



UNIVERSITAT DE
BARCELONA

Estimación de sesgos de publicación en las revistas *Atención Primaria* y *Medicina Clínica*

Carlos Campillo Artero



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – SenseObraDerivada 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – SinObraDerivada 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-NoDerivs 4.0. Spain License.**

ESTIMACIÓN DE SESGOS DE PUBLICACIÓN
EN LAS REVISTAS *Atención Primaria Y Medicina Clínica*

Carlos Campillo Artero
Departamento de Biología Celular y Anatomía Patológica
Facultad de Medicina
Universidad Central de Barcelona
División de Ciencias de la Salud

Washington, DC, 22 de marzo de 1997



AGRADECIMIENTOS

Para concebir y llevar a cabo esta memoria de investigación, he contado con el estímulo y el apoyo de varias personas y amigos a los que quiero dejar constancia de mi agradecimiento. Del Profesor Antoni Caralps, mentor y amigo, no sólo he recibido sus valiosos consejos y su apoyo incondicional como Director. Desde 1981 he gozado del privilegio de sus enseñanzas, de su amistad y de su gran talante humano. El Profesor Jaume Canela tampoco se ha limitado a desempeñar su función de Codirector. Durante los últimos años, ha sido una de las voces que más me han ayudado a mantener este proyecto en marcha. La ayuda de ambos ha sido especialmente singular, pues ha superado el escollo que representan las más de seis mil millas que nos separan. Los Profesores Charles Rohdie, Catedrático de Bioestadística, y Moyses Szklo, Profesor de Epidemiología, de la Escuela de Higiene y Salud Pública de la Universidad Johns Hopkins, me enseñaron la forma de plantear y analizar investigaciones como ésta, así como a disfrutar con ellas. Al Dr. José María Corrales, mi amigo y psiquiatra, o mi psiquiatra y amigo, debo agradecerle sus compulsivamente pormenorizados y atinados comentarios a borradores de esta memoria. Además de agradecerle al Dr. Luis Carlos Silva, Investigador Titular del Instituto Superior de Ciencias Médicas de La Habana, el alto nivel crítico de sus comentarios metodológicos, quiero destacar la honestidad inusitada de sus comentarios y sus prontas respuestas a mis preguntas, a pesar de la distancia que nos separa y de las dificultades de comunicación que nos impuso el bloqueo. El Dr. Armando Peruga, amigo y Epidemiólogo de la Organización Panamericana de la Salud, me ayudó a programar con SAS los análisis estadísticos que quise realizar y me telefoneó día tras día para recordarme, siempre con su mordaz ironía, que debía aprestarme a terminar esta memoria de investigación. Por último, mi mujer, Mariasun Contrasta, logró que la evaluación de los 200 artículos fuese ciega, tabuló los datos, introdujo correcciones en los diversos borradores y me animó sin respiro y con cariño en cada minuto que pasé trabajando en esta memoria.

ÍNDICE

1.	ANTECEDENTES CIENTÍFICOS	1
	Prolegómeno.....	2
	Primeros antecedentes históricos	4
	El desprecio de la hipótesis nula	7
	Definición y confirmación del sesgo de publicación	10
	El problema del archivero	12
	Los estudios negativos.....	15
	Primeros modelos matemáticos para estimar el sesgo de publicación	22
	El sesgo de publicación y la financiación de los estudios.....	25
	Nuevos diseños para estimar la existencia del sesgo de publicación. Los registros de estudios y el metanálisis	27
	Nuevos diseños para investigar el sesgo de publicación.....	35
	Factores de riesgo del sesgo de publicación.....	47
	El enfoque de riesgo	50
	Fuente de financiación como factor de riesgo y conflictos de intereses	55
	Avances recientes en la investigación del sesgo de publicación.....	56
	Recapitulación.....	64
2.	OBJETIVOS E HIPÓTESIS	65
	Objetivos.....	65
	Hipótesis.....	66
3.	MATERIALES Y MÉTODOS	68
	Materiales	68
	Búsqueda bibliográfica.....	68
	Criterios de inclusión y exclusión	68
	Selección de la muestra	69
	Métodos	70
	Diseño	70
	Variables estudiadas	72
	Recategorización de variables	75
	Medición de la calidad de un estudio	76
	Evaluación de los artículos.....	80
	Tabulación de datos.....	82
	Análisis estadísticos	82
	Análisis univariantes	83
	Análisis bivariantes	83
	Análisis multivariantes	83
	Programas estadísticos	85

