

## Documentos

# GUÍA PARA EL ANÁLISIS CRÍTICO DE ESTUDIOS QUE EVALÚAN EXÁMENES DIAGNÓSTICOS\*

Claudio Vera P-G.<sup>1,3</sup>, Luz María Letelier S.<sup>2,3</sup>, Jorge Carvajal C.<sup>1</sup> Ph.D.

<sup>1</sup>Departamento de Obstetricia y Ginecología, <sup>2</sup>Departamento de Medicina Interna, <sup>3</sup>Unidad de Medicina Basada en Evidencia. Facultad de Medicina, Pontificia Universidad Católica de Chile.

## RESUMEN

El análisis crítico de artículos sobre exámenes diagnósticos permite reconocer su utilidad clínica analizando su validez, resultados y aplicabilidad. En este documento se presenta la metodología para lograr ese objetivo.

**PALABRAS CLAVE:** *Análisis crítico, prueba diagnóstica, medicina basada en evidencia*

## SUMMARY

The critical analysis of articles on diagnostic test allows to recognize their clinical utility analyzing their validity, results and applicability. This document present the methodology to obtain the objective.

**KEY WORDS:** *Critical analysis, diagnostic test, medicine based on evidence*

## INTRODUCCIÓN

Frente a un paciente, portador de determinados síntomas y signos, debemos efectuar el diagnóstico de la enfermedad o proceso patológico subyacente. El proceso diagnóstico requiere tanto de razonamiento lógico como de habilidad para el reconocimiento de patrones. Un correcto y eficiente diagnóstico nos hará más eficaces como clínicos para prevenir, curar y focalizar el tratamiento de las enfermedades en nuestros pacientes. Por otro lado, procesos diagnósticos ineficientes o incorrectos tienen el riesgo de someter a los pacientes a estudios diagnósticos y/o tratamientos innecesarios con los riesgos y costos asociados. El proceso diagnóstico puede ser analizado como un ciclo de dos etapas (1, 2).

**Etapas 1:** Diagnóstico diferencial. A partir de una situación clínica el médico infiere los posibles diagnósticos, utilizando alguna(s) de las siguientes aproximaciones para organizar el diagnóstico diferencial.

a) Enumerar todas las posibilidades, según órganos o sistemas, para descartarlas en forma simultánea con una serie de exámenes; corresponde a una forma *posibilística* de aproximación.

b) Ordenar estas posibilidades diagnósticas de acuerdo a su probabilidad relativa; corresponden a una forma *probabilística* de aproximación.

c) Estudiar las posibilidades diagnósticas según su pronóstico (descartar o confirmar primero la posibilidad de peor pronóstico).

d) Evaluar las posibilidades diagnósticas se-

\* Trabajo presentado por el Dr. Claudio Vera en sesión ordinaria de la Sociedad Chilena de Obstetricia y Ginecología del 3 de mayo de 2005.

gún un criterio pragmático (diagnosticar las que respondan mejor al tratamiento).

Seleccionar la mejor posibilidad diagnóstica para el paciente implica utilizar juiciosamente las aproximaciones probabilísticas, pronósticas y pragmáticas para el diagnóstico diferencial.

**Etapas 2:** Incorporación de nueva información para reordenar las probabilidades, confirmar o descartar un diagnóstico. Para cada posibilidad diagnóstica la nueva información o estudio diagnóstico modificará su probabilidad relativa hasta llegar a un punto que permita descartarla o finalmente elegirla como la más probable. Esta nueva información puede ser: síntomas, hallazgos del examen físico, exámenes de laboratorio, estudio de imágenes o evolución clínica.

En el curso del proceso diagnóstico constantemente movemos las probabilidades relativas de diferentes diagnósticos entre dos umbrales de decisión: el umbral diagnóstico y el umbral terapéutico. Cruzamos el umbral terapéutico cuando la información disponible nos sugiere una probabilidad de diagnóstico tan alta que estamos dispuestos a iniciar tratamiento. Cruzamos el umbral de diagnóstico cuando la información disponible sugiere que determinado diagnóstico es tan improbable que descartamos esa posibilidad y no necesitamos más información. Cuando la información es insuficiente para traspasar uno de los dos umbrales, entonces se requiere de nueva información. Como se muestra en la Figura 1, rara vez los umbrales de decisión llegarán a la certeza absoluta.

