

## Análisis crítico de ensayos clínicos randomizados publicados en revistas biomédicas chilenas

Juan Manríquez M<sup>1a</sup>, Gonzalo Valdivia C<sup>2</sup>,  
Gabriel Rada G<sup>1b</sup>, Luz María Letelier S<sup>1,3,4</sup>.

### *Critical assessment of randomized controlled trials published in biomedical Chilean journals*

**Background:** Well designed clinical trials yield the strongest evidence for the effect of health care interventions. **Aim:** To assess the methodological quality of the design and report of randomized clinical trials in a sample, published in biomedical Chilean journals between 1980 and 2002. **Material and methods:** All trials identified by hand search by the Unit of Evaluation of Technologies in Health, were assessed for quality of randomization, blinding, analysis of results and other characteristics of trial design, along with the application of Jadad's Score, that assesses the methodological quality of clinical trials in a scale that ranges from 0 to 5. **Results:** Twenty eight trials were found and assessed, 75% (n=21) specified the method used for randomization, 29% (n=8) described a correct allocation concealment and 39% (n=11) were double blinded. Withdrawals and dropouts were correctly reported in 21% (n=6) of the articles, whereas intention to treat analysis was done only in one. Thirteen trials had a Jadad score equal or higher than 3 points. **Conclusions:** Several design deficiencies were found in the trials assessed. It is difficult to know if methodological weaknesses are due to incomplete reports or to methodologically poor designs. Adopting initiatives like the CONSORT can help improve the quality of randomized clinical trials published in Biomedical Chilean journals (Rev Méd Chile 2005; 133: 439-46).

**(Key Words:** Methods; Periodicals; Publications; Randomized controlled trials)

Recibido el 19 de agosto, 2004. Aceptado en versión corregida el 16 de diciembre, 2004.

<sup>1</sup>Unidad de Medicina Basada en Evidencia (UMBE-UC).

<sup>2</sup>Departamento de Salud Pública, Pontificia Universidad Católica de Chile.

<sup>3</sup>Departamento de Medicina Interna, Pontificia Universidad Católica de Chile.

<sup>4</sup>Servicio de Medicina, Hospital Sótero del Río. Santiago de Chile

<sup>a</sup>Interno de Medicina

<sup>b</sup>Residente de Medicina Interna

El ensayo clínico randomizado (ECR) representa la forma más rigurosa de establecer una relación causal entre una intervención sanitaria y su efecto<sup>1-5</sup>.

Las características fundamentales en el diseño de un ECR son: 1. La asignación aleatoria de los sujetos participantes y el adecuado ocultamiento de la secuencia de randomización. 2. El ciego de uno, de varios o de todos los participantes del ECR. 3. El reporte de las pérdidas y abandonos en cada grupo, junto con sus causas. 4. El análisis de los datos según el principio de intención de tratar<sup>2,6-8</sup>.

*Correspondencia a:* Dra. Luz María Letelier S. Departamento de Medicina Interna, Escuela de Medicina, Universidad Católica de Chile. Lira N° 44, Santiago. Fono: 3543030. Fax: 3543030. E mail: lmletel@med.puc.cl

La validez de un ECR representa la probabilidad de que éste se encuentre libre de sesgo, siendo la validez interna la medida en que los resultados obtenidos en un ECR se aproximan a la verdad buscada. La validez externa representa en qué medida los resultados de un ECR pueden ser aplicados a poblaciones similares a las del estudio<sup>1-6</sup>.

Las fallas metodológicas en el diseño de un ECR están directamente relacionadas con la probabilidad de obtener resultados sesgados, lo que afecta tanto la validez interna como la externa. El reporte completo al momento de publicar los resultados, permite al lector del artículo efectuar un análisis crítico de los resultados comunicados y decidir si éstos pueden ser incorporados en la práctica clínica<sup>2-10</sup>.

El objetivo del presente trabajo es evaluar la calidad metodológica del diseño y reporte de los ECR publicados en revistas biomédicas chilenas.