4.	RESULTADOS	86
	Análisis uni y bivariantes	86
	Estado de publicación.....	86
	Calidad de los artículos	86
	Calidad de las secciones.....	89
	Calidad de los subapartados	89
	Calidad y estado de publicación.....	97
	Procedencia de los artículos	98
	Procedencia y calidad.....	100
	Tamaño muestral.....	101
	Tamaño muestral y calidad.....	102
	Objeto de estudio.....	103
	Objeto de estudio y calidad	106
	Tipo de estudio	107
	Tipo de estudio y calidad.....	112
	Complejidad del análisis	115
	Complejidad y calidad.....	117
	Significación estadística.....	118
	Análisis multivariante.....	120
5.	DISCUSIÓN	124
	Comentarios generales	124
	Estado de publicación.....	126
	Calidad de los artículos	126
	Calidad de las secciones.....	128
	Calidad de los subapartados	129
	Procedencia de los artículos	139
	Tamaño de la muestra	141
	Objeto de estudio.....	142
	Tipo de estudio	146
	Complejidad	151
	Significación estadística.....	154
	Análisis multivariante.....	158
	Limitaciones del estudio.....	160
6.	CONCLUSIONES	165
7.	REFERENCIAS	169
8.	RESUMEN	186

CAPÍTULO 1. ANTECEDENTES CIENTÍFICOS

1.1 Prolegómeno

Los artículos científicos publicados en revistas biomédicas constituyen una de las principales fuentes de información de los profesionales de la salud. Cabe esperar, por tanto, que buena parte de la idoneidad de sus decisiones y de la calidad de la atención médica que prestan dependan de la validez de la información notificada en los artículos científicos y de su capacidad para evaluarlos críticamente.

Uno puede formularse, lícitamente, el siguiente interrogante: ¿refleja fielmente la literatura médica la información rigurosa y válida disponible en una determinada disciplina? Las reducidas dimensiones del espacio editorial no pueden dar cabida a todos los estudios que se llevan a cabo en las ciencias de la salud. Para decidir qué datos alcanzan las páginas impresas, las revistas biomédicas disponen de un sistema de revisión editorial cuyo supuesto cometido es tamizar la información considerada válida y de <<calidad>>¹ conforme al paradigma vigente de la ciencia (Khun T, 1970). Teóricamente, los estudios aceptables desde el punto de vista científico se deberían publicar y los inaceptables, rechazar. Por consiguiente, la calidad de un estudio debe ser el criterio cardinal que gobierne la publicación de estudios.

¿Es correcto pensar que sucede así? En esta memoria de investigación se intenta averiguar si la calidad prevalece como criterio rector del destino editorial de una investigación. Además, se pretende estimar por primera vez en España si en dos de las revistas españolas de mayor difusión nacional --*Atención Primaria y Medicina Clínica*-- las

¹ En esta memoria de investigación, la *calidad* de un estudio hace referencia a su pertinencia, oportunidad, originalidad, rigor metodológico, idoneidad de la presentación de datos y contribución al conocimiento.

opiniones y el proceder de investigadores, editores y revisores² introducen sesgos desde la concepción de una investigación hasta su eventual publicación.

¿Acaso la subjetividad, entreverada con el rigor científico, desempeña algún papel en la valoración de la calidad de un estudio? Un investigador puede optar por no publicar un estudio al no haber encontrado una asociación sospechada o diferencias entre variables analizadas. Es posible que frente a un resultado controvertido los lectores se decanten por aceptar, a veces de modo inopinado, la postura que sostienen ciertos autores simplemente porque gozan de renombre. Muchas revisiones de un mismo tema han arrojado conclusiones contrapuestas. En algunos casos, al ahondar en ellas comprobamos que sus autores se han limitado a yuxtaponer y transcribir hallazgos de otros estudios sin ponderar sus conclusiones a la luz de sus limitaciones metodológicas. Un ingente número de artículos publicados en revistas prestigiosas habían sido rechazados por otras que son objeto de tanto o mayor reconocimiento (Katerndahl DA, 1991; Landi G, Ciccone A, 1993; Preston NW, 1993; de Melker HE, Rosendaal FR, Vandenbroucke JP, 1993; Chalmers I, 1993; Moher D, 1993; Ernst E, Resch KL, 1994; Houston P, 1994; Ben-Shlomo Y, 1994; Vandenbroucke JP, Rosendaal FR, 1994).

