

## Análisis de subgrupos: ¿Mejoran la interpretación de los resultados o nos inducen al error?

ROBERTO CANDIA B.<sup>1,2</sup>, SOLANGE RIVERA<sup>1,3</sup>, IGNACIO NEUMANN<sup>1,4,5</sup>

### Subgroup analysis: Do they help or mislead when interpreting study results?

<sup>1</sup>Unidad de Medicina Basada en Evidencia.

<sup>2</sup>Departamento de Gastroenterología.

<sup>3</sup>Departamento de Medicina Familiar.

<sup>4</sup>Departamento de Medicina Interna.

<sup>5</sup>Unidad Docente Asociada Hospital Dr. Sótero del Río. Facultad de Medicina, Pontificia Universidad Católica de Chile. Santiago de Chile.

Correspondencia a:

Dr. Ignacio Neumann.

E-mail: ignacio.neumann@gmail.com

Los estudios clínicos con distribución aleatoria (“randomizados”) (ECR) y las revisiones sistemáticas (RS) de ECR proporcionan la evidencia más confiable acerca del efecto de las intervenciones terapéuticas<sup>1-5</sup>. Habitualmente en los ECR de gran escala y en las RS se incluyen pacientes con una amplia variedad de características, por ejemplo: distintas edades, manifestaciones de enfermedad, gravedad o etapa de progresión de la enfermedad. En la práctica diaria, los clínicos aplicamos las terapias a individuos con características particulares, por lo que necesitamos saber si el efecto beneficioso descrito para el paciente promedio es realmente aplicable a mi paciente en particular. Puede que el efecto de algunas terapias sea beneficioso o incluso perjudicial sólo en subgrupos de individuos con ciertas características. ¿Cómo saber si el beneficio es para todos o sólo para algunos pacientes? ¿Cómo saber cuáles pacientes se benefician más de una intervención determinada? Para resolver estas interrogantes se utiliza con frecuencia el análisis de subgrupo. El objetivo del siguiente artículo es difundir las herramientas disponibles para interpretar correctamente los análisis de subgrupo.

#### Concepto de grupo y subgrupo en un estudio clínico

Para clarificar la terminología primero debemos definir el concepto de “grupo” y “subgrupo” en un estudio clínico. Imaginemos un estudio diseñado para evaluar la capacidad de la droga X de disminuir el riesgo de infarto en pacientes mayores de 60 años:

- Dividimos a los pacientes en 2 grupos en forma aleatoria: A y B.
- Definimos que el grupo A va a corresponder al “grupo intervención” y recibirán la droga X.
- El grupo B va a corresponder al “grupo control” y recibirán placebo.
- Una vez obtenidos los datos nos preguntamos si tal vez en los pacientes diabéticos la droga X tiene mayor beneficio que en los no diabéticos, y decidimos analizar los datos de acuerdo a la presencia o ausencia de diabetes.

Los “grupos” en un estudio clínico son aquellos constituidos *a priori* para ser expuestos o no a una intervención, con el objetivo de comparar la ocurrencia de un determinado desenlace o *outcome*. En el ejemplo anterior, el grupo A (intervención) y el grupo B (placebo) corresponden genuinamente a los “grupos” del estudio. Un “subgrupo”, en cambio, corresponde a una subdivisión de la muestra original donde los pacientes cumplen con alguna característica de interés particular (por ejemplo: edad o gravedad de la enfermedad). Esta característica en común puede determinar una respuesta terapéutica distinta a la del resto de los pacientes de la muestra<sup>6</sup>.

En nuestro ejemplo: un análisis de toda la población incluiría la comparación de todos los pacientes del grupo A versus todos los pacientes del grupo B (“análisis de grupo”). Un análisis de subgrupos en relación a la presencia de diabetes mellitus, en cambio, compararía los pacientes diabéticos del grupo A versus los pacientes diabéticos del grupo B, para determinar si el efecto de la intervención en ellos es distinto del efecto general observado en la comparación de los grupos A y B.