#### MÉTODOS

Este es un estudio observacional retrospectivo que pretende describir las características metodológicas de los ECRs publicados en revistas biomédicas chilenas. Para ello se utilizó la base de datos de la Unidad de Evaluación de Tecnologías en Salud, dependiente del Ministerio de Salud, donde se llevó a cabo una búsqueda manual retrospectiva de ensayos clínicos controlados, a partir de un listado de revistas biomédicas chilenas publicadas entre los años 1980-2002 (Tabla 1), identificándose un total de 324 artículos<sup>11</sup>.

De esta base de datos seleccionamos, para el presente estudio, todos aquellos estudios catalogados como ECRs, en base a si su proceso de asignación de intervenciones se realizó mediante un procedimiento aleatorio.

Para determinar la calidad de los ECRs establecimos cinco ítemes principales, cada uno de ellos con más de un criterio de evaluación explícito. Para los criterios establecidos en cada ítem se asignó la categoría "sí" (presente) o "no" (ausente) según lo descrito en el reporte. Los criterios que no estaban claramente explicitados en el reporte fueron catalogados como ausentes.

Los criterios de evaluación fueron aplicados por el mismo investigador (JMM) utilizando sólo la información reportada en la publicación.

En el primer ítem se evaluaron características generales, como la declaración de los criterios de inclusión y exclusión y el lugar de realización del ensayo. Información sobre el tipo y forma de aplicación de la(s) intervención(es) y la existencia de cointervención(es). Se evaluó la descripción de los *eventos observados* primarios y secundarios, además de la declaración sobre el método utilizado para medir la ocurrencia de éstos. Se buscó información referente a las características basales de los grupos y se determinó si ellos tenían características similares al inicio del estudio. De no ser así, se evaluó si se aplicó algún método de ajuste estadístico. También evaluamos si se realizó cálculo de tamaño muestral al inicio del estudio.

En el segundo ítem, evaluamos características del proceso de randomización: descripción del método para generar la secuencia y la validez de este proceso. Finalmente valoramos la descripción de algún método para ocultar la secuencia de randomización.

En el tercer ítem evaluamos las características del ciego. Establecimos si los autores describieron el ensayo como ciego único, doble o triple, además de conocer quienes fueron objeto de ciego. En los ECRs ciegos, evaluamos si se reportaron los métodos para lograrlo y mantenerlo. Finalmente evaluamos si los autores reportaron

**Tabla 1. Revistas en las que se realizó búsqueda manual de ECR entre 1980 y 2002**

Publicación	Ensayos clínicos randomizados (n)
Revista Médica de Chile	18
Revista Chilena de Pediatría	7
Boletín Chileno de Parasitología Cardiología Clínica	
Boletín de Anestesiología	1
Pediatría/Revista Pediatría	
Revista Chilena de Salud Pública	
Revista Chilena de Fonoaudiología	
Revista Chilena de Obstetricia y Ginecología	2
Revista Chilena de Infectología	
Revista de Psiquiatría	
Boletín Hospital San Juan de Dios	
Enfermedades Respiratorias y Cirugía Torácica	1
Archivos Chilenos de Oftalmología	2
Revista Chilena de Cardiología	1

fallas en la mantención del ciego y si se evaluó el éxito de éste.

En el cuarto ítem analizamos características de la presentación de resultados, evaluando el reporte de las pérdidas y abandonos en cada grupo del estudio y los motivos de éstos. También evaluamos si el análisis de los resultados se realizó según el principio de intención de tratar, considerando este aspecto como presente, cuando fue explícito y si el denominador empleado en los cálculos correspondía a los pacientes que iniciaron el estudio.

Además aplicamos a cada ECR la escala de Jadad<sup>12</sup>. Ésta consiste en una escala validada utilizada para evaluar la calidad metodológica de ECRs incluidos en revisiones sistemáticas. Tiene un puntaje mínimo de 0 y un máximo de 5 puntos. Se otorga 1 punto en el caso que el ECR se describa como randomizado, 1 punto si cumple con el requisito de ser doble ciego y 1 punto si se detallan adecuadamente las pérdidas y abandonos en cada grupo, junto con sus causas. Se otorga 1 punto adicional si el proceso de randomización y el ciego fueron válidos, en caso contrario se descuenta 1 punto respectivamente.

