

Concordancia de lectura de test de ejercicio cardiopulmonar integrado utilizando un protocolo estándar en una población con antecedente de trauma de guerra

Concordance of integrated cardiopulmonary exercise testing interpretations using a standard protocol in a population with a background of war trauma

Alirio Rodrigo Bastidas⁽¹⁾; Gustavo Adolfo Hicapie⁽²⁾; Frank Boris Pernet⁽³⁾; Angélica Viviana Fletcher⁽⁴⁾; Luis Guillermo Bastidas⁽⁵⁾

RESUMEN

El test de ejercicio cardiopulmonar evalúa la tolerancia al ejercicio y provee una evaluación integral del sistema respiratorio, cardiovascular y muscular siendo útil en el diagnóstico diferencial de alteraciones cardiopulmonares específicas con desórdenes psicológicos o simulaciones, sin embargo, pueden existir diferencias en su interpretación.

Objetivo: evaluar el grado de concordancia interobservador para la interpretación del test de ejercicio cardiopulmonar utilizando el algoritmo diagnóstico de K. Wasserman en una población de pacientes jóvenes con heridas en combate.

Materiales y métodos: estudio de concordancia en la interpretación de 33 test de ejercicio cardiopulmonar por tres observadores y dos programas computarizados. El análisis se realizó primero entre los resultados de cada una de las observaciones con el diagnóstico final dado por consenso de los observadores y luego se realizó un análisis interobservador con los diferentes diagnósticos a los cuales se puede llegar mediante el test utilizando el coeficiente kappa, considerándose estadísticamente significativo una $p < 0,05$.

Resultados: las mejores concordancias entre observadores y el diagnóstico final fueron observadas bajo las categorías de normal y anormal, kappa con fuerza de concordancia de débil a buena (0,2725 - 0,6959 con $p < 0,05$). Cuando los diagnósticos son más específicos los valores kappa disminuyeron. La mejor concordancia entre los observadores se encontró cuando se comparan entre diagnósticos específicos, con coeficientes kappa más altos para el diagnóstico de enfermedades pulmonares.

Conclusión: el algoritmo diagnóstico utilizado para la interpretación de los test de ejercicio cardiopulmonar tiene bajos coeficientes de concordancia en general en esta población requiriendo para la misma un adecuado enfoque con historia clínica con algoritmos diagnósticos validados que pueden ser complementados con programas de computador específicos.

Palabras clave: test de ejercicio, estadística no paramétrica, heridas y lesiones.

ABSTRACT

Cardiopulmonary exercise testing assesses exercise tolerance and provides a comprehensive evaluation of the respiratory, cardiovascular, and muscular systems. It is useful in the differential diagnosis of specific cardiopulmonary alterations versus psychological disorders or malingering. However, different interpretations can arise.

Objective: to assess the degree of interobserver concordance in the interpretation of cardiopulmonary exercise testing by using K. Wasserman's algorithm in a population of young patients wounded in combat.

⁽¹⁾ Médico Internista, Neumólogo, Epidemiólogo Clínico. Clínica Universitaria Teletón, Profesor, Universidad de la Sabana. Bogotá, Colombia.

⁽²⁾ Neumólogo. Jefe del Servicio de Neumología, Hospital Militar Central, Profesor, Universidad Militar Nueva Granada. Bogotá, Colombia.

⁽³⁾ Médico Internista. Neumólogo. Hospital Militar Central. Bogotá, Colombia.

⁽⁴⁾ Médico epidemióloga. Universidad el Bosque. Bogotá, Colombia.

⁽⁵⁾ Abogado, Especialista en Derecho Médico, Candidato a Magíster, Pontificia Universidad Javeriana. Bogotá, Colombia.

Correspondencia: Doctor Alirio Rodrigo Bastidas. Correo electrónico: aliriorodrigo@yahoo.com.

Recibido: 15 de marzo de 2011. **Aceptado:** 22 de marzo de 2011.

Materials and methods: a study of concordance in the interpretation of 33 cardiopulmonary exercise tests by three observers and two computerized programs. In the first place, the analysis was done between the results of each observation with the final diagnosis given by a consensus of the observers. Secondly, an interobserver analysis was done with the different diagnoses that can be reached by way of the kappa coefficient. A $p < 0,05$ value was considered to be statistically significant.

