

## Artículo por Invitación

# Claves para el correcto análisis de la información médica

Jorge Sapunar Z. MD, Msc

Departamento de Medicina Interna y Centro de Excelencia CIGES.  
Facultad de Medicina, Universidad de la Frontera, Temuco.

## Keys for the correct analysis of the medical information

Correspondencia a:  
Dr. Jorge Sapunar Z.  
Teléfono: (56-45) 296473  
E-mail: jsapunar@ufro.cl

Recibido: 16 Marzo de 2009  
Aceptado: 27 Mayo de 2009

### Introducción

La endocrinología destaca por la amplitud de sus contenidos, la extraordinaria velocidad de generación de nuevo conocimiento y la variedad de formas en que éste puede ser analizado. Por otra parte, el desarrollo de la informática y las telecomunicaciones pone cada día a disposición del médico endocrinólogo una cantidad abrumadora de información cuya calidad no siempre está garantizada.

Todos los profesionales de la salud que investigan, enseñan o atienden pacientes en el ámbito de la endocrinología requieren información. La información proviene de la propia experiencia, de la opinión de expertos y de la literatura biomédica. La velocidad de generación de nuevo conocimiento le ha conferido a la literatura biomédica una importancia creciente, pudiendo el lector encontrar en ella estudios primarios y revisiones bibliográficas. Se considera estudio primario toda investigación cuyas conclusiones no se fundamentan en los resultados de otros estudios. El poder evaluar estudios primarios es muy valorado por aquellos que se desempeñan en áreas más específicas de la endocrinología, pero requiere de conocimientos en metodología de la investigación. Cuando el interés del profesional de la salud es más amplio habitualmente no dispone de tiempo para evaluar todos los estudios primarios y debe recurrir a revisiones bibliográficas sistemáticas o no sistemáticas y guías de práctica clínica. Estas fuentes elaboradas de información pueden estar influidas por la opinión de sus autores (revisiones no sistemáticas) o tener una calidad variable, por lo que nuevamente el lector debe contar con herramientas para seleccionar lo que es realmente útil.

La Medicina Basada en la Evidencia (MBE) surgió hace décadas frente a la necesidad de evaluar y regular la aplicación de los avances logrados por las ciencias básicas y la investigación clínica en la comprensión, diagnóstico y tratamiento de las enfermedades. Han sido fundamentales en el desarrollo de la MBE la Colaboración Cochrane y grupos de trabajo de todo el mundo pertenecientes a INCLEN (International Clinical Epidemiology Network). El aporte de la MBE ha sido generar instrumentos para la lectura crítica de la información biomédica, el análisis conjunto de datos de investigaciones individuales (meta-análisis, revisiones sistemáticas), normas de elaboración de guías clínicas, de estudios económicos y de evaluación de tecnología sanitaria<sup>1</sup>.

La MBE no reemplaza el juicio clínico fundamentado en el método científico y la experiencia; es sólo una herramienta para poder integrar la información disponible y permitir al profesional de la salud aplicarla correctamente para atender a sus pacientes, educar o generar nueva información.

El objetivo de este artículo es entregar algunos instrumentos para evaluar en forma eficiente la calidad de diversas modalidades de estudios en endocrinología.

### Analizando estudios primarios en endocrinología

Independientemente del motivo que tenga el profesional de la salud para recolectar información en endocrinología (resolver un caso clínico, hacer una revisión bibliográfica para generar nueva información o para elaborar guías clínicas) es necesario convertir el problema en una pregunta

## Artículo por Invitación

de búsqueda. En la práctica clínica las preguntas de búsqueda suelen referirse a diagnóstico, daño, prevención y terapia<sup>2-4</sup>.

Al revisar una publicación la mayor parte de los profesionales de la salud le conceden más importancia a la discusión y conclusiones que a la metodología utilizada, lo cual puede conducir a aceptar resultados incongruentes, imprecisos e inaplicables a nuestros pacientes. Por ello, luego de establecer si un determinado estudio aborda nuestra pregunta de búsqueda, y para lograr una lectura eficiente debemos revisar sucesivamente su validez interna (riesgo de sesgos) los resultados (magnitud y precisión) y la validez externa (aplicabilidad)<sup>2-4</sup>.

A continuación analizaremos la forma de evaluar críticamente publicaciones que abordan los diferentes tipos de preguntas de búsqueda bibliográfica.