Se consideraron válidos los siguientes métodos de randomización: tablas, sorteo o computacionales. En caso del ciego son válidos aquellos que

garanticen que ni quienes evalúan la ocurrencia de eventos, ni los sujetos participantes en el estudio, puedan identificar su pertenencia al grupo experimental o al grupo control. Si este aspecto no se menciona, consideramos válida la utilización de placebos idénticos a la intervención, placebos activos o la utilización de método de doble simulación.

Sólo con fines descriptivos se agruparon los estudios en 2 períodos arbitrarios, según la fecha de publicación: 1980 a 1992 y 1993 a 2002.

## RESULTADOS

De un total de 32 ECRs encontrados, se analizaron 28, con un tamaño muestral promedio de 73 sujetos por ensayo (rango 17 a 435). Cuatro artículos fueron excluidos de nuestro análisis por no contar con el reporte completo de la metodología utilizada (1 artículo editorial y 3 resúmenes de presentaciones en congresos). En el período de 1980-1992 se evaluaron 12 artículos y entre 1993 y 2002 se evaluaron 16 artículos, lo que se podría interpretar como una tendencia al aumento de este tipo de estudios.

En las Tablas 2, 3, 4 y 5 se resumen los resultados para cada uno de los ítemes descritos.

**Tabla 2. Resultados del ítem 1 en relación a características generales del ensayo. En los períodos 1980-1992 y 1993-2002 se muestra el número de artículos que describieron el criterio evaluado. Los trabajos que detallaron el criterio en evaluación se muestran, para ambos períodos, en número y porcentaje**

Aspectos evaluados	Períodos				Ambos períodos			
	(1980-1992)		(1993-2002)		(1980-2002)			
	Número de artículos que cumplen requisitos				Número (%)			
	Sí	No	Sí	No	Sí	(%)	No	(%)
Criterios de inclusión	11	1	16	0	27	(96%)	1	(4%)
Lugar de realización del estudio	11	1	16	0	27	(96%)	1	(4%)
Intervención asignada	12	0	16	0	28	(100%)	0	(0%)
Aplicación de la intervención	12	0	16	0	28	(100%)	0	(0%)
Eventos observados primario(s) y secundario(s)	8	4	10	6	18	(64%)	10	(36%)
Método de evaluación de eventos observados	11	1	16	0	27	(96%)	1	(4%)
Cálculo del tamaño muestral	0	12	6	10	6	(21%)	22	(79%)
Grupos comparables al inicio del estudio	10	2	16	0	26	(89%)	2	(8%)
Detalle de características demográficas	11	1	15	1	26	(89%)	2	(7%)
Ausencia de cointervención	12	0	16	0	28	(100%)	0	(0%)

**Tabla 3. Resultados referidos a la calidad de la randomización. En los períodos 1980-1992 y 1993-2002 se muestra el número de artículos que describieron el criterio evaluado. Los trabajos que detallaron el criterio evaluado se muestran, para ambos períodos, en número y porcentaje**

Aspectos evaluados	Períodos				Ambos períodos			
	(1980-1992)		(1993-2002)		(1980-2002)			
	Número de artículos que cumplen requisitos							
	Sí	No	Sí	No	Sí	(%)	No	(%)
Descripción del método de randomización	7	5	14	2	21	(75%)	7	(25%)
Ocultamiento de la secuencia de randomización	2	10	6	10	8	(29%)	20	(71%)

**Tabla 4. Evaluación del ciego en los artículos analizados. En los períodos 1980-1992 y 1993-2002 se muestra el número de artículos que describieron el criterio evaluado. Los trabajos que detallaron el criterio en evaluación se muestran, para ambos períodos, en número y porcentaje**

Aspectos evaluados	Períodos				Ambos períodos			
	(1980-1992)		(1993-2002)		(1980-2002)			
	Número de artículos que cumplen requisitos							
	Sí	No	Sí	No	Sí	(%)	No	(%)
Doble ciego	4	8	7	9	11	(39%)	17	(61%)
Quiénes fueron ciegos:								
Individuos participantes	4	0	6	1	10	(91%)	1	(9%)
Clínicos tratantes	4	0	6	1	10	(91%)	1	(9%)
Adjudicadores de eventos	4	0	7	1	11	(91%)	1	(9%)*
Analistas de datos								
Métodos para mantener ciego	4	0	7	1	11	(91%)	1	(9%)*
Evaluación del éxito del ciego	0	4	0	8	0	(0%)	12	(100%)*
Reporte de fallas en el ciego	0	4	0	8	0	(0%)	12	(100%)*

\*Hubo un ECR no descrito como doble ciego, sin embargo la adjudicación de eventos fue enmascarada.