### ¿Por qué realizar análisis de subgrupo?

La aplicabilidad de los resultados de un ECR depende de las características o del perfil de riesgo de los pacientes incluidos en el estudio<sup>1,6,7</sup>. Cuando un estudio está constituido por individuos cuyas características son variadas, la extrapolación de sus resultados es factible para una amplia gama de pacientes. Esto, que parece una fortaleza, nos impide precisar si los resultados son realmente aplicables a pacientes con características específicas<sup>2</sup>. Una alternativa es estratificar la muestra en distintos subgrupos, con el objetivo de identificar con mayor precisión los pacientes que tendrán mayor o menor beneficio. Para comprender mejor esta situación recurriremos a un ejemplo:

Los ECR que evaluaron la cirugía versus el tratamiento médico en la estenosis carotídea sintomática, incluyeron pacientes con distintos grados de estenosis. Dado que la intervención (cirugía carotídea) es un procedimiento riesgoso que puede terminar en la muerte o en la ocurrencia del *outcome* que se quería prevenir (accidente vascular encefálico, AVE), es posible pensar que los pacientes que obtienen el mayor beneficio de la intervención son aquellos con el mayor riesgo de presentar el *outcome* adverso producto de la enfermedad. Los resultados de ECRs y RS<sup>8-10</sup> avalan esta hipótesis. En pacientes con alto grado de estenosis carotídea, es decir, alto riesgo de un *outcome* desfavorable, la cirugía resulta en la reducción de la mortalidad e incidencia de AVE. Al contrario, en pacientes con bajo grado de estenosis, es decir, bajo riesgo de un *outcome* desfavorable, la intervención resulta en un aumento de la mortalidad e incidencia de AVE. En este caso, el beneficio observado con la intervención ocurre sólo en un subgrupo de los pacientes estudiados. Si consideráramos el resultado global del estudio tendríamos una impresión equivocada del real efecto de la cirugía en estenosis carotídea sintomática. Por esta razón, parece razonable realizar análisis estratificados en subgrupos dentro de un mismo estudio, para identificar con precisión los individuos que concentran el beneficio o el riesgo.

### Diferencia entre riesgo basal y efecto de subgrupo

El riesgo basal corresponde al riesgo que tendría un paciente de presentar el *outcome* de interés

si no recibiera la intervención. En el caso de un estudio, se puede equiparar al riesgo en el grupo control. El problema está en que distintos pacientes tienen distintos riesgos basales, dependiendo de sus características personales o antecedentes mórbidos, lo que frecuentemente se confunde con el efecto de subgrupo. Para comprender mejor este fenómeno recurriremos a ejemplos:

- Un hombre diabético de 65 años, fumador, hipertenso sin tratamiento, con presión sistólica de 150 mmHg, colesterol total de 200 mg/dL, HDL 40 mg/dL, tiene riesgo basal de un evento cardiovascular mayor de 66,2% a 10 años<sup>11</sup>.
- Una mujer diabética de 50 años, no fumadora, sin hipertensión, con presión sistólica de 130 mmHg, colesterol total de 200 mg/dL, HDL 40 mg/dL, tiene riesgo basal de un evento cardiovascular mayor de 11,2% a 10 años<sup>11</sup>.
- Si ofrecemos terapia con estatinas para prevenir eventos cardiovasculares mayores, la reducción relativa de riesgo es de 30%<sup>12</sup>. En el primer paciente este riesgo relativo se traducirá en una reducción absoluta de 19,8% de eventos cardiovasculares mayores (la proporción correspondiente al 30% del riesgo basal de 66,2%), mientras que en la segunda paciente la reducción absoluta del riesgo será de 3,3% (la proporción correspondiente al 30% del riesgo basal de 11,2%).

Si comparamos los riesgos absolutos, se observa una clara diferencia entre el primer y segundo paciente (disminución de riesgo de 19,8% vs 3,3%). Esta diferencia está dada por los riesgos basales y no por un efecto de subgrupo, ya que la intervención tiene el mismo *efecto relativo* en ambos pacientes: reduce en 30% la probabilidad de presentar el evento adverso.