Results: the best correlations between the observers and the final diagnosis were observed under the categories "normal" and "abnormal", kappa strength of agreement from weak to good (0,2725 - 0,6959 with $p < 0,05$). Kappa values diminished with more specific diagnoses. The best interobserver concordance was found when comparing specific diagnoses, with higher kappa coefficients for the diagnosis of pulmonary diseases.

Conclusion: the diagnostic algorithm used for interpreting cardiopulmonary exercise testing has low overall concordance coefficients for this population. As a consequence, this population requires adequate focusing by way of a clinical history with validated diagnostic algorithms that can be complemented with specific computer programs.

Key words: exercise testing, nonparametric statistics, wounds and injuries

Rev Colomb Neumol 2011; 23(1): 2-7.

MÉTODOS

Se utilizaron 33 test de ejercicio cardiopulmonar realizados en el servicio de neumología del Hospital Militar Central a un grupo de militares quienes consultaban por intolerancia al ejercicio de etiología no clara después de requerir manejo hospitalario por trauma de guerra.

Medición del test de ejercicio cardiopulmonar

La espirometría fue realizada de acuerdo a los parámetros establecidos por La Sociedad Americana de Tórax (ATS) (vmsx 2130 spiromethrer, EU). A todos los pacientes se les realizó cicloespirometría (sensormedics viatis v6200 autobox (dimax) vmsx 2130 spiromethrer) y valoración electrocardiográfica de doce derivaciones adaptadas al mismo equipo. Tras la colocación de la boquilla se determinaron los parámetros en reposo. Posteriormente, los sujetos realizaron un periodo de calentamiento de 2 a 4 minutos, seguido de trabajo en rampa con un incremento constante de 20 W/min hasta terminar la prueba por cansancio, dolor o disnea. Se estableció como tiempo mínimo de prueba un lapso de 6 minutos. Los gases arteriales fueron obtenidos de muestras de sangre en reposo y después del ejercicio, y fueron analizados inmediatamente. Las variables a medir, incluyendo el consumo de oxígeno máximo fueron obtenidas en el equipo de sensormedic de California USA en concordancia con las recomendaciones de la ATS.

Cálculo de los valores normales y definiciones

Las mediciones directas realizadas son el electrocardiograma, consumo de oxígeno máximo ($\text{VO}_2 \text{ max}$), umbral anaeróbico, frecuencia cardiaca, frecuencia respiratoria, presión arterial de oxígeno (PaO_2) y la presión arterial de dióxido de carbono (PaCO_2), las mediciones

indirectas calculadas son la relación del delta de consumo de oxígeno con el delta de trabajo ($\Delta\text{VO}_2/\Delta\text{WR}$), oxígeno latido ($\text{O}_2 \text{ latido}$), ventilación minuto ($\text{Ve} = \text{Vd} + \text{Va}$), reserva respiratoria = $\text{MVV} - \text{Ve}$, espacio muerto Vd/Vt , $\text{D(a-a)}\text{O}_2$, $\text{P(a-et)}\text{CO}_2$. Los valores de normalidad más importantes utilizados se muestran en la tabla 1.

Interpretación y lectura del test de ejercicio cardiopulmonar

Todos los test de ejercicio cardiopulmonar fueron leídos de manera individual por dos médicos neumólogos expertos, un médico neumólogo en formación y dos programas de computador diseñados para la interpretación del test; uno sigue los parámetros de diagnóstico utilizados en el libro Principles of Exercise Testing and Interpretation de K. Wasserman y el otro es un programa para la interpretación del test de ejercicio cardiopulmonar de la ATS/ACCP.

Se tomó como estándar de oro la interpretación en consenso de los tres evaluadores de cada test de ejercicio cardiopulmonar con el objeto de realizar una comparación con la interpretación individual de cada observador. La reunión de consenso se realizó con una diferencia de 2 meses a la primera evaluación individual en el Hospital Militar Central.