### Estudios de pruebas diagnósticas

El proceso diagnóstico de una enfermedad se inicia en el paciente, siendo nuestra incertidumbre respecto a su estado mórbido (probabilidad pre-test) lo que justifica la solicitud de estudio adicional. Por ello, antes de iniciar una búsqueda bibliográfica para determinar el valor diagnóstico de un determinado examen, debemos conocer la probabilidad de que nuestro paciente tenga esa determinada enfermedad en base a los elementos clínicos disponibles. La estimación de la probabilidad pre-test de una enfermedad es subjetiva por lo que una alternativa es extrapolar a nuestro paciente la prevalencia de la enfermedad de su población de origen<sup>2</sup>. El año 2007 Heemstra y colaboradores<sup>5</sup>, comunicaron un estudio cuya pregunta de investigación fue: ¿Cuál es el valor diagnóstico de la concentración sérica de tiroglobulina previo a la ablación con I<sup>131</sup> en pacientes con tiroidectomía por carcinoma diferenciado del tiroides, para predecir la recurrencia del tumor?

La lectura crítica de una publicación de diagnóstico se inicia con la determinación de su validez interna y para ello basta chequear 2 propiedades del estudio:

- Comparación “a ciegas” del resultado de la prueba diagnóstica con un estándar de referencia apropiado.
- Uso de una muestra de población con todo el espectro de pacientes a los cuales se aplicaría la prueba diagnóstica en la práctica clínica.

En nuestro ejemplo la medición de tiroglobulina sérica previa a la ablación con radioyodo se realizó rutinariamente en todos los pacientes, independientemente de la aplicación del estándar de referencia. Se incluyeron pacientes operados por carcinoma diferenciado del tiroides con indicación de radioyodo, pero que no tuvieran enfermedad residual evidente luego de la cirugía. Se excluyeron también sujetos con título elevado de anticuerpos antitiroglobulina.

Si la publicación no consideró estos dos aspectos podemos desecharla y si lo hizo debemos chequear dos propiedades adicionales:

- Aplicación del estándar de referencia independientemente del resultado de la prueba diagnóstica evaluada.

- Descripción detallada de la metodología para aplicar la prueba diagnóstica evaluada y permitir su replicación

En el estudio de Heemstra todos los pacientes fueron sometidos al estándar de referencia y se realizó una descripción adecuada de la técnica de laboratorio requerida para cuantificar la tiroglobulina sérica. También se definió explícitamente el desenlace (recurrencia tumoral).

Si la publicación cumple con estos requisitos podemos considerar válidos sus resultados y estamos en condiciones de analizarlos. Habitualmente las propiedades de una prueba diagnóstica son expresadas en la literatura como sensibilidad (proporción de sujetos con prueba diagnóstica positiva entre los que están enfermos) y especificidad (proporción de sujetos con prueba diagnóstica negativa entre los que están sanos). Sin embargo, lo que interesa en la práctica clínica es conocer la proporción de sujetos enfermos entre los que tienen la prueba positiva (valor predictivo positivo) y la proporción de sujetos que están sanos entre los que tienen la prueba diagnóstica negativa (valor predictivo negativo). Los valores predictivos (VP) dependen de la prevalencia de la enfermedad en la muestra de sujetos en los que se evaluó la prueba diagnóstica, por lo que su aplicación directa sólo es válida si la probabilidad pre-test de enfermedad de los pacientes del estudio es la misma que la de nuestra población. Esto habitualmente no es así por lo que el lector debe calcular el valor predictivo de la prueba diagnóstica para sus propios pacientes y para ello éste debe conocer la probabilidad pre-test de la enfermedad en su paciente y la publicación debe aportar los datos para el cálculo de las razones de verosimilitud (LR). El valor predictivo de la prueba diagnóstica es igual al producto de la probabilidad pre-test por LR, sin embargo, como la primera es una proporción y la segunda una razón no se pueden multiplicar por lo debemos convertir la probabilidad pre-test en Odds pre-test. En la mayor parte de los casos en que la prueba diagnóstica es una variable continua no se conoce el valor de corte para definir normalidad, por ello es recomendable que el estudio defina este valor mediante una curva receptor-operador (ROC). Por lo tanto, si en la comunicación de los resultados falta una tabla de sensibilidad y especificidad para distintos valores de la prueba diagnóstica o un gráfico con ROC, es un aspecto negativo respecto de la calidad metodológica del estudio.