**Tabla 5. Evaluación del análisis de los resultados. Los trabajos que detallaron el criterio en evaluación se muestran, para ambos períodos, en número y porcentaje**

Aspectos evaluados	Períodos				Ambos períodos			
	(1980-1992)		(1993-2002)		(1980-2002)			
	Número de artículos que cumplen requisitos							
	Sí	No	Sí	No	Sí	(%)	No	(%)
Pérdidas y abandonos	3	9	3	13	6	(21%)	22	(79%)
Descrito como con intención de tratar	0	12	2	14	2	(7%)	26	(93%)
Intención de tratar según criterios establecidos	0	12	1	15	1	(4%)	27	(96%)

En relación al puntaje de Jadad, en el período 1980-1992, 4 de 12 ECR (33,3%) obtuvieron puntaje mayor o igual a tres. En el período 1993-2002 9 de 16 ECRs (56,3%) obtuvieron tres o más puntos. Esto podría traducir una tendencia a una mejor calidad de diseño en las publicaciones más recientes.

## DISCUSIÓN

A continuación discutimos nuestros hallazgos, según los ítemes propuestos para el análisis.

*Características generales del estudio.* La elegibilidad de los pacientes, el lugar de realización del ensayo y la información sobre las intervenciones en estudio, son esenciales para reproducir la experiencia en la práctica diaria. Pese a la importancia de estos puntos, a menudo no son adecuadamente reportados. En una revisión de ECRs en cirugía, sólo 25% de 364 ECRs lo reporta<sup>13</sup>. En nuestro análisis, encontramos que los criterios de inclusión y el lugar de realización del estudio se describió en el 96% (n=27) del total de artículos, mientras que la intervención realizada y su modalidad de aplicación se reportan en el 100% (n=28).

El *evento observado* primario queda claramente establecido sólo en 27% de 48 artículos analizados, según un reporte publicado en 1985<sup>1</sup>. En nuestro análisis el *evento observado* primario queda establecido en 64% de los trabajos (n=18), estando en la mayoría de ellos (96%) claramente explicitados los métodos para evaluarlos.

El 89% de los artículos (n=26) informó claramente sobre las características demográficas. Sólo un artículo mostró diferencias basales entre los grupos, por lo que se realizó ajuste estadístico. En una publicación no pudimos definir la similitud entre los grupos por la ausencia de esta información. No encontramos cointervenciones declaradas en nuestro análisis.

El cálculo de tamaño muestral se reportó sólo en 21% (n=6). Esto podría explicar el alto número de ECRs con hallazgos negativos o ausencia de efecto (bajo poder del estudio con alto riesgo de error  $\beta$ ). Se ha reportado que sólo 43% de 37 ECR con resultados negativos reportan algún método para calcular tamaño muestral<sup>3</sup>. Varias revisiones han mostrado que es infrecuente que los autores indiquen cómo se determinó el tamaño muestral<sup>6-9</sup>.

*Randomización.* La asignación aleatoria (cuando se realiza con un tamaño muestral adecuado) equilibra los grupos respecto de los factores pronósticos que puedan determinar mayor o menor riesgo de presentar el *evento observado*, sean éstos conocidos o desconocidos.

Se consideran válidos aquellos métodos de randomización que aseguran que cualquier paciente tiene igual probabilidad de ingresar a cualquiera de los grupos. Esto se logra utilizando secuencias generadas computacionalmente, tablas de randomización o sorteo. Métodos como ingreso de sujetos con alternación, fecha de nacimiento u otros, son considerados cuasi aleatorios y se ha demostrado que este tipo de estudios pueden entregar resultados sesgados<sup>3,14,15</sup>.

Sólo 32% de los artículos publicados en revistas de especialidades y 48% en revistas de medicina general<sup>16-20</sup> especifican el método utilizado para generar la secuencia de randomización. En nuestro análisis el 75% (n=21) de los autores describieron la forma en que generaron el orden de asignación.

