080154 (B4)
		Término independiente — 0,049887 (C)

Varianza explicada: 48%

Coefficiente de correlación múltiple = 0,692

*Función discriminante final, según la S de Tukey*

El cuadro de los coeficientes de las cuatro variables que se retienen en la ecuación que mejor ajusta los datos a los grupos A y B de pacientes controles.

**CUADRO VI**

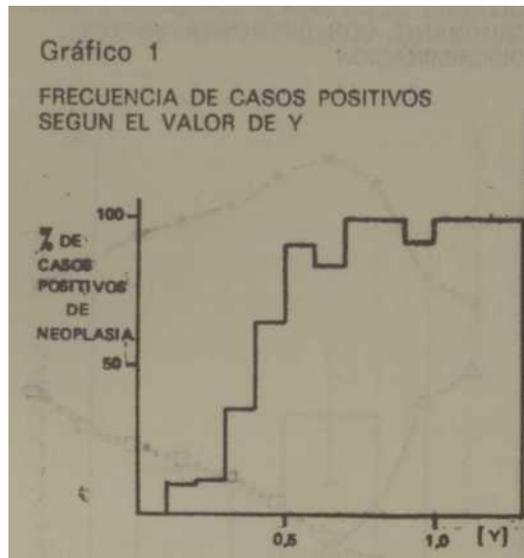
**FUNCIONES DISCRIMINANTES OBTENIDAS CON TODOS LOS PACIENTES INVESTIGADOS**

$$Y = 0,00436 \text{ (eritrosed.)} + 0,00283 \text{ (Cpl)} + 0,21525 \text{ (PCR)} + 0,06982 \text{ (catalasa)} + 0,00072 \text{ (LDH)} - 0,13216$$

$$Y = 0,00443 \text{ (eritrosed.)} + 0,00249 \text{ (Cpl)} + 0,21600 \text{ (PCR)} + 0,00114 \text{ (calcio)} + 0,07419 \text{ (catalasa)} + 0,000719 \text{ (LDH)} - 0,12762$$

$$Y = 0,000443 \text{ (eritrosed.)} + 0,00214 \text{ (Cpl)} + 0,20448 \text{ (PCR)} + 0,00116 \text{ (calcio)} - 0,00055 \text{ (zinc)} + 0,07792 \text{ (catalasa)} + 0,000723 \text{ (LDH)} - 0,04673$$

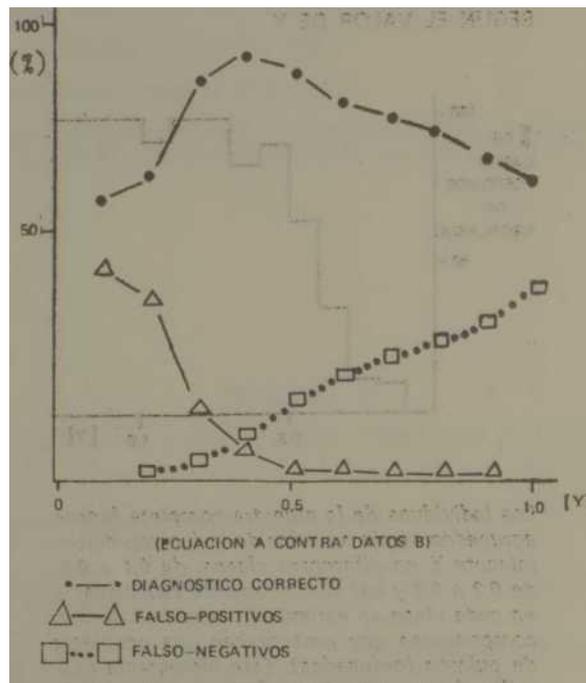
Por último, el gráfico 4 muestra la distribución de frecuencia de la función discriminante Y. Nótese que en el grupo de individuos controles la distribución es aguda con una moda entre 0,2 y 0,3, mientras que en el grupo de pacientes de cáncer broncopulmonar la distribución se extiende, uniformemente, en un amplio rango de valores de Y. La suma de estas dos curvas produce la distribución de valores en toda la muestra que se aprecia en el gráfico.