Este proceso lo realizamos habitualmente de modo intuitivo, utilizando la información de los exámenes diagnósticos para mover la probabilidad (*pre-test*) de un determinado diagnóstico a una nueva probabilidad (*post-test*), que nos acerque a alguno de los dos umbrales, el proceso se completa al superar el umbral diagnóstico o terapéutico.

Las *propiedades* de un examen diagnóstico o "test" reflejan su capacidad de llevarnos a una nueva probabilidad *post-test*. Por lo tanto, conocer las propiedades de un examen diagnóstico, previo a solicitarlo, es esencial para decidir su realización, especialmente si se asocia a costos o riesgos, tanto para la madre como para el feto.

En este artículo mostramos una guía para el análisis de artículos que reporten estudios de las propiedades de un examen diagnóstico, permitiéndonos una mejor decisión respecto de su uso en una situación clínica determinada.

GUÍA PARA EL ANÁLISIS

El análisis crítico de artículos sobre exámenes diagnósticos permite reconocer su utilidad clínica analizando su validez (descartar posibilidad de sesgo o error sistemático en su metodología), resultados y aplicabilidad. Se han diseñado guías que permiten realizar esta evaluación, el siguiente ejemplo esta basado en una de ellas (3).

Escenario clínico

Paciente de 32 años, primigesta, cursando un embarazo de 33 semanas, a quien se efectúa el diagnóstico de síndrome hipertensivo del embarazo. Se decide su hospitalización para evaluar la severidad del cuadro hipertensivo y evaluación de la unidad fetoplacentaria. En el examen físico del ingreso destaca altura uterina en percentil menor a 10 para la edad gestacional, formulándose el diagnóstico de sospecha de restricción de crecimiento fetal.

*¿Cual será la capacidad de la ultrasonografía para detectar restricción de crecimiento fetal (RCF) en una paciente con síndrome hipertensivo del embarazo?*

La búsqueda de información relevante en PubMed (<http://www.pubmed.gov>) utilizando los



Figura 1. Umbrales para realizar un examen diagnóstico de acuerdo a la probabilidad del diagnóstico clínico.

términos: "pregnancy toxemias", "ultrasonography" y "fetal growth retardation", aporta varios estudios atingentes de los cuales elegimos uno para analizar (4).

Sugerimos efectuar este análisis dando respuesta a tres preguntas fundamentales: 1. ¿Son válidos los resultados?; 2. ¿Cuáles son los resultados?; y 3. ¿Cómo aplicar los resultados al cuidado del paciente?

### **Paso 1. ¿Son válidos los resultados?**

#### *1. ¿Los clínicos enfrentaron incertidumbre diagnóstica?*

Un examen diagnóstico es útil si es capaz de diferenciar trastornos o condiciones que pueden ser confundidos entre sí. La capacidad de un examen diagnóstico se evalúa realmente cuando es expuesto a condiciones clínicas similares a la realidad.

La selección de pacientes con cuadros muy sugerentes de la enfermedad a diagnosticar (en los cuales prácticamente no se necesitaría estudio diagnóstico) comparados con voluntarios sanos, entrega poca información sobre la utilidad de un estudio diagnóstico y puede sobreestimar la capacidad diagnóstica de un examen. Es por esta razón que la población elegida para la evaluación de un examen diagnóstico debiera incluir a pacientes con todo el espectro de presentación de la enfermedad o condición que se pretende diagnosticar y pacientes con otras condiciones que habitualmente se confundan con ella.

En este ejemplo, el artículo elegido para análisis (4) corresponde a un estudio caso control donde se seleccionaron 287 casos con embarazos únicos, de vértice, sin anomalías, con edad gestacional corroborada por ultrasonografía antes de las 24 semanas y síndrome hipertensivo del embarazo. Los controles fueron pareados (1:1) por edad gestacional con pacientes sin síndrome hipertensivo, sin otras restricciones. Las mediciones realizadas fueron diámetro biparietal, longitud femoral, perímetros craneano y abdominal, en ecografías efectuadas dentro de 2 semanas previas al parto. En el artículo se evaluó la capacidad diagnóstica de la ultrasonografía para RCF en casos y controles.

#### *2. ¿Existió comparación ciega con un estándar de referencia independiente?*

Los resultados del examen en evaluación deben ser comparados de forma ciega e independiente

con un estándar de referencia. Ciego implica que los resultados de un examen (el evaluado y el estándar) no son conocidos por quien aplica el otro examen. Independiente significa que los resultados de un examen no interfieren con los resultados del estándar o viceversa.