A diferencia del ejemplo anterior, en el caso de la cirugía carotídea observamos lo siguiente:

- En los pacientes con alto grado de estenosis (alto riesgo de presentar el *outcome* adverso) la intervención reduce a la mitad el riesgo de presentar un AVE o morir<sup>10</sup>. Si consideramos que el riesgo de presentar tales *outcomes* en estos pacientes es aproximadamente 30%<sup>9</sup> la intervención produce una reducción absoluta de 15%.
- En los pacientes con bajo grado de estenosis (bajo riesgo de presentar el *outcome* adverso), la cirugía aumenta en 25% el riesgo de presentar un AVE o morir<sup>10</sup>. Si consideramos que

el riesgo basal de los pacientes con bajo grado de estenosis de presentar los *outcomes* mencionados es de aproximadamente 12%, la intervención provoca un aumento absoluto de 3%.

En el ejemplo de las estatinas la intervención tiene el mismo efecto en los dos subgrupos de pacientes (reduce la probabilidad de presentar el *outcome* adverso en 30%) y las diferencias observadas en términos absolutos son sólo producto de un *riesgo basal diferente*, por tanto, no hay un *efecto de subgrupo*. En el segundo ejemplo la intervención tiene distintos efectos dependiendo del subgrupo de pacientes (en los pacientes de alto riesgo produce una reducción del riesgo relativo de 50%, mientras que en los pacientes de bajo riesgo produce un aumento del riesgo relativo de 25%). En este caso, estamos frente a un *efecto de subgrupo*. La única forma de diferenciar ambas situaciones es mediante la comparación de los efectos en términos relativos, los que permanecen constantes independientemente del *riesgo basal* y varían sólo si hay *efecto de subgrupo*. Los efectos medidos en términos absolutos varían en ambas situaciones, por lo que no permiten su diferenciación (Figura 1).

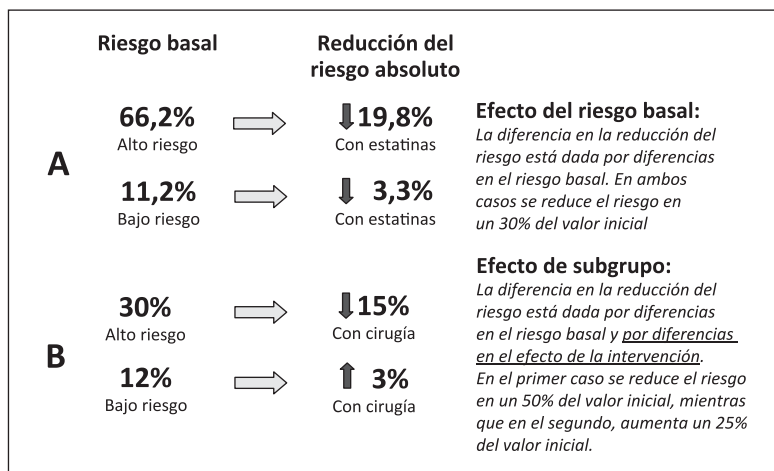
**¿Por qué los análisis de subgrupo nos pueden llevar a conclusiones erróneas?**

Los análisis de subgrupo nacen con el objetivo de predecir con mayor precisión cuáles pacientes se benefician con las terapias. Sin embargo, en ocasiones su aplicación poco cuidadosa puede

inducirnos a error. Para documentar lo anterior, en la Tabla 1 se enuncian algunos ejemplos de conclusiones obtenidas exclusivamente desde análisis de subgrupo. Todas estas afirmaciones fueron obtenidas de estudios clínicos publicados en revistas de alto impacto y posteriormente refutadas por múltiples estudios clínicos<sup>2,13-20</sup>. Estas afirmaciones estaban avaladas por análisis estadísticos que mostraban diferencia significativa entre los subgrupos, pero aún así la información fue errónea e indujo a los clínicos a conductas equivocadas ¿Cuál es la razón?.