Por todo ello cabe preguntarse: ¿conforman los estudios publicados una muestra representativa del universo de estudios de calidad realizados hasta la fecha? ¿Es posible que los juicios y las acciones de autores, revisores y editores puedan menoscabar dicha representatividad? El interés en desentrañar los sesgos que pueden incorporarse en todas las etapas de la publicación de un estudio y en la interpretación de los estudios publicados

² En adelante, los especialistas que revisan la calidad científica y la idoneidad de los artículos enviados a una revista, y asesoran a los editores en materia de publicación, se denominarán *revisores*. Este término intenta ser una traducción, más o menos afortunada, de su homónimo inglés *referee* o *peer*, que también se ha traducido al castellano como *árbitro*, *experto* o *par*, y no debe confundirse con el de corrector de estilo de la revista (*copyeditor*).

estriba en el principio de replicabilidad³ de los resultados, un criterio de inferencia causal. Buena parte de las investigaciones epidemiológicas pueden contemplarse como experimentos aleatorios encaminados a contrastar determinada hipótesis (Szklo M, 1991).

En esta memoria de investigación se revisarán estudios a fin de constatar la tendencia de autores, editores y revisores, respectivamente, a enviar a publicación, aceptar y emitir evaluaciones más favorables de los artículos en que se notifican diferencias estadísticamente significativas (en adelante, *estudios positivos*) en detrimento de aquellos en los que éstas no se han detectado (en adelante, *estudios negativos*). Dicha tendencia viola el principio de replicabilidad: los artículos publicados no conforman una muestra representativa de los conocimientos disponibles y, al compilarlos, no es posible obtener una estimación promediada e insesgada de los parámetros que pretendemos estimar al acometer investigaciones.

Para hacer referencia a la tendencia de autores, revisores y editores a enviar a publicación, recomendar la aceptación y aceptar para publicación, respectivamente, los estudios sobre la base de la magnitud y dirección de los resultados (entendidas desde el prisma de la significación estadística), se acuñó el término *sesgo de publicación*. Con arreglo al análisis epidemiológico, este error sistemático corresponde a un sesgo de selección (Schlesselman JJ, 1982; Michael M, Boyce WT, Wolcox AJ, 1984; Mausner JS, Kramer S, 1985; Kelsey JL, Thompson WD, Evans AS, 1986; Last JM, 1988).

Asimismo, en esta memoria de investigación se tratará de confirmar si la calidad no es el principal factor que rige la decisión de publicar o rechazar un estudio. Por el contrario, en numerosas investigaciones se ha comprobado que tanto la significación estadística como

³ Según este principio, que constituye uno de los criterios de causalidad de Hill (Hill AB, 1965), la supuesta asociación, para considerarse causal debe, entre otros requisitos, poder detectarse en estudios realizados en diversos momentos y en distintas circunstancias.

otras variables prevalecen a menudo en la revistas médicas como criterios que gobiernan la publicación de estudios.

Al revisar por orden cronológico la literatura sobre este tema, se presencia claramente la construcción progresiva del concepto de sesgo de publicación. Dicha revisión ejemplifica un hecho nada infrecuente en medicina: la incorporación de métodos de investigación procedentes de otras disciplinas del saber con objeto de solucionar problemas importantes en el campo de las ciencias de la salud.

Para no perder la riqueza informativa que ofrece esta visión histórica, se ha optado por estructurar la introducción que sigue conforme a un criterio cronológico. Ello explica que algunos pasajes puedan parecer inconexos. Esta falta de hilación pretende ser fiel reflejo del avance escalonado del conocimiento científico en esta pequeña parcela de la investigación médica.