*Los individuos de la muestra completa fueron agrupados según su valor de la función discriminante Y en diferentes clases, de 0,1 a 0,2, de 0,2 a 0,3 y así sucesivamente (abscisas) y en cada clase se estimó el porcentaje de sus componentes que presentaban una neoplasia de pulmón (ordenadas). Este histograma describe el comportamiento de la muestra e indica el aumento de la probabilidad de neoplasia con el aumento del valor de la función Y, pero sus porcentajes no pueden ser extrapolados a la población en general.*

Gráfico 2

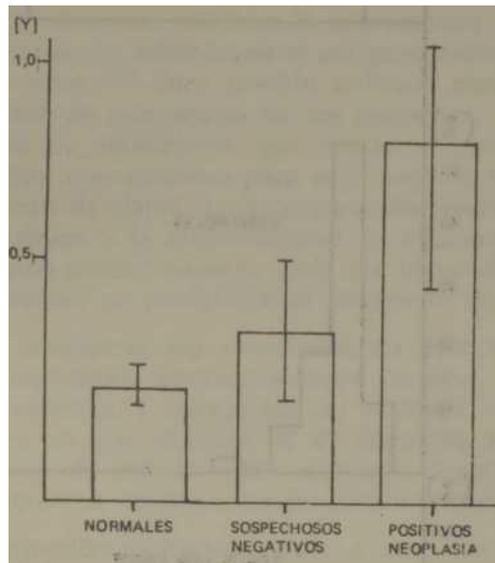
RESULTADOS DE LA APLICACION DE LA FUNCION DISCRIMINANTE CON DIFERENTES NIVELES DE DISCRIMINACION



El cuadro de la eficiencia del diagnóstico de los casos del grupo B, al aplicar la ecuación obtenida del grupo A: si se tomase como criterio de "positivo" tener valores de Y superiores a 0, i, a 0,2, y así sucesivamente (abscisa). Nótese cómo a medida que se sube el nivel arbitrario de discriminación disminuye el porcentaje de falso-positivos a expensas de un aumento de falsonegativos, y el porcentaje de diagnósticos correctos hace un máximo para el nivel de discriminación de 0,4.

**Gráfico 3**

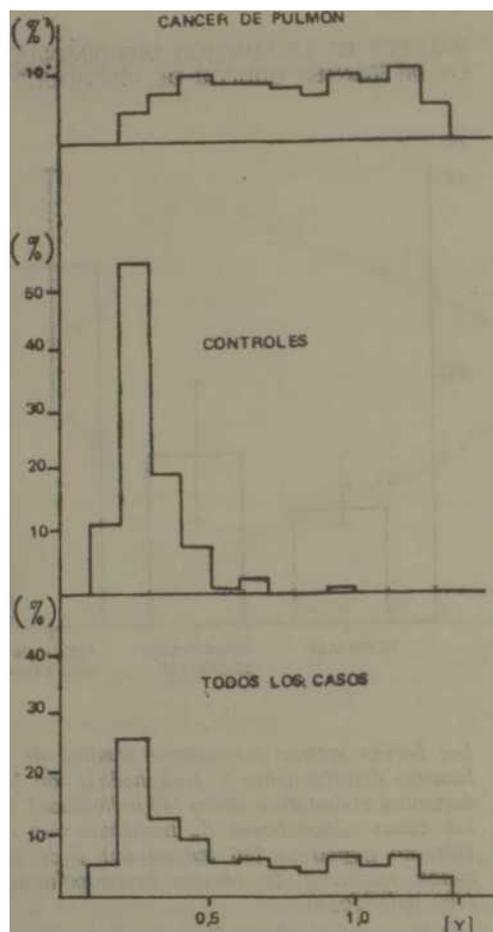
**VALORES DE LA FUNCION DISCRIMINANTE EN DIFERENTES GRUPOS DE INDIVIDUOS**



*Las barras indican los valores medios de la función discriminante Y (ordenadas), de los donantes voluntarios sanos (40 individuos), de los casos sospechosos de neoplasia que resultaron negativos (36 individuos) y de los casos positivos de cáncer broncopulmonar (102 individuos). Las líneas verticales indican las desviaciones estándares.*