A continuación expondremos como hacerlo utilizando datos del estudio de Heemstra y cols:

Resultado del test de tiroglobulina	Condición mórbida		
	Enfermo (recurrencia)	Sano (no recurrencia)	
Positivo: > 27,5 ug d/L	A (29)	B (19)	A+B (48)
Negativo: < 27,5 ug d/L	C (4)	D (170)	C+D (174)
	A+C (33)	B+D (189)	A+B+C+D (222)

## Artículo por Invitación

1. Estimación de la razón de verosimilitud para test positivo (LR+) y para test negativo (LR-):  
Sensibilidad:  $A/A+C = 0,878$   
Especificidad:  $D/B+D = 0,899$   
LR (+):  $\text{Sensibilidad}/1\text{-Especificidad} = 9$   
LR (-):  $1\text{-Sensibilidad}/\text{Especificidad} = 0,13$
2. Estimación de la Odds pre-test de enfermedad: En nuestro centro la probabilidad de recidiva es 0,3 (30%)  
Odds pre-test:  $\text{Probabilidad pre-Test}/1\text{- Probabilidad pre-Test} = 0,428$
3. Estimación de la Odds Post-test de enfermedad para test positivo y negativo  
Odds post-test (+):  $\text{LR (+)} \times \text{Odds pretest} = 3,85$   
Odds post-test (-):  $\text{LR (-)} \times \text{Odds pretest} = 0,0556$
4. Estimación de la probabilidad post-test (valores predictivos)  
VP (+):  $\text{Odds post-test (+)}/\text{Odds post-test (+)} + 1 = 0,793 (79,4\%)$   
VP (-):  $\text{Odds post-test (-)}/\text{Odds post-test (-)} + 1 = 0,052 (5,2\%)$

Del análisis anterior se puede concluir que los valores predictivos son probabilidades post-test y que además de depender de las propiedades intrínsecas de la prueba diagnósticas dependen también de la prevalencia de la enfermedad en la muestra de población estudiada<sup>2</sup>.

La interpretación directa de las razones de verosimilitud es que cifras extremas como  $> 10$  ó  $< 0,1$  producen grandes cambios en la probabilidad pre-test y reflejan gran potencia diagnóstica de un examen. Valores entre 5-10 y 0,1-0,2 indican potencia moderada en tanto que LR entre 1-2 y 0,5-1 no producen cambios significativos en la probabilidad pre-test y sugieren que el examen no tiene poder de discriminación.

En el estudio de Heemstra el valor de corte de tiroglobulina se fundamentó en datos de la literatura. No se realizó una curva receptor-operador.

Al aplicar el valor de corte propuesto por Heemstra a nuestros pacientes, cuando la tiroglobulina sérica previa a la ablación es mayor de 27,5 ug/dL la probabilidad de recurrencia aumenta de 30 a 79,4% y cuando ella es menor que dicho valor el riesgo de recurrencia baja a 5,2%. Basados en el significado del valor de LR el examen evaluado tiene una capacidad diagnóstica moderada.

Una vez que hemos establecido la validez de los resultados del estudio y obtenido las razones de verosimilitud debemos juzgar la aplicabilidad del estudio. Para ello debemos contestar las siguientes preguntas: ¿Puedo replicar el examen en mis pacientes con las mismas características de reproducibilidad y validez que en el estudio? ¿Los resultados del estudio son aplicables a mis pacientes? ¿Los resultados del examen permitirán modificar conductas terapéuticas?<sup>2</sup>.

### Estudios de daño

En endocrinología clínica es frecuente que nuestros pacientes tengan desenlaces adversos atribuibles a factores genéticos, ambientales o intervenciones terapéuticas. Para