El estándar de referencia es un estimador de la "verdad", y por lo tanto debe ser apropiado para la condición estudiada. En el artículo evaluado, la estimación de peso ultrasonográfica (examen en evaluación) se compara contra el peso del recién nacido dentro de 2 semanas posterior al examen, estándar que consideramos apropiado. Si la estimación ecográfica hubiese sido previa al nacimiento; la evaluación de la capacidad del examen habría sido comparada con la "verdad", condición ideal para la evaluación de un examen diagnóstico. Si el estándar de referencia hubiese sido el peso del recién nacido 1 mes después de la estimación ultrasonográfica, probablemente no lo consideraríamos un adecuado estándar de referencia. Cuando esto sucede es improbable que el resultado del estudio sea válido para la pregunta clínica.

Si el estándar de referencia es aceptado entonces debemos evaluar si fue aplicado en forma ciega e independiente. Respecto del ciego, el estudio no aporta datos suficientes para evaluar, pues no se describe si los que asignaron el peso del recién nacido conocían los resultados de la ultrasonografía, y tampoco se indica si los que efectuaban la ultrasonografía llegaban a conocer el peso real de los recién nacidos. Es posible, por ejemplo, que el conocimiento de los pesos de los recién nacidos sucesivos por parte de los que realizaban la estimación de peso ecográfica influya sobre las estimaciones futuras, constituyendo una fuente de sesgo. En términos generales, mientras más subjetiva e influenciada sea la interpretación de un nuevo examen por el conocimiento del estándar de referencia, o vice versa, mayor es la importancia de que se mantenga ciega la persona que interprete los exámenes.

Respecto de la aplicación independiente de los exámenes esto si se cumplió, pues la estimación de peso ecográfica no es capaz de alterar la medición del peso al nacimiento.

#### *3. ¿Los resultados del examen diagnóstico evaluado influyeron la realización del estándar de referencia?*

La evaluación de las propiedades de un examen diagnóstico puede ser distorsionada si es que el resultado del examen influye la decisión de

efectuar el estándar de referencia. El estudio evaluado en este ejemplo si cumple con este criterio, pues todos los fetos fueron evaluados con el examen en estudio y el estándar. Habría sido un error estudiar el peso del recién nacido sólo en aquellos casos en que la ecografía sugiriera el diagnóstico de restricción de crecimiento. De presentarse, esta situación se conoce como sesgo de verificación.

**Paso 2. ¿Cuáles son los resultados?**

En el proceso diagnóstico, antes de realizar un examen, debemos conocer la probabilidad pre-test de presentar la enfermedad. Para una embarazada de 3er trimestre sana con altura uterina en percentil 50 no es razonable solicitar una ultrasonografía buscando RCF, dado su baja probabilidad pre-test. Intuitivamente, la probabilidad pre-test cambia si se agrega la presencia de síndrome hipertensivo del embarazo y altura uterina en p10; sin embargo, aun así la probabilidad de RCF no es tan alta como para cruzar el umbral terapéutico, y se requiere de un examen adicional. Si la paciente presenta síndrome hipertensivo del embarazo y la altura uterina está en p < 2 las dudas sobre un RCF son pocas y es posible iniciar un tratamiento, en este caso, decidir la interrupción del embarazo.

En el proceso diagnóstico los exámenes tienen utilidad en la medida que son capaces de cambiar la probabilidad pre-test a una nueva probabilidad post-test significativamente distinta para la toma de decisiones clínicas. El análisis de los resultados de un estudio para evaluar un examen diagnóstico está dirigido a evaluar en que magnitud los resultados del examen diagnóstico cambian la probabilidad pre test.

En el estudio elegido los resultados de la evaluación ultrasonográfica en pacientes con síndrome hipertensivo del embarazo, son presentados en la Tabla I, una tabla de doble entrada, la cual permite fácilmente calcular la sensibilidad, especificidad, valor predictivo positivo y negativo del examen diagnóstico. Sin embargo, estos datos, no

**Tabla I  
RESULTADOS DEL EXAMEN DIAGNÓSTICO  
(DATOS ESTIMADOS CON LOS RESULTADOS  
DEL ESTUDIO) (4)**

		Peso del RN			
		PEG		No PEG	
Ultrasonografía	RCF Si	4	a	4	b
	RCF No	39	c	240	d
Total		43		244	

Sensibilidad:  $a/(a+c) = 9\%$   
 Especificidad:  $d/(b+d) = 98\%$   
 Valor predictivo positivo:  $a/(a+b) = 50\%$   
 Valor predictivo negativo:  $d/(c+d) = 86\%$

PEG: recién nacido pequeño para la edad gestacional (menor a p10)

RCF: restricción de crecimiento fetal

nos permiten conocer en cuanto cambia la probabilidad pre-test. Para conocer este dato, debemos calcular el Likelihood Ratio (Tabla II).