Las interpretaciones obtenidas fueron ingresadas a una base de datos en Excel. Se tomaron como resultados posibles todas las categorías descritas por Wasserman, a saber: normal, obesidad, pobre esfuerzo o desorden musculoesquelético, enfermedad pulmonar obstructiva, enfermedad pulmonar restrictiva, enfermedad vascular pulmonar, enfermedad pulmonar con alteración de la oxigenación periférica, isquemia miocárdica, falla ventricular izquierda, enfermedad arterial periférica y anemia. Se añadió una categoría adicional cuando se consideró que el test no era diagnóstico.

ESTADÍSTICA

Se realizó un primer análisis de correlación de la lectura de test de ejercicio cardiopulmonar entre cada observador y el estándar de oro, agrupándose los resultados inicialmente en normal y anormal, posteriormente en 5 categorías que incluían normal, obesidad, pobre esfuerzo, enfermedad cardiaca y enfermedad pulmonar y finalmente se realizó una última agrupación incluyendo todas las categorías resultantes de la lectura de cada observador.

El segundo análisis de correlación fue realizado comparando la lectura del test de ejercicio cardiopulmonar entre todos los observados, sin incluir la lectura del segundo programa de computador (validado por la ATS) ya que se considero no comparable, en el contexto del objetivo del estudio, puesto que era el único que seguía parámetros de lectura diferentes a los propuestos por Wasserman. Para este análisis se utilizaron las mismas 3 agrupaciones de las categorías descritas previamente, sólo que la última incluía todas las categorías diagnosticas resultantes de la lectura de los 33 test de ejercicio cardiopulmonar por parte de los 4 observadores.

Se utilizó el coeficiente Kappa de correlación para la valoración entre las variables; todas categóricas. El análisis se realizó con el programa Epi Dat 3.1, disponible de manera gratuita en la página de la Organización Mundial de la Salud. Se consideró significativo un valor de $p < 0,05$.

RESULTADOS

Correlación de la lectura del test de ejercicio cardiopulmonar entre observador y estándar de oro

Comparación entre los resultados de normalidad y anormalidad de cada observador contra el estándar de oro.

Comparación de todas las categorías resultantes de la lectura de cada observador contra el estándar de oro incluyendo: normal, obesidad, pobre esfuerzo o desorden musculoesquelético, enfermedad pulmonar obstructiva, enfermedad pulmonar restrictiva, enfermedad vascular pulmonar, enfermedad pulmonar con alteración de la oxigenación periférica, falla ventricular izquierda, enfermedad arterial periférica y test no diagnóstico.

Comparación entre los resultados de 5 categorías de cada observador contra el estándar de oro, se incluyó: normal, obesidad, pobre esfuerzo o desorden musculoesquelético, enfermedad pulmonar y enfermedad cardiovascular.

Al realizar la prueba de concordancia con el estándar de oro, las correlaciones más altas fueron observadas cuando se utilizaron para el diagnóstico las categorías de normal y anormal, encontrándose los valores de kappa con una fuerza de concordancia de débil a buena (entre 0,2725 y 0,6959). Los valores p fueron significativos ($p < 0,05$) excepto en un observador, aunque es de anotar que todos los intervalos de confianza fueron amplios.

Tabla 1. Valores de normalidad.

Variable	Valor
VO2 (L/min.)	< 84% del predicho se considero anormal.
Umbral anaerobio (AU)	Punto de quiebre gráfica de consumo de oxígeno y producción CO ₂ valor normal > 60%
Reserva de la frecuencia cardiaca	FC pico del predicho - FC observada, mayor de 15 L/min como normal.
Reserva ventilatoria	Valoración gráfica espirometria y/o valor mayor de 11 L/min como normal.
Máximo de la frecuencia respiratoria	55 respiraciones por minuto
O ₂ latido	< 80% fue considerado como anormal
Vd/Vt ejercicio máximo	Menor de 0,3
D(A-a)O ₂	Menor de 20 mmHg como normal
P(a-et)CO ₂	Menor de 5 mmHg como normal.
FEV1 y CVF	Según los valores publicados por la American Thoracic Society
VO ₂ /WR	Valor normal > 10 ml/min/W

Tabla 2. Correlación observador individual con estándar de oro entre las categorías de normal y anormal.