establecer una relación causa-efecto entre una exposición o intervención y un desenlace adverso el mejor diseño de investigación son los estudios experimentales controlados y aleatorizados. En este caso el investigador aplica la exposición (intervención) a los sujetos experimentales comparando la ocurrencia del desenlace adverso con un grupo control. Utilizar estudios experimentales para evaluar daño no es éticamente aceptable a menos que éste no sea el objetivo primario del estudio, por ello se prefieren los estudios observacionales de cohortes en que el investigador selecciona sujetos expuestos y no expuestos a un determinado factor para comparar la incidencia de una o más formas de daño entre los grupos (cohortes). La principal limitación de este tipo de estudio es la pérdida de seguimiento de sujetos y el tiempo requerido para obtener resultados. Otra forma de evaluar daño es mediante estudios observacionales de casos y controles donde los sujetos son seleccionados en base al daño y retrospectivamente se indaga respecto a la exposición a una o más variables. Este diseño permite valuar formas infrecuentes de daño y no requiere seguimiento, sin embargo, es muy susceptible a sesgo por depender del recuerdo o registro de la exposición. Al no tener la certeza de que la exposición haya precedido al daño los estudios de casos y controles no permiten establecer causalidad a menos que las exposiciones sean inherentes a la persona como género, etnia, mutaciones y polimorfismos. Finalmente, muchas exposiciones o intervenciones y desenlaces adversos son comunicados en series de casos clínicos, diseño que no permite establecer asociaciones entre exposición y desenlace por carecer de grupo control<sup>6</sup>.

“The Million Women Study” (MWS) fue un gran estudio de cohortes diseñado para contestar la pregunta: ¿Cuál es el efecto de la terapia de reemplazo hormonal en el riesgo de cáncer de mama incidente en mujeres post-menopáusicas de Gran Bretaña? Como ejemplo de un artículo de daño analizaremos un estudio post hoc derivado del MWS cuya pregunta de investigación fue: ¿Cuál es el efecto de la terapia de sustitución hormonal en el riesgo de cáncer endometrial en mujeres post-menopáusicas no hysterectomizadas de Gran Bretaña<sup>7</sup>.

En la primera etapa del análisis crítico de una publicación de daño es determinar su validez interna, para lo cual debemos identificar 4 propiedades del estudio<sup>3</sup>:

- Descripción en los grupos de comparación, de variables distintas a la exposición que puedan afectar la ocurrencia del desenlace adverso.
- Que la frecuencia o magnitud de las mencionadas variables sea semejante en los grupos de comparación.
- Que exposiciones y desenlaces sean evaluados de la misma forma en los grupos de comparación.
- Seguimiento adecuado.

En el estudio MWS fueron seguidas 5 cohortes expuestas a diferentes modalidades de terapia hormonal de reemplazo y una sin la exposición. En todos los sujetos se registró la edad, nivel socio-económico, paridad, uso previo

## Artículo por Invitación

de anticonceptivos orales, IMC, actividad física, consumo de alcohol, hábito de fumar y antecedente de hipertensión arterial. Estas variables se distribuyeron de manera similar entre las cohortes. Sin embargo, el MWS no consideró todas las variables relevantes que afectan el riesgo del desenlace, particularmente el motivo de elección de una determinada forma de terapia. Exposición y desenlace fueron registrados de la misma forma en expuestos y no expuestos. El principal defecto metodológico del MWS reside en el que la información respecto al estado de la exposición durante el seguimiento fue obtenida por una encuesta, la cual sólo retornaron a los investigadores una proporción minoritaria de los sujetos.

Si estas preguntas fueron abordadas satisfactoriamente podemos continuar con el análisis de la publicación. Hay que destacar que los estudios de casos y controles carecen de seguimiento y ninguna serie clínica cumple con las propiedades anteriormente señaladas.

Para que un estudio sea totalmente válido debe existir una clara relación temporal entre exposición y desenlace, lo que sólo pueden lograr los estudios prospectivos o estudios de casos y controles con variables de exposición inherentes al sujeto como género, etnia y marcadores genéticos. Finalmente, es deseable que se establezca una relación dosis-respuesta entre exposición y desenlace. Aunque el estudio MWS exhibe una adecuada relación temporal entre exposición y desenlace sus defectos metodológicos no permiten continuar adelante con el análisis.

Los resultados de un estudio de daño deben establecer si existe asociación entre la exposición y el desenlace (prueba de hipótesis), expresar la magnitud de la asociación (RR o HR para estudios prospectivos, OR para estudios retrospectivos) y estimar el efecto del error de muestreo o precisión (intervalos de confianza). A continuación expondremos como hacerlo utilizando algunos datos del estudio MWS referentes al efecto de Tibolona en el riesgo de carcinoma endometrial incidente:

Exposición a tibolona	Desenlace (+) Cáncer de endometrio	Desenlace (-)
Expuesto	A (86)	B (27.942)
No expuesto	C (763)	D (395.022)