La capacidad de un test para cambiar la probabilidad pre test a una nueva probabilidad post test se llama Likelihood Ratio (LR) y nos informa sobre la dirección y magnitud de este cambio. El LR tiene alta relevancia para la práctica clínica.

¿Cuán frecuentemente las pacientes con RCF en ultrasonografía tienen un recién nacido PEG?, ¿Cuán frecuentemente las pacientes sin recién nacidos PEG tienen RCF en ultrasonografía?. La razón de estas dos probabilidades constituye el LR. Tal como se muestra en la Tabla II, hemos calculado que si la ultrasonografía sugiere una RCF el LR es 5,7, y si no sugiere RCF, el LR es 0,9.

La consideración de qué tan relevante es la magnitud de un LR, depende de la probabilidad pretest. Sin embargo, como regla general se puede decir que:

– LR > de 10 o <0,1 son capaces de generar cambios importantes y a menudo concluyentes entre la probabilidad pre y la post test.

**Tabla II**

**CÁLCULO DE LIKELIHOOD RATIO (DATOS ESTIMADOS CON LOS RESULTADOS DEL ESTUDIO) (4)**

Likelihood ratio positivo (LR+)		Likelihood ratio negativo (LR-)	
$\frac{\text{sensibilidad}}{1 - \text{especificidad}}$	$= \frac{a/(a+c)}{b/(b+d)} = 5,7$	$\frac{1 - \text{sensibilidad}}{\text{especificidad}}$	$= \frac{c/(a+c)}{d/(b+d)} = 0,9$

– LR 5-10 y 0,1-0,2 generan cambios moderados entre las probabilidades pre y post test.

– LR 2-5 y 0,5-0,2 generan cambios pequeños, pocas veces importantes en la probabilidad post test.

– LR 1-2 y 0,5-1 generan muy poco cambio de la probabilidad y rara vez relevante.

### **Paso 3. ¿Cómo aplicar los resultados al cuidado del paciente?**

#### *1. ¿Son reproducibles los resultados e interpretaciones del test diagnóstico en mi medio?*

Un examen es útil si es reproducible en la población en la cual debe ser aplicado. Los factores que pueden afectar esta reproducibilidad dependen del examen (Ej. resolución de la ultrasonografía) o de su realización e interpretación (Ej. medición correcta de la biometría fetal, operador dependiente). Un artículo que evalúe un examen diagnóstico debiera idealmente referirse a la reproducibilidad del examen con alguna medida que evalúe la concordancia frente al azar. Esto es relevante cuando se evalúa un examen que requiere experiencia para su realización e interpretación. Si la reproducibilidad del examen es baja y la discordancia en las interpretaciones es frecuente y aún el examen es capaz de discriminar los afectados por la condición en estudio, el examen es muy útil. Si la reproducibilidad es alta y la variabilidad interobservador es muy baja, el examen puede ser simple y poco ambiguo, de modo que el desempeño de observadores con menos experiencia afecta poco la capacidad del examen. En el estudio analizado no hay datos suficientes para poder evaluar, pues las ecografías fueron realizadas por residentes, becados de medicina materno fetal y sonografistas, y no se entrega alguna evaluación de su concordancia o variabilidad interoperador.

#### *2. ¿Cómo aplico los resultados en mi paciente?*

Las propiedades de un examen pueden variar con diferentes poblaciones de pacientes (efecto del espectro de pacientes) (5). Si el examen se evalúa en un espectro de enfermos de mayor gravedad harán que el LR del examen se mueva lejos de 1 (aumenta la sensibilidad). En grupos de pacientes con enfermedad menos severa los LR se acercarán a 1 (menor sensibilidad). Por ello, el primer paso para aplicar los resultados del estudio a nuestros pacientes, es evaluar si el estudio de