**Tabla 1. Ejemplos de hallazgos en análisis de subgrupo que finalmente fueron irreales. Todas las afirmaciones fueron refutadas por estudios posteriores**

• La aspirina es inefectiva como prevención secundaria de AVE en mujeres <sup>2,26,27</sup>
• La terapia antihipertensiva es inefectiva como prevención cardiovascular primaria en mujeres <sup>30,31</sup>
• La terapia antihipertensiva es inefectiva o dañina en ancianos <sup>30</sup>
• IECA no reducen la mortalidad y hospitalización en pacientes con ICC que toman AAS <sup>32</sup>
• Los beta bloqueadores son inefectivos tanto en IAM de pared inferior como en ancianos <sup>29,33</sup>
• La trombolisis es inefectiva en IAM de > 6 h <sup>34</sup>
• El tamoxifeno es inefectivo en mujeres con cáncer de mama < 50 años <sup>35</sup>
• El amlodipino reduce la mortalidad en pacientes con ICC de origen no isquémico pero no en los de origen isquémico <sup>36</sup>



**Figura 1.** Efecto del riesgo basal y de subgrupos. A. Riesgo de un evento cardiovascular mayor en pacientes con riesgo cardiovascular aumentado. B. Riesgo de accidente vascular o muerte en pacientes con estenosis carotídea sintomática.

### **La potencial falacia de los análisis de subgrupo**

La interpretación de un hallazgo generado por un análisis de subgrupo debe ser cautelosa. La principal razón por la cual tienen un alto riesgo de *error* está relacionada con la necesidad de establecerlos *a priori*, en el momento en que se diseña el estudio. Cuando un subgrupo es establecido o definido después de la “randomización” o en un análisis “*post hoc*”, existe un alto riesgo de pesquisar diferencias inexistentes que aparecen solamente como producto del azar. Para comprender lo anterior necesitamos conocer algunos conceptos estadísticos generales:

En una situación ideal, un test estadístico nos mostraría un resultado significativo toda vez que exista una diferencia real y nos mostraría un resultado no significativo toda vez que las diferencias observadas se deban al azar. En la vida real esto no siempre ocurre y existen básicamente 2 tipos de errores en estadística: uno es que un test no sea capaz de mostrar una diferencia significativa cuando esta existe, y el otro es que muestre una diferencia significativa cuando esta no existe.

Los análisis de subgrupos están también expuestos a estos errores. En algunas ocasiones, habiendo diferencias reales entre los subgrupos de una población respecto del efecto de la intervención, los ECR o las RS no van a ser capaces de identificarlos (*error*). Esta situación suele ocurrir cuando se dispone de un tamaño muestral pequeño o la diferencia en el efecto de la intervención es de poca magnitud entre los subgrupos.

Una situación más peligrosa, por las consecuencias que trae en la aplicación de la evidencia, ocurre cuando un análisis de subgrupo muestra un efecto significativamente distinto de la intervención en 2 subgrupos (efecto de subgrupo) cuando este en realidad no existe. Respecto de esto es importante considerar que al repetir una prueba estadística, en múltiples ocasiones aumenta la probabilidad de encontrar diferencias que no existen, sólo por azar (*error*). Esta situación se ve reflejada en estudios con análisis en múltiples subgrupos. Cuando en un estudio se evalúa sólo un subgrupo (con el habitual nivel de significación de 0,05), existe una probabilidad de 5% de obtener un resultado significativo sólo por azar. Si evaluamos cinco subgrupos, la probabilidad de encontrar al menos un resultado significativo aumenta a 23%, al evaluar diez subgrupos, esta probabilidad es de

40%. Lo anterior refleja que mientras más subgrupos se consideren, mayor es la probabilidad de encontrar una diferencia significativa sólo por azar.

Desafortunadamente, en muchas ocasiones, en los estudios clínicos los análisis no se plantean *a priori*, por lo que su tamaño no es considerado en el diseño del estudio; el análisis de un alto número de subgrupos aumenta el riesgo de detectar diferencias inexistentes. En ocasiones los hallazgos de un análisis de subgrupo desencadenan conclusiones exageradas, a pesar de la ausencia de un fundamento biológico que lo explique<sup>2,21-23</sup>. En base a lo anterior, se han diseñado estrategias para interpretar las conclusiones obtenidas desde análisis de subgrupo.