1. Riesgo Relativo (RR)  
 $[A/(A+B)] / [C/(C+D)]$   
 $86/28.028 / 763/395.785 = 1,59$
2. Riesgo Atribuible (RA)  
 $A/(A+B) - C/(C+D)$   
 $86/28.028 - 763/395.785 = 0,0012$
3. Número necesario a tratar (NNT)  
 $1/RA$   
 $1/0,0012 = 833$
4. Razón de Odds (OR)  
 No aplicable a estudio MWS  
 $(A/C) / (B/D)$

No es aceptable que un estudio de daño no comunique intervalos de confianza, sin embargo, éstos se pueden calcular mediante programas estadísticos (Epi Info, STATA). El estudio MWS para el efecto de Tibolona en el riesgo de carcinoma endometrial incidente reportó un intervalo de confianza de 1,43 a 2,25.

Una vez establecida la validez, magnitud y precisión de los resultados debemos evaluar la aplicabilidad del estudio considerando las semejanzas entre la muestra de la población del estudio y nuestros pacientes. También es necesario establecer la relevancia clínica de los hallazgos para evitar la exposición o intervención evaluada. Cuando en un estudio la magnitud del efecto se expresa como riesgo relativo el lector puede tener una impresión equivocada de la importancia clínica del hallazgo, por lo que es recomendable obtener indicadores de magnitud que consideren efectos absolutos como reducción de riesgo atribuible o NNT. El análisis post hoc del estudio MWS para el desenlace carcinoma endometrial incidente carece de validez interna, sin embargo, es interesante hacer notar que el efecto relativo de la Tibolona pierde importancia clínica al expresarlo como NNT.

### Estudios de terapia y prevención

Cuando el profesional de la salud desea obtener información respecto al efecto de una intervención preventiva o terapéutica en endocrinología, el diseño de investigación que mejor contesta la pregunta son los estudios experimentales controlados y aleatorizados<sup>4</sup>. HERS, es un estudio experimental cuya pregunta de investigación fue: ¿Cuál es el efecto de la terapia de reemplazo hormonal combinada continua en el riesgo de nuevos eventos cardiovasculares en mujeres post-menopáusicas, no hysterectomizadas, menores de 80 años con cardiopatía coronaria?<sup>8</sup>

La primera etapa en la lectura crítica de una publicación de prevención o terapia es establecer su validez interna identificando las siguientes 4 propiedades del estudio<sup>4</sup>:

- Asignación aleatoria de la intervención.
- Ocultamiento de la secuencia de aleatorización.
- Comunicación de pérdidas de seguimiento.
- Análisis de los sujetos en los grupos a los que originalmente fueron asignados (análisis por intención de tratar).

En el estudio HERS la intervención fue asignada en forma aleatorizada y estratificada por centro clínico involucrado. Los autores comunicaron las pérdidas de sujetos durante los 4 años de seguimiento y éstas fueron sorprendentemente bajas. La edad avanzada y la cardiopatía coronaria de la población de interés explican la mortalidad mayor de 10% en ambos grupos. El seguimiento podría ser considerado adecuado dado el alto riesgo de eventos cardiovasculares en los sujetos experimentales; sin embargo, al analizar el comportamiento del desenlace primario en el tiempo debería haberse prolongado algo más el estudio. Los resultados del estudio HERS se analizaron como intención de tratamiento.

## Artículo por Invitación

Otros elementos que a veces pueden no estar presentes por la naturaleza de la pregunta de investigación, pero que indican buena calidad metodológica son<sup>4</sup>:

- Enmascaramiento de la intervención a los sujetos, investigadores y tratantes.
- Control de las co-intervenciones no censuradas.

En el estudio HERS la intervención fue enmascarada a todos los participantes. Se hizo una completa caracterización de los sujetos experimentales. La frecuencia y magnitud de otras variables de exposición así como de co-intervenciones no censuradas fueron similares en ambos grupos.

Los resultados de un estudio de prevención o terapia deben establecer si existe asociación entre la intervención y el desenlace (prueba de hipótesis), expresar la magnitud de la asociación (RR) y estimar el efecto del error de muestreo o precisión (intervalos de confianza).

El desenlace primario del estudio HERS fue la incidencia de eventos cardiovasculares. El efecto de la terapia hormonal de reemplazo no fue significativamente diferente al del placebo ( $p$  0,91) y correspondió a un HR de 0,99 con un intervalo de confianza de 0,88 a 1,22. El efecto de la intervención tampoco fue significativamente diferente al del placebo en otros desenlaces primarios y secundarios.