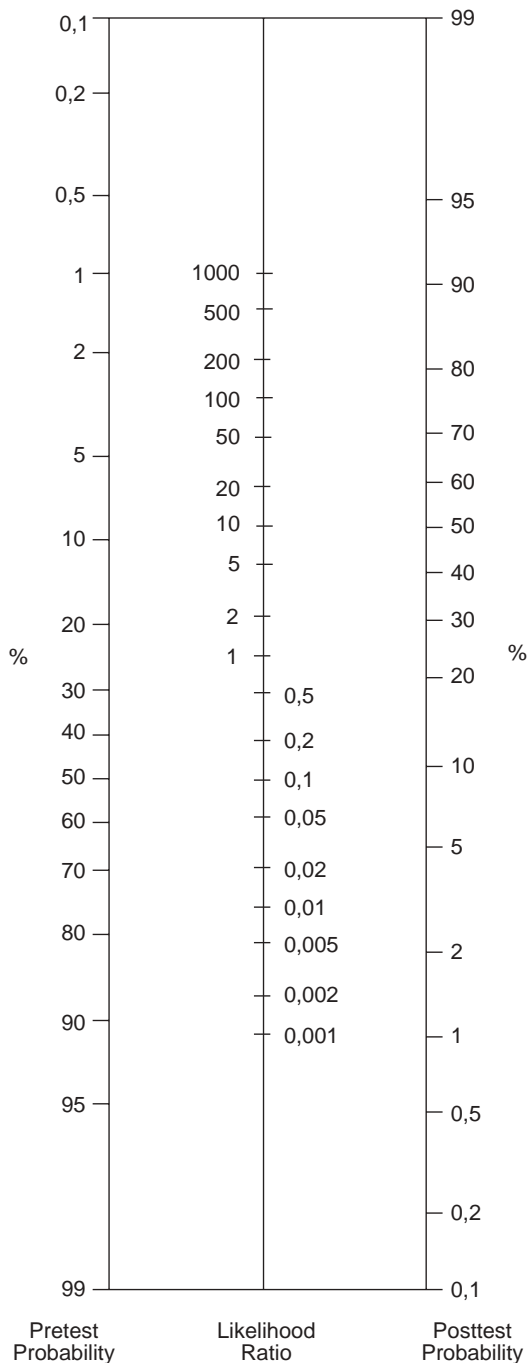
las características del examen diagnóstico fue efectuado en el espectro adecuado de pacientes. En este ejemplo esto si se cumple, pues el estudio se efectuó en pacientes embarazadas con diagnóstico de síndrome hipertensivo del embarazo. También se efectuó en una cohorte de pacientes sin hipertensión, pero este dato no es útil para nuestro escenario clínico.

El nomograma de Fagan (Figura 2) nos ayudará a aplicar los LR en el contexto de los pacientes individuales (6). En la columna izquierda se representa la probabilidad pretest, la columna intermedia representa el LR y la columna de la derecha la probabilidad post test. Una vez confirmado que el paciente se asemeja a los pacientes incluidos en el estudio y determinado clínicamente su probabilidad pre test, debemos usar el LR calculado. En el caso expuesto la probabilidad pre test de RCF es cercana al 15% y el LR (+) de la ecografía es 5,7 en el caso de examen positivo para RCF, con lo cual la probabilidad post test se elevará a 50%, cambio suficiente para cruzar el umbral terapéutico. Por otro lado el LR (-) es 0,9 si la ecografía descarta RCF, con lo cual la probabilidad post test de RCF alcanza al 12% lo cual no nos ayuda a descartar el diagnóstico (Figura 3).

Mayor información podríamos obtener si el estudio hubiese considerado los LR para distintos percentiles de peso fetal estimado por la ultrasonografía. Así mientras menor percentil estimado, mayor sería el LR y la probabilidad de RCF. Cuando los estudios entregan esta información se habla de múltiples niveles de LR, situación que puede aportar aún más para la individualización de la probabilidad post test en el paciente específico.

#### *3. ¿Los resultados cambiarán mi estrategia de manejo?*

Si la ecografía sugiere una RCF, claramente cambiará la estrategia de manejo, pese a reconocer que un 50% de los fetos no estarán afectados. Por el contrario, la pobre capacidad para descartar RCF hace que la vigilancia y el seguimiento se extiendan a todos los fetos de pacientes con síndrome hipertensivo del embarazo, lo cual resalta la importancia de la sospecha clínica, incluso frente al resultado de una ecografía con biometría fetal normal. En esta situación, si la sospecha clínica sigue siendo alta, necesitamos de otra estrategia que permita cambiar la probabilidad pre test, como por ejemplo realizar una curva de crecimiento fetal (7).



Copyrighted © (2005), American Medical Association. All Rights reserved.