1.2 Primeros antecedentes históricos

El interés por el sesgo de publicación ha ido cobrando prominencia en la literatura médica de las últimas cuatro décadas. Como se describe en adelante, el concepto de sesgo de publicación y las medidas propuestas para minimizarlo han ido madurando paulatinamente. Ello se ha debido principalmente a dos factores. Primero, a la aplicación de distintos métodos estadísticos y epidemiológicos en su investigación. Segundo, y sobre todo, a los conocimientos que numerosos estudios han aportado sobre su naturaleza y sus causas.

La primera mención indirecta del sesgo de publicación se atribuye al químico Robert Boyle, quien en 1661 afirmó: <<Muchos experimentos y nociones son suprimidos por hombres soberbios y modestos>> (Hall MB, 1965). Dos siglos más tarde, un editorial de la revista *Boston Medical Journal* incluía el siguiente comentario: <<Con mucha

asiduidad vemos referencias a "tantos casos exitosos" en las cuales se hace inevitable hincapié en la palabra "exitoso" [...] Sin lugar a dudas, todas esas publicaciones imprimen un falso énfasis que tiende a elevar la reputación del autor, pero no a aumentar la seguridad del público>> (Hall MB, 1965).

En el presente siglo se han hecho numerosas alusiones al sesgo de publicación, aunque hasta principios de los sesenta no se comenzaron a acopiar pruebas fehacientes de su existencia. Las primeras investigaciones sobre el tema se llevaron a cabo en el campo de la psicología. En la publicación que podría considerarse primigenia Sterling afirmó que diversas revistas habían decidido fijar el nivel de significación estadística de los artículos que aceptaban (o el dintel de rechazo de la hipótesis nula) en 0,05 (Sterling, 1959). Esta convención no parecía acarrear *per se* mayores consecuencias. Empero, su aplicación como criterio de aceptación de artículos encerraba --como se comprobó más tarde-- serios peligros.

Para respaldar su afirmación, Sterling revisó 362 artículos de cuatro prestigiosas revistas de psicología publicados en 1955 y 1956. En 294 (81%) se habían aplicado pruebas de significación. En 97% (recorrido de valores: 95% - 99%) de estos últimos se había rechazado la hipótesis nula y sólo en 2,7% (recorrido: 0,9% - 4,8%) se había aceptado. Esta última cifra es menor del 5% esperado en general, en virtud del nivel de significación estadística acordado de 5%. En su opinión, dicha decisión era más atribuible al juicio de los autores que a la política editorial de las revistas. A raíz de esos resultados concluyó que, si bien se desconocía el número de investigaciones en que se había rechazado la hipótesis nula, la probabilidad de que un estudio se publicase era más alta si se había rechazado que si se hubiese aceptado; que la probabilidad de que un estudio se replicara era muy baja una vez publicado (ninguno de los artículos que revisó constituían repeticiones de estudios

publicados con anterioridad), y que se realizaban muchos más estudios de los que se publicaban.

Obviamente, la comunidad científica no puede conocer todos los estudios que se llevan a cabo. Pero, a juzgar por los hallazgos de Sterling, ésto parecía ser especialmente cierto en aquellos estudios que no habían rechazado la hipótesis nula, ya que su probabilidad de ser publicados parecía ser menor. Resultaba paradójico que en la literatura aparecieran *resultados negativos* con menor frecuencia que en la práctica científica, incluso suponiendo que en todos los estudios las hipótesis se formularan y seleccionasen con absoluto rigor.

Una de las consecuencias de la adopción de la política editorial criticada por Sterling es que, con el correr de los años, los lectores esperan que en los artículos que consultan se rechacen hipótesis nulas, a menudo sin tener presente que algunos de esos resultados corresponden a errores de tipo I (rechazar la hipótesis nula cuando es cierta). De hecho, los hallazgos de Sterling atestiguaron que la literatura contenía más errores de tipo I de los que cabía esperar a tenor del nivel alfa escogido en los estudios.