Se ha determinado que los ensayos no aleatorizados pueden sobrestimar la magnitud del efecto de la intervención en 41% cuando el método de aleatorización es inadecuado y en 30% cuando el método no está claro<sup>10,21</sup>.

El ocultamiento de la secuencia de randomización impide que el investigador a cargo de incorporar sujetos a un ECR conozca la intervención asignada, de modo de no verse influenciado a aceptar o rechazar sujetos de un ensayo<sup>16</sup>.

Los métodos considerados apropiados para este fin son: uso de componente externo (como una farmacia o central generadora de códigos), contenedores o sobres opacos y sellados. Se ha determinado que sólo 11% de ECR sobre artritis reumatoide, 52% de los ECR en revistas de obstetricia y ginecología y 64% de los ECR publicados en revistas médicas generales, describen un mecanismo apropiado para ocultar la secuencia de randomización<sup>8,16,22</sup>. En nuestro análisis encontramos que 29% (n=8) describió algún método con este fin.

Se ha determinado que los ECR en que no se especifica un adecuado ocultamiento, pueden llegar a sobrestimar la medida del efecto en un rango que va entre 30 y 40%<sup>2,3,14,15</sup>.

*“Ciego”.* En un ECR se entiende por ciego el desconocimiento de la intervención asignada a los

grupos en estudio, ya sea por parte de uno, de varios o de todos los participantes en un ECR<sup>23-25</sup>.

Las principales ventajas del "ciego" son: prevenir el sesgo de información, mejorar la adherencia de los sujetos y disminuir las pérdidas en el seguimiento<sup>25-27</sup>.

Aquellos estudios que no son "ciegos" tienden a sobrestimar el efecto de la intervención en 17% comparado con ECRs "ciegos"<sup>10</sup>. También se han encontrado resultados de ECR totalmente distintos en cuanto a la dirección del efecto entre aquellos que fueron o no "ciegos" a la intervención<sup>28</sup>.

Para establecer la existencia del ciego en un ECR es necesario conocer quienes lo fueron durante el desarrollo del estudio, los métodos utilizados para lograrlo y el éxito o fracaso de éste<sup>29</sup>. Cincuenta y uno por ciento de 506 ECR sobre fibrosis quística, 22% de 196 ECR en artritis reumatoide y 38% de 68 ECR en dermatología, no informan si se incluyó el "ciego" en su metodología<sup>6,22,30</sup>. Por otro lado, sólo 45% de 31 ECR doble ciego en revistas de ginecología reportan similitud entre la intervención en el grupo experimental y control y sólo el 5% reporta si el ciego fue o no exitoso<sup>31</sup>.

En nuestro análisis, encontramos que 39% de los artículos evaluados (n=11) fueron descritos como doble ciego, y, de éstos, 91% (n=10) reportan características similares entre la intervención del grupo experimental y el grupo control. Un artículo no se describió como ciego, sin embargo la adjudicación del *evento observado* se realizó sin conocimiento de la asignación. Ninguno de los ECR evaluó el éxito del ciego ni reportó fallas en su mantención.

**Análisis de los resultados.** Los sujetos excluidos luego de la asignación pueden ser diferentes de quienes continuaron en el ensayo en cuanto a características vinculadas con la probabilidad de ocurrencia del evento evaluado<sup>31-33</sup>. Por ello se debe reportar y diferenciar adecuadamente las pérdidas en el seguimiento de aquellos sujetos que fueron excluidos del análisis por poca adherencia al protocolo. El número y las causas de las pérdidas o la exclusión de sujetos debiesen ser explícitos, de modo que el lector pueda evaluar la ocurrencia de sobre o subestimación de los resultados.

Esta información no siempre se encuentra disponible aun en revistas relevantes de medicina

interna<sup>34</sup>. En nuestro análisis encontramos que 53% (n=15) de los artículos reportó algún tipo de información sobre las pérdidas y abandonos, de ellos, 32% lo hizo en forma incompleta y sólo 21% (n=6) informó el número de pérdidas y abandonos en ambos grupos, junto con sus causas.

El análisis por intención de tratar se refiere a que los sujetos sean analizados en los grupos a los que inicialmente fueron asignados, independiente de si recibieron la intervención. Este tipo de análisis permite conservar los beneficios de la randomización<sup>1-10</sup>. Existen ensayos que muestran resultados beneficiosos de una intervención al analizar sólo los pacientes que completan o adhieren al protocolo, sin embargo, al analizar con intención de tratar, los beneficios no son tan evidentes<sup>35</sup>.