Una vez establecida la validez, magnitud y precisión de los resultados debemos evaluar la aplicabilidad del estudio considerando las semejanzas entre la muestra de población del estudio y nuestros pacientes. También es necesario establecer si se consideraron todos los resultados clínicamente relevantes y si los beneficios de la intervención compensan sus efectos adversos y costos.

Las pacientes evaluadas en el HERS tenían una edad promedio de  $67 \pm 7$  años, eran predominantemente caucásicas, el 13% fumaba, el 18% padecía DM2, más de la mitad tenía un IMC  $> 27\%$  y el 60% no hacía ejercicio. El estudio consideró todos los desenlaces relevantes al objetivo y no se demostró que la intervención fuera efectiva en reducir el riesgo cardiovascular en mujeres postmenopáusicas con enfermedad coronaria previa.

### Analizando información elaborada en endocrinología

Tomar decisiones en clínica fundamentados en la evidencia puede ser difícil porque no tenemos acceso a toda la información ni disponemos de tiempo para reunirla. Aunque pudiéramos encontrar toda la información referente a un tema es necesario tener las destrezas para evaluar su calidad y convertirla en una recomendación final. Tradicionalmente, este problema ha sido abordado por los textos de medicina o las revisiones literarias de las revistas científicas; sin embargo, los contenidos de los textos no se renuevan con la velocidad que se genera nuevo conocimiento y al igual que las revisiones literarias representan la opinión de expertos respaldada por referencias bibliográficas.

Uno de los aportes de la MBE ha sido el desarrollo de revisiones sistemáticas y guías clínicas, sin embargo, estas modalidades de información elaborada pueden tener calidad variable y deben ser sometidas a lectura crítica. A continuación revisaremos mediante ejemplos los aspectos que debemos considerar para una lectura eficiente de información elaborada<sup>1</sup>.

### Revisiones sistemáticas y meta-análisis

El objetivo de las revisiones sistemáticas es contestar una pregunta de investigación obteniendo toda la evidencia disponible mediante una estrategia de búsqueda reproducible. A continuación la información es seleccionada en base a criterios de contenido y metodología. Si los estudios primarios seleccionados son suficientes y homogéneos es posible obtener un resultado de conjunto mediante meta-análisis. La Revisión sistemática es un diseño de investigación retrospectivo cuyo sujeto de investigación es cada estudio primario incluido en una revisión sistemática<sup>9</sup>.

Las revisiones sistemáticas pueden contestar preguntas de diagnóstico, daño, terapia y prevención, por lo que su lectura crítica también nos obliga a revisar sucesivamente su validez interna (riesgo de sesgos), resultados (magnitud y precisión) y validez externa (aplicabilidad)<sup>9</sup>.

Nissen y cols<sup>10</sup>, comunicaron una revisión sistemática y meta-análisis en que se preguntaron: ¿Cuál es el efecto de la rosiglitazona en el riesgo de infarto agudo al miocardio y muerte de causa cardiovascular?

La primera etapa en la lectura crítica de una revisión sistemática es establecer su validez interna identificando las siguientes 2 propiedades del estudio:

- Definición precisa del problema clínico abordado por el estudio.
- Descripción de criterios de selección de estudios originales utilizados (tipo de pacientes, exposiciones, resultados de interés, aspectos metodológicos).

Nissen y cols, definieron en forma precisa el problema clínico abordado. Se consideraron para el análisis ensayos clínicos aleatorizados que evaluaran rosiglitazona y un comparador (placebo u otro hipoglicemiente oral) por un período superior a 24 semanas. No se menciona la población de interés ni si los desenlaces mencionados en la pregunta de investigación debían ser primarios, secundarios o reportados como efectos adversos. Si la revisión sistemática cumple con estos requisitos podemos continuar su análisis chequeando los siguientes aspectos:

- Inclusión de todos los estudios relevantes.
- Evaluación reproducible de la validez de los estudios primarios (ej. escala de Jaddad).
- Homogeneidad de los resultados de los estudios originales (evaluación visual de los gráficos de bosque, prueba de homogeneidad, prueba de consistencia).