Gráfico 4

DISTRIBUCION DE FRECUENCIAS DE LOS VALORES DE LA FUNCION DISCRIMINANTE Y



*El gráfico muestra las distribuciones en porcentajes (ordenadas de los diferentes valores de la función Y (abscisa) en el grupo de 102 pacientes de cáncer broncopulmonar en el grupo de 76 controles que no presentan esta enfermedad y en la muestra agrupado de 178 individuos. Nótese la distribución estrecha de los controles y la dispersión de los pacientes de cáncer broncopulmonar.*

## DISCUSION

En los últimos años ha aumentado el interés por los marcadores bioquímicos en los procesos neoplásicos. Por una parte, los métodos clásicos de diagnóstico han resultado con algunas excepciones, como cuello uterino y mama, incapaces de identificar pacientes oncológicos en etapas susceptibles de tratamiento curativo. Esto es especialmente cierto en el caso del cáncer broncopulmonar. Por otra parte, el desarrollo de métodos más eficientes, pero también riesgosos de tratamiento, hacen imprescindible disponer de mejores indicadores del pronóstico de un paciente, de su respuesta al tratamiento y de la evolución de la enfermedad, que posibiliten una individualización de la estrategia terapéutica.

Así, en los últimos 20 años han continuado apareciendo y evaluándose nuevos marcadores bioquímicos, tales como el antígeno carcinoembrionario, las alfafetoproteínas y otros.<sup>12,13</sup> Otro posible enfoque para aumentar la precisión de la evaluación de laboratorio de los pacientes, es el análisis de conjunto de pruebas de laboratorio que muestran cierta asociación con el proceso neoplásico, que emplean para este análisis técnicas multi- variadas de procesamiento de datos. La incorporación masiva de la computación a la gestión médica y la disponibilidad de minicomputadoras en hospitales, hace que estos procedimientos cada día dejen de ser un ejercicio teórico y se conviertan en posibilidades concretas de uso práctico.

En este trabajo se analizaron los resultados de determinaciones de velocidad de eritrosedimentación concentraciones de zinc, cobre, calcio, magnesio y proteína C reactiva, y actividades de catalasa, ceruloplasmina y deshidrogenasa láctica en una muestra de 40 donantes voluntarios, 36 sospechosos de neoplasia de pulmón que resultaron después no poseer la enfermedad y 102 pacientes confirmados de cáncer broncopulmonar.

En el proceso de regresión multilínea, paso a paso las variables que resultaron tener mayor asociación con la presencia de cáncer broncopulmonar fueron la eritrosedimentación, la actividad de deshidrogenasa láctica, la proteína C reactiva y la actividad de catalasa.

Con estas cuatro determinaciones es posible formar una función discriminante (cuadro IV) dada por:

$Y = 0,005062 \text{ (eritrosed)} + 0,000852 \text{ (LDH)} + 0,2395 \text{ (PCR)} + 0,80154 \text{ (catalasa)} - j- 0,049887$   
la cual si se consideran positivos los individuos con

Y mayor que 0,4 debe clasificar correctamente a los sujetos que presentan o no una neoplasia de pulmón con más del 80% de certeza. La posibilidad de que un sujeto presente una neoplasia de pulmón, aumenta con los valores de Y, y es especialmente elevada para valores superiores a 0,5. El límite de discriminación a emplear es obviamente arbitrario y depende del empleo que se le pretenda dar a una función de este tipo. Claramente, si el objeto del empleo de la función discriminante es establecer una sospecha más firme de neoplasia para investigar más profundamente a los sujetos catalogados por ella como "positivos", sería lógico fijar un límite de discriminación bajo y considerar positivas, por ejemplo, las cifras mayores de 0,3, y aceptar de entrada un gran número de

falsos positivos, pero garantizando perder un mínimo de pacientes por falsonegativos. Por el contrario, si la implicación de la función discriminante tuviera que ver con la instauración de determinada medida terapéutica, tendría sentido fijar un nivel de discriminación alto de modo de no tener falsos positivos. De cualquier forma, es el uso práctico de la función que se propone el que habrá de fijar su verdadero valor.