Sólo 119 de 249 ECRs publicados en 1997 en connotadas revistas de medicina general mencionan análisis por intención de tratar, sin embargo, sólo cinco de éstos establecieron que todos los pacientes randomizados fueron analizados según el grupo de asignación inicial<sup>36</sup>. Nosotros encontramos que sólo en 7% (n=2) de los artículos, los autores describen un análisis por intención de tratamiento. Sin embargo, sólo uno de ellos considera a todos los pacientes, independiente del cumplimiento del protocolo. En 72% (n=20) de los artículos no queda claro cuál es el tipo de análisis empleado.

**Puntaje de Jadad.** La escala de Jadad<sup>12</sup> evalúa, fundamentalmente, características de la randomización y del ciego, junto con la adecuada descripción de los abandonos y de las pérdidas. Según esta escala, un estudio es considerado de baja calidad si obtiene un puntaje menor o igual a dos, habiéndose demostrado que aquellos ECR que obtienen 2 puntos o menos pueden sobrestimar la magnitud del efecto hasta en 35%, comparado con aquellos ECR con 3 o más puntos de esta escala<sup>37</sup>. En nuestro análisis encontramos que 54% (n=15) de los artículos alcanzaron un puntaje de Jadad menor a 3 puntos (rango de 1 a 2), mientras que 46% (n=13) logró un puntaje mayor o igual a 3 (rango de 3 a 4).

Aunque es imposible diferenciar en qué medida las deficiencias metodológicas encontradas son consecuencia de reportes incompletos, o si obedecen a falencias reales en el diseño de los ECRs,

esperamos que el análisis presentado en este reporte contribuya a mejorar la calidad del diseño de los ECRs publicados en nuestro país en aquellas áreas identificadas como deficitarias. La adhesión a iniciativas como el CONSORT<sup>29</sup>, puede ayudar a mejorar la calidad de la ejecución, análisis y reporte en la publicación de ensayos clínicos. La incorporación de expertos estadísticos y metodólogos en la etapa de planificación del ECR podría ser fundamental para este fin.

Entre las limitaciones de nuestro trabajo encontramos el no evaluar todas las revistas biomédicas publicadas en el período analizado y la ausencia de un segundo revisor de los puntos

evaluados, aunque éstos correspondían a criterios objetivos, la mayoría sin segundas lecturas.

#### CONCLUSIÓN

Las publicaciones consideradas en la revisión cubren razonablemente la oferta nacional y representan ampliamente áreas disciplinarias en las cuales se investiga utilizando diseños experimentales y se publican con evaluación de pares.

Creemos que la información contenida en este trabajo puede constituir un aporte al mejoramiento de las publicaciones biomédicas en nuestro país.

#### REFERENCIAS

1. POCKOCK SJ, HUGHES MD, LEE RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987; 317: 426-32.
2. KUNZ R, VIST G, OXMAN AD. Randomization to protect against selection bias in healthcare trials (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 2, 2004. Chichester, UK: John Wiley & Sons, Ltd.
3. MOHER D. Consort: an evolving tool to help improve the quality of reports of randomized controlled trials. Consolidated Standards of Reporting Trials. *JAMA* 1998; 279: 1489-91.
4. MOHER D, DULBERG CS, WELLS GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; 272: 122-4.
5. FREIMAN JA, CHALMERS TC, SMITH H JR, KUEBLER RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978; 299: 690-4.
6. ADETUGBO K, WILLIAMS H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol* 2000; 136: 381-5.
7. GLAVAN B, SHEWAKRAMANI S, HOLLANDER J. Incomplete data reporting in studies of emergency department patients with potential acute coronary syndromes using troponins. *Acad Emerg Med* 2003; 10: 943-8.
8. ALTMAN DG, DORE CJ. Randomization and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-53.
9. THORNLEY B, ADAMS C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998; 317: 1181-4.
10. SCHULZ KF, CHALMERS I, HAYES RJ, ALTMAN DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408-12.
11. Unidad de Evaluación de Tecnologías de Salud. Ministerio de Salud de Chile. 2004. Disponible en: <http://www.minsal.cl/> [Consultado el 01 de junio de 2004].
12. JADAD AR, MOORE RA, CARROLL D, JENKINSON C, REYNOLDS DJ, GAVAGHAN DJ ET AL. Assessing the quality of reports on randomized clinical trials: Is blinding necessary? *Controlled Clin Trials* 1996; 17: 1-12.
13. HALL JC, MILLS B, NGUYEN H, HALL JL. Methodologic standards in surgical trials. *Surgery* 1996; 119: 466-72.
14. KJAERGARD L, VILLUMSEN J, GLUUD C. Quality of randomized clinical trials affects estimates of intervention efficacy. Abstracts for Workshops and Scientific Sessions, 7<sup>th</sup> *International Cochrane Colloquium*, Rome, Italy, 1999.
15. JUNI P, ALTMAN DG, EGGER M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42-6.
16. SCHULZ KF, CHALMERS I, GRIMES DA, ALTMAN DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994; 272: 125-8.