Nissen y Wolski, no detallaron su estrategia de búsqueda pero recurrieron a varias fuentes de información como



## Artículo por Invitación

el sitio web de la FDA y Glaxo Smith Kline. Otro grupo de ensayos clínicos fueron identificados en la literatura (DREAM y ADOPT). No mencionaron la forma en que evaluaron la metodología de cada estudio primario, pero consignaron la referencia, número de registro, fase (II, III ó IV), duración, intervención y número de pacientes, calidad. En la publicación no aparecen gráficos donde se pueda apreciar la heterogeneidad de los estudios originales individualmente, agrupados por categorías de estudios, por desenlace o por comparador activo. Podría haber reparos en comparar estudios en pre-diabéticos (DREAM) con estudios en diabéticos (ADOPT) o agrupar datos con comparadores como la insulina y el placebo o entre estudios donde no se señala el control metabólico de la diabetes. Los autores aplicaron la estadística Q de Cochrane en un modelo de efectos fijos y no encontraron heterogeneidad.

Si la revisión sistemática tiene validez interna podemos analizar sus resultados, que deben establecer si existe asociación entre la exposición y el desenlace (prueba de hipótesis), expresar la magnitud de la asociación y estimar el efecto del error de muestreo o precisión (intervalos de confianza).

La magnitud de asociación en meta-análisis se puede expresar como OR o RR cuando éste aborda preguntas de daño, prevención o terapia. Recientemente se han diseñado meta-análisis que expresan efecto mediante NNT. También, cuando se evalúan variables continuas, un meta-análisis nos puede entregar promedios ponderados y cuando se evalúan pruebas diagnósticas de diagnóstico sus propiedades se pueden resumir como LR.

Cuando se demuestra heterogeneidad en un meta-análisis los autores deben realizar una prueba de sensibilidad, excluyendo estudios primarios del análisis basados en su menor calidad metodológica, o bien se puede dividir el meta-análisis original en análisis de sub-grupos de población, exposición y desenlace. Si la heterogeneidad es muy alta es recomendable no realizar el meta-análisis.

En la Tabla 1 se pueden apreciar los resultados del estudio de Nissen. La magnitud del efecto expresado como OR fue 1,29 pero la falta de precisión de los resultados sugiere tamaño muestral insuficiente (IC 95% 0,93 a 1,78). La prueba de heterogeneidad no fue significativa (p 0,89) y el análisis fue consistente ( $I^2$  de Higgins 0%).

**Tabla 1. Efecto de la Rosiglitazona en el riesgo de infarto agudo al miocardio<sup>10</sup>**

Estudio	OR (efectos aleatorios)	IC 95%
Adopt 2006	1,32	0,81 - 2,15
Dream 2006	1,50	0,65 - 3,53
Estudios pequeños	1,19	0,71 - 1,98
Total	1,29	0,93 - 1,78

Una vez establecida la validez, magnitud y precisión de los resultados debemos establecer si son aplicables a nuestros pacientes evaluando si se consideraron todos los resultados clínicamente relevantes y si los beneficios de la intervención compensan sus efectos adversos y costos.

### Guías para la práctica clínica

Hay enfermedades del ámbito de la endocrinología que destacan por su importancia epidemiológica y en los cuales existen controversias diagnósticas o terapéuticas. En estos casos es necesario que grupos de expertos, disponiendo de toda la evidencia, generen recomendaciones que se comunican en guías de práctica clínica. Considerando la extensión que pueden tener algunas guías de práctica clínica es conveniente considerar los siguientes aspectos metodológicos antes de aplicarlas<sup>11</sup>:

1. Descripción del objetivo general de la guía.
2. Descripción de los pacientes a quienes se pretende aplicar la guía.
3. Grupo de elaboración de la guía con representación de todos los profesionales relevantes.
4. Definición del usuario de la guía.
5. Búsqueda sistemática de la información.
6. Definición de criterios de selección de la evidencia.
7. Definición de los métodos para formular recomendaciones.
8. Calificación de las recomendaciones de acuerdo a la calidad de la evidencia que la sustenta.
9. Revisión de la guía por expertos previa a su publicación.
10. Descripción de los costos de aplicación de las recomendaciones.
11. Independencia editorial de la entidad que financia la iniciativa.
12. Registro de conflictos de interés de los miembros del grupo de elaboración.

La fortaleza de las recomendaciones de una guía de práctica clínica se pueden expresar de acuerdo a la escala de Oxford<sup>12</sup> o mediante el sistema GRADE<sup>13</sup>. En la Tabla 2 se detallan los niveles de calidad de la evidencia y en la Tabla 3 los grados de recomendación en una guía clínica.