Resulta interesante notar, en las distribuciones de frecuencias, que si bien los individuos controles mostraron una distribución estrecha alrededor de cifras de 0,2-0,3, los pacientes de neoplasia mostraron una amplia distribución con cierta tendencia a la aparición de dos picos modales, alrededor de 0,4 y de 1,0, respectivamente. Pudieran avanzarse especulaciones sobre la posible significación de esta distribución de los valores de los pacientes, pero ello debe esperar a un análisis ulterior que permita establecer si entre los pacientes con neoplasia confirmada que tuvieron valores bajos y los que tuvieron valores elevados de la función Y, existe o no alguna diferencia sustancial en cuanto al curso clínico de la enfermedad. Este aspecto, y con él las posibles implicaciones de la función discriminantes, serán objeto de otro estudio.

Una limitación de estos resultados que debe destacarse, es la ausencia en la muestra de estadios iniciales del cáncer broncopulmonar (nódulos solitarios), lo cual impide extraer conclusiones sobre las posibilidades de diagnóstico temprano con la función descrita.

Los resultados, sin embargo, justifican la inclusión de la función discriminante hallada; en la práctica sistemática del diagnóstico de enfermedades del pulmón en el Instituto y su evaluación posterior con una serie más extensa que incluya una muestra importante de estadios I y II.

El cuadro V, de la opción de utilizar, además de la función discriminante recomendada según la "S" de Tukey, cuadro IV, tres funciones discriminantes más donde están incorporadas, sucesivamente, la ceruloplasmina, el calcio y el zinc. Se le concede la importancia a estos tres discriminantes, en el sentido de que a mayor número de determinaciones realizadas en un paciente, mayor es la posibilidad de seleccionar éste como portador de un cáncer broncopulmonar, entre otros pacientes con otras enfermedades, incluidas las de las vías respiratorias, o sea, hace más preciso el diagnóstico. Podría surgir el inconveniente de crear dificultades en usar más pruebas, pero en los discriminantes propuestos la determinación de la ceruloplasmina y calcio está al alcance de cualquier laboratorio, no así la determinación de zinc.

Resulta de interés el hecho de que la función discriminante obtenida en el trabajo sobre neoplasia malignas de mama<sup>2</sup>, que es:

$$Y = 0,00518 \text{ (eritrosed.)} + 0,09930 \text{ (catalasa)} + 0,00077 \text{ (cobre)} \\ - 0,26866 \text{ (calcio)} - 0,66844$$

y la obtenida en este trabajo:

$$Y = 0,005062 \text{ (eritrosed.)} - 0,000852 \text{ (LDH)} - 0,23950 \text{ (PCR)} + \\ 0,08015 \text{ (catalasa)} - 0,049887$$

muestran distintas pruebas de laboratorio clínico, indicándose cuán distinto es el comportamiento bioquímico de ambas enfermedades.

#### SUMMARY

Grueiro Azcano, E. et al. *Multivariate analysis of clinical laboratory data for the diagnosis of malignant neoplasias. II. Bronchopulmonary cancer.* Rev Cub Med 23: 2, 1984.

Techniques for multivariate analysis of data offer the possibility of evaluating simultaneous association of a set of laboratory tests with some determined diseases and to produce discriminating functions composed by several laboratory results that may have an association with disease at issue, being it superior to each one of the laboratory test, separately. Characteristics of the discriminating function consist in its ability to take into consideration the premonitory trend, since the value of the test, although being in the normal range, if it is moved toward the alteration sense, should influence on Y value (dependent variable). Laboratory studies were realized to more than 3 000 patients at the time of registration at the IOR. In this paper results corresponding to a group of 138 patients with or without bronchopulmonary cancer and 40 "healthy" individuals to whom the following determinations were performed: ceruloplasmin, erythrocytation, copper, zinc, calcium, magnesium, reactive protein C, lactate dehydrogenase, and catalase. The samples studied were divided at random into two groups, A and B, in order to obtain two separate discriminating functions, and after that, function obtained from group A data was evaluated and compared with function obtained from group B data, and conversely. With this study a discriminating function with results from erythrocytation was obtained, and serum determinations of lactate dehydrogenase, catalase, and reactive protein C, which gave a right diagnosis for 88% of the cases studied, when Y took a value over 0,4.

#### RÉSUMÉ

Grueiro Azcano, E. et al. *Analyse multivariée de données de laboratoire clinique dans le diagnostic de néoplasies malignes. II. Cancer bronchopulmonaire.* Rev Cub Med 23 : 2, 1984.