	Kappa	Estándar de oro IC 95%	Valor p
Observador 1	0,3709	0,1086 - 0,6332	0,013
Observador 2	0,2725	-0,0603 - 0,6054	0,1102
Observador 3	0,56	0,286 - 0,834	0,0007
Programa 1	0,4261	0,0918 - 0,7604	0,0144
Programa 2 (ATS)	0,6959	0,422 - 0,9697	0,0001

Tabla 3. Correlación observador Individual con estándar de oro entre todas las categorías resultantes

	Kappa	Estándar de oro IC 95%	Valor p
Observador 1	0,349	0,1504 - 0,5476	0
Observador 2	0,2168	0,0118 - 0,4217	0,0104
Observador 3	0,2048	0,0434 - 0,3662	0,0036
Programa 1	0,2486	0,0082 - 0,489	0,0045
Programa 2 (ATS)	0,2687	-0,0012 - 0,5386	0,0015

Cuando se realizó la comparación con todas las categorías resultantes de la lectura de cada observador o agrupándolas en 5 categorías, todos los valores de kappa disminuyeron. Llama la atención que el segundo programa de computador validado por la ATS tiene una buena concordancia para diferenciar lo normal de lo anormal, sin embargo, pierde su concordancia al ser utilizado para diagnósticos más específicos.

Correlación interobservador de la lectura del test de ejercicio cardiopulmonar

Al estudiar la correlación interobservador entre normalidad y anormalidad se encontró un kappa de 0,2134 con un IC 95% entre 0,0165 y 0,4096 con una p de 0,0027.

Hubo un total de 9 categorías diagnósticas resultantes de la lectura de los 33 test de ejercicio cardiopulmonar con las que se realizó una correlación interobservador incluyendo: normal, obesidad, pobre esfuerzo o desorden musculoesquelético, enfermedad pulmonar obstructiva, enfermedad pulmonar restrictiva, enfermedad vascular pulmonar, enfermedad pulmonar con alteración de la oxigenación periférica, falla ventricular izquierda, y test no diagnóstico. La categoría de enfermedad arterial periférica no se encuentra incluida ya que sólo el estándar de oro realizó este diagnóstico, por lo que fue incluido en el primer análisis. Los resultados de la concordancia diagnóstica entre los observadores se muestran en la tabla 5.

Al agrupar los resultados en 5 categorías que incluyen normal, pobre esfuerzo o desorden musculoesquelético, obesidad, enfermedad cardiovascular o enfermedad pulmonar se obtuvieron los resultados dados en la tabla 6.

Al realizar la prueba de concordancia interobservador utilizando las categorías de normal y anormal se encuentra una concordancia débil aunque con un intervalo de confianza amplio que indica que el valor verdadero oscila entre una fuerza de concordancia de pobre a débil. Cuando se comparan todas las categorías, se encuentra una correlación de débil a moderada en alteraciones pulmonares tipo enfermedad pulmonar con alteración de la oxigenación periférica, enfermedad obstructiva y restrictiva pulmonar, mientras que todas las otras categorías muestran una correlación pobre, incluso algunas no muestran correlación. Cuando se realizó la comparación con menor número de categorías no existieron diferencias significativas.

DISCUSIÓN

La lectura del test de ejercicio cardiopulmonar al utilizar los parámetros establecidos en los algoritmos diagnósticos de Wasserman varía tanto para la certeza diagnóstica cuando se compara con un estándar de oro, así como cuando se compara su lectura entre varios observadores. Esta situación es similar a la observada con otras pruebas diagnósticas como las radiografías, citologías, electrocardiogramas, espirometrías, etc [6-8].