### Conclusiones

Para obtener información en el ámbito de la endocrinología en forma eficiente, el profesional de la salud debe abordar las publicaciones de la literatura biomédica desde un punto de vista metodológico. Para ello es necesario establecer el tipo de pregunta que el estudio pretende contestar y de acuerdo a ello establecer su validez interna.

Las revisiones sistemáticas y guías clínicas también pueden tener defectos metodológicos y antes de aplicar sus conclusiones a nuestra práctica clínica deben ser evaluadas.

## Artículo por Invitación

**Tabla 2. Niveles de Evidencia (Centro de Medicina Basada en la Evidencia de Oxford)**

Nivel	Diagnóstico	Daño	Prevención/Terapia
1a	Meta-análisis homogéneo de estudios de diagnóstico nivel 1	Meta-análisis homogéneo de ensayos clínicos	Meta-análisis homogéneo de ensayos clínicos
1b	Estudios de cohortes de validación	Ensayos clínicos con IC estrecho	Ensayos clínicos con IC estrecho
1c	Estudios de diagnóstico con LR extremos		
2a	Meta-análisis homogéneo de estudios de diagnóstico nivel 2	Meta-análisis homogéneo de estudios de cohortes	Meta-análisis homogéneo de estudios de cohortes
2b	Estudio de cohortes exploratorio	Estudios de cohortes Ensayos clínicos de baja calidad	Estudios de cohortes Ensayos clínicos de baja calidad
3a	Meta-análisis homogéneo de estudios de diagnóstico nivel 3	Meta-análisis homogéneo de estudios de casos y controles	Meta-análisis homogéneo de estudios de casos y controles
3b		Estudios de casos y controles	Estudios de casos y controles
4	Estudios de casos y controles	Series de casos Evidencia 2b y 3b de mala calidad	Series de casos Evidencia 2b y 3b de mala calidad
5	Opinión de expertos no basada en evidencia	Opinión de expertos no basada en evidencia	Opinión de expertos no basada en evidencia

**Tabla 3. Grados de Recomendación (Centro de Medicina Basada en la Evidencia de Oxford)**

Grado	Recomendación
A	Basada en estudios de nivel 1
B	Basada en estudios de niveles 2 y 3, o extrapolaciones de estudios de nivel 1
C	Basada en estudios de nivel 4, o extrapolaciones de estudios de nivel 2 y 3
D	Basada en nivel de evidencia 5, o estudios inconsistentes o no concluyentes de cualquier nivel

## Referencias

- Letelier LM, Moore P. 2003. La Medicina Basada en Evidencia. Visión después de una década. *Rev Méd Chile* 131: 939-946.
- Jaeschke R, Guyatt GH, Sackett DL. 1994. 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* 271: 703-707
- Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. 1994. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA* 271: 1615-1619.
- Guyatt GH, Sackett DL, Cook DJ. 1993. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 270: 2598-2601.
- Heemstra KA, Liu YY, Stokkel M, Kievit J, Corssmit E, Pereira AM, et al. 2007. Serum thyroglobulin concentrations predict disease-free remission and death in differentiated thyroid carcinoma. *Clin Endocrinol* 66: 58-64.
- Hennekens CH, Buring JE. 1987. Design strategies in epidemiologic research. In Mayrent SL (ed.), *Epidemiology in Medicine*. Boston/Toronto: Little, Brown and Company, pp. 16-29.
- Million Women Study Collaborators. 2005. Endometrial cancer and hormone-replacement therapy in The Million Women Study. *Lancet* 365: 1543-1551.
- Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, et al. 1998. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 280: 605-613.
- Oxman AD, Cook DJ, Guyatt GH. 1994. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. *JAMA* 272: 1367-1371.
- Nissen SE, Wolski K. 2007. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*; 356: 1-15.
- Hanson D, Hoss BL, Wesorick B. 2008 Evaluating the evidence: guidelines. *AORN J* 88:184-196.
- Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, et al. 2004. Levels of evidence and grades of recommendations. Oxford: Oxford Centre of Evidence-Based Medicine [citado 8 Feb 2004]. Disponible en: [www.cebm.net/levels\\_of\\_evidence.asp](http://www.cebm.net/levels_of_evidence.asp)
- Guyatt GH, Oxman AD, Vist G, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schunemann HJ and the GRADE Working Group. 2008. Rating quality of evidence and strength of recommendations GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336: 924-926.