### **Recomendaciones para interpretar análisis de subgrupo**

Existen varias iniciativas para evaluar e interpretar si la información obtenida desde un análisis de subgrupo es real o una falacia<sup>2,6,21-24</sup>. La más conocida se resume en la Tabla 2. Esta estrategia funciona como una lista de chequeo, a través de la cual se evalúa el resultado obtenido desde un análisis de subgrupo, considerándolo como “altamente plausible” o “extremadamente improbable”<sup>22</sup>. La evaluación de un análisis de subgrupo se realiza en 3 niveles: *Diseño del estudio*, *Análisis de los datos* y *Contexto general*. Sintetizaremos un análisis de subgrupo cuyos hallazgos son “altamente plausibles”:

#### **Diseño del estudio**

*Establecimiento a priori del subgrupo:* Un análisis de subgrupo debe ser establecido *a priori*, previo a la randomización. Muy rara vez las interacciones entre subgrupos no definidas ni sospechadas *a priori* son reales y pueden ser influenciadas por características pronósticas desbalanceadas<sup>2,6,21-24</sup>. Conocido es el ejemplo del estudio ISIS-2, que demuestra el beneficio global de la aspirina en el infarto agudo al miocardio<sup>18</sup>. Una vez finalizado el estudio, y a petición de algunos revisores del manuscrito, se realizaron múltiples análisis de subgrupo. Los investigadores de ISIS-2 incluyeron (entre ellos) un análisis según el signo zodiacal, para demostrar el riesgo potencial de la exploración de múltiples subgrupos a la vez. En este análisis se pesquió que el beneficio global

**Tabla 2. Criterios para evaluar la credibilidad de análisis de subgrupo (adaptado de BMJ 2010; 340: 850-54)**

Criterios para evaluar la credibilidad de análisis de subgrupo
<p><b>Diseño:</b></p> <ul style="list-style-type: none"> <li>- ¿Es la variable del subgrupo una característica medida basalmente o después de la randomización?</li> <li>- ¿Es el efecto sugerido por comparaciones dentro más que entre estudios?*</li> <li>- ¿Fue la hipótesis especificada a priori?</li> <li>- ¿Fue la dirección del efecto en el subgrupo especificada a priori?</li> <li>- ¿El efecto de subgrupo fue uno de entre un pequeño número de efecto de subgrupo evaluados?</li> </ul>
<p><b>Análisis:</b></p> <ul style="list-style-type: none"> <li>- ¿La prueba de interacción sugiere con baja probabilidad que el efecto de subgrupo se deba al azar?</li> <li>- ¿Es el efecto de subgrupo un efecto estadísticamente independiente?</li> </ul>
<p><b>Contexto:</b></p> <ul style="list-style-type: none"> <li>- ¿Es de magnitud relevante el tamaño del efecto de subgrupo?</li> <li>- ¿La interacción del efecto de subgrupo es consistente entre distintos estudios?</li> <li>- ¿La interacción del efecto de subgrupo es consistente entre outcomes cercanamente relacionados dentro del estudio?</li> <li>- ¿Existe evidencia indirecta que apoya la interacción de subgrupo (fundamento biológico)?</li> </ul>

\*Criterio aplicable sólo a revisiones sistemáticas de ensayos clínicos randomizados.

detectado no era efectivo en los pacientes de signo Géminis y Libra, lo cual no tiene sustento biológico y no fue confirmado por estudios posteriores.

**Establecimiento a priori de la dirección del efecto:** La dirección del efecto en el subgrupo debe ser establecida *a priori*. Muy rara vez las interacciones de subgrupo con dirección del efecto en el sentido opuesto al preestablecido son reales<sup>6,21-24</sup>.