Tabla 4. Correlación Observador Individual con estándar de oro con diagnósticos agrupados en 5 categorías

	Kappa	Estándar de oro IC 95%	Valor p
Observador 1	0,338	0,1425 - 0,5334	0,0001
Observador 2	0,2287	0,0054 - 0,4521	0,0378
Observador 3	0,2326	0,0565 - 0,4086	0,0037
Programa 1	0,234	-0,009 - 0,4771	0,0198
Programa 2 (ATS)	0,3456	0,0151 - 0,0016	0,0016

Tabla 5. Correlación interobservador con todas las categorías resultantes

Categoría	Kappa	IC 95%	Valor p
Normal	0,2075	0,0064 - 0,4079	0,0035
Pobre esfuerzo	0,1598	0,0719 - 0,2473	0,0246
Obesidad	0,2016	0,0643 - 0,3382	0,0046
Falla VI	-0,0233	- 0,0502 - 0,0037	0,7435
EPOP	0,4833	- 0,3418 - 1,3	0
EPR	0,2457	- 0,3558 - 0,8414	0,0005
EVP	-0,0233	- 0,0502 - 0,0037	0,7435
EOP	0,4844	- 0,5667 - 1,5152	0
Test no diagnóstico	-0,0313	- 0,0623 - - 0,0002	0,6601
Kappa global	-0,0233	0,0732 - 0,3538	0

EPOP: enfermedad pulmonar con alteración oxigenación periférica, EPR: enfermedad restrictiva pulmonar, EVP: enfermedad vascular pulmonar, EOP: enfermedad obstructiva pulmonar.

Las principales mediciones del test de ejercicio cardiopulmonar se basan en medidas cuantitativas directas de parámetros fisiológicos, los cuales pueden variar dependiendo de la calibración del equipo e individuos evaluados, lo que podría corresponder en cierto momento a un sesgo de medición (9, 10). Así mismo se debe tener en cuenta que los test de ejercicio cardiopulmonar interpretados fueron realizados en Bogotá; a 2640 msnm, hecho que puede alterar los parámetros fisiológicos obtenidos en el test y tener un impacto sobre el diagnóstico, ya que para la lectura se siguieron los valores estrictos dados en el libro, que no contemplan datos de normalidad en la altura (5).

Las consideraciones descritas son de vital importancia, sin embargo, no es lo anterior una explicación a las variaciones de la interpretación de los resultados, ya que lo que se evalúa en este caso es la correlación al interpretar un test de ejercicio cardiopulmonar ya realizado con base en un algoritmo diagnóstico. Si bien, la concordancia se ve afectada y está directamente relacionada con el conocimiento particular de cada uno de los

observadores, no es este el único factor a considerar cuando las correlaciones no son las esperables teniendo en cuenta que esta prueba diagnóstica puede catalogarse como relativamente objetiva (11, 12).

El análisis entre agrupar más o menos categorías muestra una mayor correlación al distinguir entre lo anormal de lo normal, sin embargo, el número de categorías adicionales posterior casi no influye en el resultado de correlación. A pesar de lo anterior, la correlación entre el diagnóstico de normalidad es baja para lo esperable de un diagnóstico que teóricamente es de fácil interpretación (5). Vale la pena destacar que de todas las correlaciones interobservador, los mejores coeficientes kappa se presentaron en los diagnósticos de enfermedades pulmonares específicas, hecho relacionado con la interpretación por observadores especialistas en el área de neumología.

Con los resultados obtenidos al evaluar un algoritmo específico, se identifica la necesidad de plantear otra forma de interpretación de los test de ejercicio

Tabla 6. Correlación interobservador agrupándose los diagnósticos en 5 categorías

Categoría	Kappa	IC 95%	Valor p
Normal	0,2075	0,0064 - 0,4079	0,0035
Pobre esfuerzo	0,1598	0,0719 - 0,2473	0,0246
Obesidad	0,2016	0,0643 - 0,3382	0,0046
Enfermedad cardiovascular	-0,0394	- 0,0741 - 0,0047	0,5796
Enfermedad pulmonar	0,3585	0,0258 - 0,6896	0
Kappa global	-0,0394	0,07 - 0,3702	0

cardiopulmonar en búsqueda de una mejor concordancia. En el momento se encuentran disponibles otros programas validados para la interpretación de esta prueba diagnóstica (13) que tienen un buen rendimiento, sin embargo, van a estar siempre subyugados a la evaluación clínica, ya que como se observa en los resultados, cuando se realizan diagnósticos más específicos se pierde la concordancia.