Cinco años más tarde, Smart replicó el estudio de Sterling con artículos publicados en 1961 y 1962 en las mismas revistas (Smart RG, 1964). El porcentaje de estudios negativos en esta ocasión osciló entre 7 y 12%. Una pregunta que se deriva de este hallazgo es ¿hasta qué punto reflejan estos porcentajes la proporción real de estudios negativos realizados, al margen de que se hayan publicado? Para responderla, y abrigando la sospecha de que dicho porcentaje fuese más alto en los estudios no publicados, revisó 100 estudios presentados en la reunión de la *American Psychological Association* de 1962. El porcentaje de estudios negativos ascendió a 20% y en las tesis doctorales no publicadas, a 30%. Este resultado induce a pensar que algún sesgo de selección debe de interferir en la selección de artículos para publicación. Paralelamente constató que la probabilidad de que una tesis

doctoral con resultados negativos se publicara era mucho menor que la de aquellas con resultados positivos.

Lamentablemente, Smart desconocía la proporción de estudios negativos que no se intentaban publicar y la de los que se rechazaban. Acaso muchos de ellos no se publican porque se rechazan de entrada a consecuencia del sesgo mencionado o porque se someten a revisiones más rigurosas que los positivos. Como solución, propuso publicar los resúmenes de los estudios negativos junto con las señas de los autores y se mostró remiso a la propuesta lanzada tiempo atrás de publicar revistas de resultados negativos (Black, 1958) ante las reducidas dimensiones del espacio editorial (Greenwald, 1975).

1.3 El desprecio de la hipótesis nula

En la década de los setenta se constataron los prejuicios de muchos investigadores que menospreciaban la aceptación de la hipótesis nula. Analizadas desde un prisma metodológico, las expresiones de esas aprensiones son hasta cierto punto irrisorias. Como ejemplos de ellas valga citar las siguientes: <<los resultados negativos son fuente de incertidumbre>>; <<las conclusiones sobre relaciones entre variables sólo deben basarse en rechazos de la hipótesis nula>>; <<observar que dos variables no están relacionadas no aporta prácticamente conocimiento alguno>>; <<la ciencia avanza descubriendo relaciones entre variables>>; <<la comprobación de efectos estadísticamente significativos es casi un garante del rigor de la investigación>>... (Bacan, 1966; Wilson WR, Miller H, 1964; Greenwald, 1975).

Para confirmar esa actitud contra la hipótesis nula, Greenwald administró un cuestionario a 48 autores y a 47 revisores de la *Journal of Personality and Social Psychology* (Greenwald, 1975). El estudio reveló la preferencia de los autores por publicar estudios positivos (las probabilidades estimadas de enviar a publicación un estudio positivo

y uno negativo fueron, respectivamente, 0,49 y 0,06) y de abandonar la investigación frente a la imposibilidad de rechazar la hipótesis nula.

Sobre la base de los resultados obtenidos, Greenwald construyó un modelo matemático con el propósito de estimar las probabilidades de que los autores seleccionasen un problema de investigación sospechando que la hipótesis nula fuese cierta, prosiguieran las investigaciones a la luz de los resultados, intentaran publicarlos, y la de que los editores aceptasen publicar estudios positivos y negativos. A pesar de las restricciones de algunos supuestos del modelo, los resultados pusieron de manifiesto la baja cifra de estudios publicados sobre temas en los cuales la hipótesis nula era cierta y que de ellos una elevada proporción la rechazaban erróneamente (error de tipo I). El intervalo de confianza del 95% (IC95% en adelante) de la proporción de artículos publicados en los que se aceptaba la hipótesis nula fue 7,6% - 18,6%.

Entre las medidas que se barajaron por aquel entonces para encarar el problema destacan la adecuación de la potencia estadística a los propósitos del estudio y la sustitución de las pruebas de hipótesis por pruebas de estimación y técnicas bayesianas (Grant D, 1962; Bakan, 1966). A éstas recomendaciones Greenwald añadió las siguientes: formulación de hipótesis nulas de intervalo en lugar de puntuales; selección de un tamaño muestral congruente con el error estándar del estimador estadístico; inclusión de pruebas convincentes de las medidas empleadas y de las manipulaciones realizadas en el estudio; cálculo de la probabilidad *a posteriori* de la hipótesis nula, y notificación de todos los resultados relevantes obtenidos (o pruebas de hipótesis realizadas) y no sólo de los que mejor encajan con los objetivos.

El estudio de Greenwald representó un avance sustancial en la investigación de los sesgos de publicación. Primero, por utilizar técnicas convencionales como la encuesta para desentrañar los factores que los explican. Segundo, por la primicia que significó la

construcción de un modelo matemático destinado a simular el intrincado binomio investigación-publicación.