17. SCHULZ KF. Randomized trials, human nature, and reporting guidelines. *Lancet* 1996; 348: 596-8.
18. AH-SEE KW, MOLONY NC. A qualitative assessment of randomized controlled trials in otolaryngology. *J Laryngol Otol* 1998; 112: 460-3.
19. BATH FJ, OWEN VE, BATH PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke* 1998; 29: 2203-10.
20. ALTMAN DG, DORE CJ. Randomization and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-53.
21. CARROLL D, TRAMER M, MCQUAY H, NYE B, MOORE A. Randomization is important in studies with pain outcomes: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain. *Br J Anaesth* 1996; 77: 798-803.
22. GOTZSCHE PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989; 10: 31-56.
23. DAY S, ALTMAN D. Statistics notes: blinding in clinical trials and other studies. *BMJ* 2000; 321: 504.
24. DEVEREAUX PJ, MANNIS BJ, GHALI WA ET AL. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001; 285: 2000-3.
25. SCHULZ K, CHALMERS I, ALTMAN D. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002; 136: 254-9.
26. BANG H, NI L, DAVIS C. Assessment of blinding in clinical trials. *Control Clin Trials* 2004; 25: 143-56.
27. SCHULZ K, GRIMES D. Blinding in randomized trials: hiding who got what. *Lancet* 2002; 359: 696-700.
28. ERNST E, WHITE AR. Acupuncture for back pain: A meta-analysis of randomized controlled trials. *Arch Int Med* 1998; 158: 2235-41.
29. ALTMAN DG, SCHULZ KF, MOHER D, EGGER M, DAVIDOFF F, ELBOURNE D ET AL. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663-94.
30. CHENG K, SMYTH RL, MOTLEY J, O'HEA U, ASHBY D. Randomized controlled trials in cystic fibrosis (1966-1997) categorized by time, design, and intervention. *Pediatr Pulmonol* 2000; 29: 1-7.
31. SCHULZ KF, GRIMES DA, ALTMAN DG, HAYES RJ. Blinding and exclusions after allocation in randomized controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996; 312: 742-4.
32. SCHULZ KF. Randomized controlled trials. *Clin Obstet Gynecol* 1998; 41: 245-56.
33. SACKETT DL, GENT M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979; 301: 1410-2.
34. EGGER M, JUNI P, BARTLETT C. Value of flow diagrams in reports of randomized controlled trials. The CONSORT Group. *JAMA* 2001; 285: 1996-9.
35. FIELDS WS, MASLENIKOV V, MEYER JS, HASS WK, REMINGTON RD, MACDONALD M. Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient cerebral ischemic attacks and cervical carotid artery lesions. *JAMA* 1970; 211: 1993-2003.
36. HOLLIS S, CAMPBELL F. What is meant by intention to treat analysis? Survey of published randomized controlled trials. *BMJ* 1999; 319: 670-4.
37. MOHER D, PHAM B, JONES A, COOK DJ, JADAD AR, MOHER M ET AL. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352: 609-13.

*Agradecimientos:*

A los miembros de la Unidad de Medicina Basada en Evidencias (UMBE) de la Escuela de Medicina de la Pontificia Universidad Católica por sus aportes en la revisión crítica del artículo.