Ejemplo: En un estudio que comparó el efecto de vasopresina versus norepinefrina en la mortalidad de pacientes con shock séptico, los autores establecieron *a priori* la hipótesis que la vasopresina sería más efectiva en pacientes más graves<sup>25</sup>. El análisis definitivo mostró justamente lo contrario: mayor beneficio en los pacientes menos graves. Estos resultados, opuestos a los establecidos *a priori*, son menos creíbles.

**Número acotado de subgrupos:** El número de subgrupos a analizar debe ser limitado, para evitar detectar diferencias significativas inexistentes, sólo por azar. A mayor número de subgrupos, mayor es el riesgo de detectar diferencias que no existen (*error*)<sup>2,21-24</sup>.

**Las comparaciones intraestudio son más confiables que aquellas interestudio:** En las RS de ECRs existe una dificultad adicional. En estas se analizan en conjunto varios ECRs, por lo que es posible llevar a cabo un análisis de subgrupo en 2 niveles: El primero es analizar sólo un subgrupo de cada ECR. Volviendo al ejemplo de la cirugía de carótida, sólo los pacientes con estenosis de alto grado en cada estudio. Para realizar este análisis es necesario que cada ECR reporte datos por separado para el subgrupo de interés, cosa que rara vez ocurre. Otra posibilidad, más factible, es analizar los estudios que incluyeron mayor cantidad con estenosis de alto grado versus los estudios con menor cantidad de estos pacientes (por tanto, con mayor proporción de pacientes con bajo grado de estenosis). A pesar que el objetivo es comparar el efecto de la intervención en pacientes con alto grado de estenosis versus pacientes con bajo grado de estenosis, en rigor lo que se está realizando es una comparación entre estudios. Si encontramos una diferencia en el efecto de la intervención, es difícil atribuir este hallazgo sólo a la proporción de pacientes con alto grado de estenosis, ya que los estudios no sólo se diferencian en sus poblaciones, también lo hacen en los procedimientos, cointervenciones, métodos de análisis, etc. Considerando lo anterior, es más probable que las diferencias observadas sean explicadas por las diferencias del diseño más que por la característica de la población de interés. Por esta razón, en una RS los análisis de subgrupo que consideran subpoblaciones específicas dentro de cada estudio (intraestudio) suelen ser más confiables que los análisis de subgrupos basados en comparaciones entre estudios (interestudios)<sup>21,22</sup>.

**Análisis de los datos:**

**Pruebas estadísticas adecuadas:** Un aspecto importante es descartar el azar como causa de las diferencias encontradas entre subgrupos. Lo ideal es realizar este análisis con test estadísticos especialmente diseñados con ese propósito: los test de interacción. Ellos nos permiten saber si las diferencias entre subgrupos son significativas<sup>2,6,22-24</sup>.

Los test de interacción relacionan la diferencia entre 2 o más subgrupos y el *outcome* de interés sin ajustar por otros factores (análisis univariado).

*Demostración que el efecto de subgrupo es independiente:* Cuando se detecta simultáneamente más de un efecto de subgrupo estadísticamente significativo en un estudio (hay diferencias significativas en el test de interacción para más de un subgrupo), es necesario demostrar que estos efectos son independientes entre sí. Si el efecto continúa siendo estadísticamente significativo, una vez que se ajusta por el efecto en los otros subgrupos, es más probable que este corresponda a un efecto real. Para realizar estos análisis habitualmente se requiere de modelos matemáticos complejos, como los modelos de regresión<sup>22,24</sup>.

#### Contexto general

*Magnitud del efecto:* El tamaño del efecto debe ser de una magnitud importante, ya que mientras mayor es la magnitud también lo es la credibilidad del análisis de subgrupo.

*Consistencia:* La credibilidad del análisis de subgrupo es mayor si el efecto en un subgrupo es replicado en otros estudios de diseño similar. La replicación y consistencia de los resultados de un análisis de subgrupo en múltiples estudios es la mayor evidencia de que esa diferencia es real y supera en credibilidad a cualquier prueba estadística.