Les techniques d'analyse multivariée de données offrent la possibilité d'évaluer l'association simultanée d'un ensemble d'épreuves de laboratoire avec des maladies spécifiques, et de produire des fonctions discriminantes composées par plusieurs résultats de laboratoire, qui peuvent avoir une association avec la maladie en question. Ceci est supérieur à chacune des épreuves de laboratoire séparément. La caractéristique de la fonction discriminante consiste dans le fait qu'elle est capable de tenir compte de la tendance du phénomène, car la valeur de l'épreuve, quoiqu'elle soit dans les limites normales, si elle se déplace dans le sens de l'altération, elle aura une influence sur le résultat de Y (variable dépendante). Les études de laboratoire ont été faites sur plus de 3 000 patients au moment de leur inscription à l'Institut d'Oncologie et de Radiobiologie (IOR). Dans cet article, nous présentons les résultats correspondant à un groupe de 138 patients qui présentaient ou non cancer bronchopulmonaire et à 40 sujets "sains", chez lesquels on a déterminé: la céruloplasmine, la

sédimentation des érythrocytes, le cuivre, le zinc, le calcium, le magnésium, la protéine C réactive, la déshydrogénase lactique et la catalase. Les échantillons étudiés ont été distribués au hasard dans deux groupes, A et B, qui ont été employés pour obtenir deux fonctions discriminantes séparément, et ensuite la fonction obtenue avec les données du groupe A a été évaluée contre la fonction obtenue avec les données du groupe B, et vice versa. Avec cette étude on a obtenu une fonction discriminante avec les résultats de la sédimentation des érythrocytes et les déterminations sériques de la déshydrogénase lactique, de la catalase et de la protéine C réactive, qui a apporté un diagnostic correct dans 88% des cas étudiés, lorsque V prenait une valeur supérieure à 0,4.

#### BIBLIOGRAFIA

1. *Heberman, R. B.*: Overview on news markers for diagnosis of Cáncer 42: 1590-1600, 1978.
2. *Grueiro, S. Esquenazi; A. Lage*: Análisis multivariado de datos de laboratorio clínico en el diagnóstico de neoplasias malignas. I. Cáncer mamario. Enviado para su publicación en la Revista Cubana de Ginecología y Obstetricia.
3. *Martín, M. y colaboradores*: Laboratorio clínico. La Habana, Ed. Universitaria, 1966.
4. *Pye Unicam*: Manual de espectrofotometría de absorción atómica.
5. Técnica establecida por el MINSAP: Set distribuido por EPB Carlos J. Finlay, 1982.
6. *Henry, R. J.*: Química clínica, principios y técnicas. 2da. ed., T. 1. Barcelona. Jims, 1980. Pp. 705-706.
7. *Bergmeyer, H. U.; E. Bernz*: Methods of enzymatic analysis. Hans Ulrich Bergmeyer Ed Second ed. Vol. 2 Academic Press Inc. New York, 1969. Pp. 574-578.
8. *Thielmann, K.*: Principios de Metodología en Bioquímica. Editorial Organismos, Cuba, 1973.
9. *Armitage, P.*: Statistical Methods in Medical Research. Blackwill Scientific Publications, Oxford, 1971.
10. *Pascual, R.*: Centro Nacional de Investigaciones Científicas. Comunicación personal.
11. *Masteller, F.J. IV. Tukey*: Data analysis and regression. A second courses in statistics addison. Wesley Publishing Co., Massachusetts, 1977.
12. Proceedings of The First International Conference on the Clinical Uses of Carcinoembryonic Antigen. June 1-3, 1977, Lexington, Kentucky, Supplement to Cáncer. Vol. 42, No. 3 September, 1978.
13. *Javadpour, N.; K. R. Me Infere; T. A. Waldmann*: Human Chorionic Gonadotropin (HCG) and alpha-fetoprotein (AFP) in sera and tumor cells of patients with testicular seminoma. Cáncer 42: 2768-2772, 1978.

Recibido: 30 de noviembre de 1982.

Aprobado: 18 de abril de 1983.

Dr. *Emilio Grueiro*

Instituto de Oncología y Radiobiología

Calle 29 y E, Vedado

Ciudad de La Habana.