Parece ser que entre más complejo el algoritmo de interpretación del test de ejercicio cardiopulmonar, se presenta una mayor variabilidad en el resultado diagnóstico final, por lo que es necesario evaluar la posibilidad de desarrollar mecanismos diagnósticos más simples que ayuden a optimizar no sólo la concordancia sino también la certeza diagnóstica (14).

La principal debilidad del estudio es el poco tamaño de muestra y el grupo seleccionado de militares poco representativo de la población general, no obstante, debe tenerse en cuenta que el test de ejercicio cardiopulmonar se encuentra indicado precisamente para situaciones de difícil diagnóstico.

CONCLUSIÓN

El algoritmo diagnóstico utilizado en este caso para la interpretación de los test de ejercicio cardiopulmonar tiene un coeficiente de correlación pobre con todos los diagnósticos posibles y aumenta ligeramente cuando se utilizan menos categorías, mostrando la mayor correlación entre las categorías diagnósticas de normal y anormal. Sin embargo, considerando que el test de ejercicio cardiopulmonar se encuentra indicado para diagnósticos difíciles, es necesario considerar otros algoritmos más sencillos y de mejor rendimiento, que deben ser validados en cada servicio de neumología, y siempre interpretados a la luz de la historia clínica como en cualquier prueba diagnóstica.

REFERENCIAS

1. Palange P, Ward SA, Carlsen KH, Casaburi R, Gallagher CG, Gosselink R, et al. Recommendations on the use of exercise testing in clinical practice. *Eur Respir J* 2007; 29 (1): 185-209.
2. Ross RM. ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med* 2003; 167 (10): 1451.
3. Eschenbacher WL, Mannina A. An algorithm for the interpretation of cardiopulmonary exercise tests. *Chest* 1990; 97 (2): 263-7.
4. Schmid A, Schilter D, Fengels I, Chhajed PN, Strobel W, Tamm M, et al. Design and validation of an interpretative strategy for cardiopulmonary exercise tests. *Respirology* 2007; 12 (6): 916-23.
5. Wasserman K, Hansen J, Sue D, Stringer W, Whipp B. Principles of Exercise Testing and Interpretation. Fourth Edition, 2005. p. 1-235. (GENERIC) Ref Type: Generic
6. Leung SK, Yew WW, Leung CC. Discordance in spirometric interpretations using different reference equations. *Chest* 2009; 136 (1): 324-5.
7. Hong Y, Ra SW, Shim TS, Lim CM, Koh Y, Lee SD, et al. Poor interpretation of pulmonary function tests in patients with concomitant decreases in FEV1 and FVC. *Respirology* 2008; 13 (4): 569-74.
8. Aizer A, Holmes DS, Garlitski AC, Bernstein NE, Smyth-Melsky JM, Ferrick AM, et al. Standardization and validation of an automated algorithm to identify fractionation as a guide for atrial fibrillation ablation. *Heart Rhythm* 2008; 5 (8): 1134-141.
9. Ross RM. ATS/ACCP statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med* 2003; 167 (10): 1451.
10. Brawner CA, Ehrman JK, Aldred H, Schairer JR, Keteyian SJ. Quality assurance and cardiopulmonary exercise testing in clinical trials. *J Card Fail* 2008; 14 (4): 283-9.
11. Schmid A, Schilter D, Fengels I, Chhajed PN, Strobel W, Tamm M, et al. Design and validation of an interpretative strategy for cardiopulmonary exercise tests. *Respirology* 2007; 12 (6): 916-23.
12. Sridhar MK, Carter R, Banham SW, Moran F. An evaluation of integrated cardiopulmonary exercise testing in a pulmonary function laboratory. *Scott Med J* 1995; 40 (4): 113-6.
13. Ross RM, Corry DB. Software for interpreting cardiopulmonary exercise tests. *BMC Pulm Med* 2007; 7: 15.
14. Schmid A, Schilter D, Fengels I, Chhajed PN, Strobel W, Tamm M, et al. Design and validation of an interpretative strategy for cardiopulmonary exercise tests. *Respirology* 2007; 12 (6): 916-23.