Figura 2. Nomograma de Fagan. Adaptado de Fagan TJ. Nomogram for Bayes's theorem N Engl J Med Jul 31, 1975; 293(5):257. En Jaeschke, R. Users' guide to the medical literature: III. How to use an article about diagnostic test: B. What are the results and will they help me in caring for my patients? JAMA, 1994; 271(9):703-7.

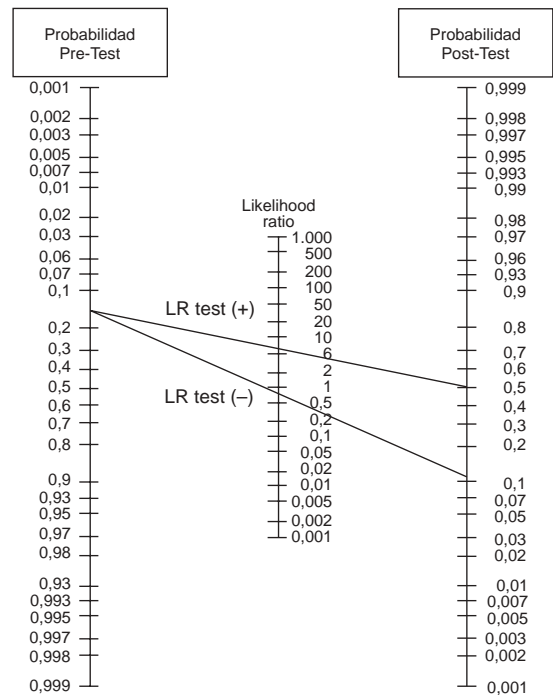


Figura 3. Cálculo de la probabilidad post test con el nomograma de Fagan.

4. ¿Los pacientes estarán mejor como resultado del examen?

El último criterio para establecer la utilidad de un examen es determinar si los beneficios esperados son mayores que los riesgos asociados. La respuesta a esta pregunta es pensar en el examen diagnóstico en conjunto con la maniobra terapéutica posible asociada. La respuesta a esta pregunta pudiera alcanzarse con estudios que:

- a) Evalúen aleatoriamente estrategias en donde el estudio diagnóstico esté disponible, comparado con estrategias donde el estudio no lo esté.
- b) Evalúen la frecuencia de resultados clínicamente relevantes en el seguimiento en el tiempo.

Ejemplos de este tipo de estudios son relevantes en el embarazo de alto riesgo y es así como han sido evaluadas las pruebas de evaluación de la unidad fetoplacentaria como la medición del Doppler umbilical para la primera alternativa (8) o el perfil biofísico fetal en la segunda (9). Respecto del ejemplo analizado en esta guía, el estudio en si mismo no permite conocer si el manejo del paciente será beneficiado por el resultado del examen diagnóstico, para ello requerimos de otros estudios diseñados en base a las estrategias mencionadas.

**BIBLIOGRAFÍA**

1. Baerheim A. The diagnostic process in general practice: has it a two-phase structure? *Fam Pract* 2001; 18(3): 243-5.
  2. Richardson WS, Wilson M, Gordon G. The Process Of Diagnosis. En: Guyatt G and Rennie D (eds). *User's Guides To The Medical Literature*. Chicago: American Medical Association 2002: 101-108
  3. Jaeschke R, Gordon G, Lijmer J. Diagnostic Tests. En: Guyatt G and Rennie D (eds). *User's Guides To The Medical Literature*. Chicago: American Medical Association 2002: 121-140
  4. Chauhan SP, Scardo JA, Magann EF, Devoe LD, Hendrix NW, Martin JN, Jr. Detection of growth-restricted fetuses in preeclampsia: a case-control study. *Obstet Gynecol* 1999; 93(5 Pt 1): 687-91.
  5. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002; 137(7): 598-602.
  6. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; 271(9): 703-7.
  7. Owen P, Maharaj S, Khan KS, Howie PW. Interval between fetal measurements in predicting growth restriction. *Obstet Gynecol* 2001; 97(4): 499-504.
  8. Neilson JP, Alfirevic Z. Doppler ultrasound for fetal assessment in high risk pregnancies. *Cochrane Database Syst Rev* 2000; 2): CD000073.
  9. Dayal AK, Manning FA, Berck DJ, Mussalli GM, Avila C, Harman CR y cols. Fetal death after normal biophysical profile score: An eighteen-year experience. *Am J Obstet Gynecol* 1999; 181(5 Pt 1): 1231-6.
-