*Comportamiento en outcomes relacionados:* Si un análisis de subgrupo muestra un efecto consistente en *outcomes* con fisiopatología similar o común, el efecto es más creíble. Por ejemplo, en la actualidad sabemos que la aspirina es efectiva en profilaxis secundaria de infarto agudo al miocardio y accidente vascular isquémico en toda la población, independiente del sexo. Un estudio publicado a fines de la década del '70 mostró en un análisis de subgrupo que la aspirina era efectiva en prevenir infartos al miocardio en mujeres, pero resultaba inefectiva como profilaxis secundaria de accidente vascular isquémico<sup>26</sup>. Este efecto de subgrupo no fue confirmado por estudios posteriores<sup>27</sup>, pero *a priori* es poco creíble, ya que la fisiopatología de ambos *outcomes* es similar. Un efecto de subgrupo tiene mayor credibilidad si el efecto es consistente en *outcomes* fisiopatológicamente relacionados.

*Plausibilidad biológica:* Mayor es la credibilidad si el efecto de subgrupo tiene un fundamento indirecto que lo apoya, especialmente la racionalidad biológica o fisiopatológica. Los análisis de subgrupo que muestran efectos que no son consistentes con el conocimiento biológico existente, con frecuencia son irreales. Por ejemplo, hallazgos tales como que la aspirina o la terapia antihipertensiva no son efectivos en mujeres<sup>26,27</sup> y ancianos<sup>28</sup> o que el beta bloqueo no es efectivo en infartos al miocardio de pared inferior<sup>29</sup>, no tienen fundamento biológico y fueron posteriormente refutados por estudios de mejor diseño<sup>2,6,21,22,23</sup>.

#### Conclusiones

Un análisis de subgrupo puede entregarnos información valiosa al momento de aplicar la evidencia a grupos específicos de pacientes, sin embargo, existe un riesgo importante de que nos lleve a conclusiones erróneas. Los criterios mencionados pueden ser utilizados como una guía para su interpretación: mientras más criterios de cumplan, más probable es que el efecto de subgrupo sea real. A la hora de aplicar la evidencia proveniente de análisis de subgrupos, los clínicos y pacientes debieran balancear la confianza en el efecto de subgrupo con los potenciales beneficios y riesgos de la intervención.

#### Referencias

1. Guyatt G, Rennie D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chapter 1A. First Edition. 2002.
2. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005; 365 (9454): 176-86.
3. Isaacs D, Fitzgerald D. Seven alternatives to evidence based medicine. *BMJ* 1999; 319 (7225): 1618.
4. Letelier LM, Moore P. [Evidence based medicine: a view after a decade]. *Rev Med Chile* 2003; 131 (8): 939-46.
5. Pantoja T, Letelier LM, Neumann I. [A critical appraisal of clinical reports in medical journals]. *Rev Med Chile* 2004; 132 (4): 513-5.
6. Guillemin F. Primer: the fallacy of subgroup analysis. *Nat Clin Pract Rheumatol* 2007; 3 (7): 407-13.
7. Rada GG, Andrade AM. [Should the results of this study be applied to my patient?]. *Rev Med Chile* 2006; 134 (1): 115-9.
8. North American Symptomatic Carotid Endarterectomy

- Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. North American Symptomatic Carotid Endarterectomy Trial Collaborators. *N Engl J Med* 1991; 325 (7): 445-53.
9. European Carotid Surgery Trialists' Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet* 1998; 351 (9113): 1379-87.
  10. Rerkasem K, Rothwell PM. Carotid endarterectomy for symptomatic carotid stenosis. *Cochrane Database Syst Rev* 2011; (4): CD001081
  11. D'agostino RB SR, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; 117 (6): 743-53.
  12. Taylor F, Ward K, Moore TH, Burke M, Davey Smith G, Casas JP, et al. Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 2011; (1): CD004816.
  13. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy--I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 1994; 308 (6921): 81-106.
  14. Gueyffier F, Boutitie F, Boissel JP, Pocock S, Coope J, Cutler J, et al. Effect of antihypertensive drug treatment on cardiovascular *outcomes* in women and men. A meta-analysis of individual patient data from randomized, controlled trials. The INDANA Investigators. *Ann Intern Med* 1997; 126 (10): 761-7.
  15. Gueyffier F, Bulpitt C, Boissel JP, Schron E, Ekblom T, Fagard R, et al. Antihypertensive drugs in very old people: a subgroup meta-analysis of randomised controlled trials. INDANA Group. *Lancet* 1999; 353 (9155): 793-6.
  16. Flather MD, Yusuf S, Køber L, Pfeffer M, Hall A, Murray G, et al. Long-term ACE-inhibitor therapy in patients with heart failure or left-ventricular dysfunction: a systematic overview of data from individual patients. ACE-Inhibitor Myocardial Infarction Collaborative Group. *Lancet* 2000; 355 (9215): 1575-81.
  17. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985; 27 (5): 335-71.
  18. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; 2 (8607): 349-60.
  19. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer. *Cochrane Database Syst Rev* 2001; (1): CD000486.
  20. Wijeyesundera HC, Hansen MS, Stanton E, Cropp AS, Hall C, Dhalla NS, et al; Praise II Investigators. Neurohormones and oxidative stress in nonischemic cardiomyopathy: relationship to survival and the effect of treatment with amlodipine. *Am Heart J* 2003; 146 (2): 291-7.
  21. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116 (1): 78-84.
  22. Sun X, Briel M, Walter S D, Guyatt G. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *BMJ* 2010; 340: 850-54.
  23. Brookes ST, Whitley E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004; 57 (3): 229-36.
  24. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001; 5 (33): 1-56.
  25. Russell JA, Walley KR, Singer J, Gordon AC, Hébert PC, Cooper Dj, et al; Vast Investigators. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med* 2008; 358 (9): 877-87.
  26. The Canadian Cooperative Study Group. A randomized trial of aspirin and sulfinpyrazone in threatened stroke. *N Engl J Med* 1978; 299 (2): 53-9.
  27. Fields WS, Lemak NA, Frankowski RF, Hardy RJ. Controlled trial of aspirin in cerebral ischemia. *Stroke* 1977; 8 (3): 301-14.
  28. Amery A, Birkenhäger W, Brixko P, Bulpitt C, Clement D, De Leeuw P, et al. Influence of antihypertensive drug treatment on morbidity and mortality in patients over the age of 60 years. European Working Party on High blood pressure in the Elderly (EWPHE) results: subgroup analysis on entry stratification. *J Hypertens Suppl* 1986; 4 (6): S642-7.
  29. Andersen MP, Bechsgaard P, Frederiksen J, Hansen DA, Jørgensen HJ, Nielsen B, et al. Effect of alprenolol on mortality among patients with definite or suspected acute myocardial infarction. Preliminary results. *Lancet* 1979; 2 (8148): 865-8.
  30. Anastos K, Charney P, Charon RA, Cohen E, Jones CY, Marte C, et al. Hypertension in women: what is really known? The Women's Caucus, Working Group on Women's Health of the Society of General Internal Medicine. *Ann Intern Med* 1991; 115 (4): 287-93.

31. Medical Research Council Working Party. MRC trial of treatment of mild hypertension: principal results. *Br Med J (Clin Res Ed)* 1985; 291 (6488): 97-104.
32. Cleland JG, Bulpitt CJ, Falk RH, Findlay IN, Oakley CM, Murray G, et al. Is aspirin safe for patients with heart failure? *Br Heart J* 1995; 74 (3): 215-9.
33. Reduction in mortality after myocardial infarction with long-term beta-adrenoceptor blockade. Multicentre international study: supplementary report. *Br Med J* 1977; 2 (6084): 419-21.
34. Gruppo Italiano Per Lo Studio Della Streptochinasi Nell'infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; 1 (8478): 397-402.
35. Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. *N Engl J Med* 1988; 319 (26): 1681-92.
36. Packer M, O'connor CM, Ghali JK, Pressler ML, Carson PE, Belkin RN, et al. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. Prospective Randomized Amlodipine Survival Evaluation Study Group. *N Engl J Med* 1996; 335 (15): 1107-14.