A la luz de sus hallazgos cabe preguntarse: ¿cómo influye la publicación preferente de resultados estadísticamente significativos en la exactitud de las estimaciones de la magnitud de los efectos (por ejemplo, diferencias entre proporciones, medias, tiempos de supervivencia o riesgos)? (Greenwald, 1975). Probablemente, la presión que ejerce el empleo de la significación estadística como criterio de publicación desemboque en sobreestimaciones de la magnitud real de los efectos. En un estudio de simulación se constató que cuanto mayor era la sobreestimación del efecto tanto mayor era la probabilidad de que fuese publicado (Lane DM, Dunlap WP, 1978). Como solución, Greenwald recomendó que las decisiones editoriales se basaran exclusivamente en las secciones de introducción y métodos. Es decir, que un estudio debía publicarse si estaba justificado teóricamente y bien diseñado e independientemente de la dirección de los resultados de las pruebas de hipótesis (Greenwald, 1979).

A menudo, las inferencias incorrectas derivadas del sesgo de publicación dependen más de la fuerza de una asociación o de la magnitud de una diferencia que de su dirección (por ejemplo, que el riesgo relativo sea mayor o menor de 1,0). Pero, con frecuencia --y esta es la consecuencia más deletérea del sesgo-- la discrepancia entre los estudios publicados y los no publicados radica tanto en la magnitud del efecto estimado como en su dirección. Como se comprobó, esas discrepancias no son atribuibles a diferencias entre la <<calidad>> de los estudios publicados y no publicados. De hecho, en algunas especialidades médicas se ha demostrado, como se verá más adelante, que la calidad de algunos estudios no publicados excede la de los publicados, lo cual debe resaltarse por la importancia que entrañan sus consecuencias en la toma de decisiones médicas (Juhl E, Christensen E, Tygstrup N, 1977).

1.4 Definición y confirmación del *sesgo de publicación*

Frente a los hechos descritos, Smith acuñó el término *sesgo de publicación* para designar la selección que se produce cuando la publicación de un estudio se basa primordialmente en la significación estadística o en la dirección de los resultados. Todo confirmaba la capacidad agorera de Francis Bacon, pues con más de tres siglos de antelación ya había aseverado: <<el intelecto humano [...] se mueve y excita más por afirmativos que por negativos>> (Smith, 1980).

¿Cuál es, entonces, el origen del sesgo de publicación? ¿Por qué se distorsiona el conocimiento de ese modo? El proceso de adquisición de conocimiento ha sido objeto de análisis en una amplia amalgama de estudios epistemológicos y psicológicos. Aunque este tema escapa al objetivo de esta memoria de investigación, parece lícito pensar que el sesgo de publicación estribe en alguna alteración de dicho proceso.

Por su especial pertinencia al tema abordado destaca el denominado *sesgo de confirmación*: la tendencia del hombre a buscar, prestar atención y ornamentar las experiencias que respaldan o confirman sus ideas. Por el contrario, los hechos que refutan sus pareceres tienden a ser ignorados o desacreditados. Así lo indican las opiniones de los científicos que integraron una muestra (aceptablemente representativa) extraída a nivel nacional en los Estados Unidos. La encuesta reveló su marcada preferencia por las investigaciones <<confirmatorias>> (Mahoney MJ, Kimper TP, 1976). La publicación de resultados es condición *sine qua non* del avance de cualquier campo del conocimiento; sin difusión de resultados no es posible el conocimiento. La interpretación errada de este hecho explica dos hechos desafortunados. Primero, que los peligros que encierra el sesgo de confirmación sean especialmente prominentes en el campo de la publicación. Segundo, que el uso del número de publicaciones como uno de los criterios cardinales para valorar a los

profesionales y decidir su progreso refuerce la tendencia a publicar al margen de la calidad de los estudios y con ello a introducir sesgos de publicación.

La contribución potencial de un estudio al conocimiento puede verse seriamente amenazada por una decisión equivocada del editor de una revista. El primer estudio *controlado* que se publicó sobre los aspectos relacionados con el proceso de revisión editorial (en inglés, *editorial peer review*) se centró en el contenido de los artículos y su consiguiente interpretación (Mahoney JM, 1977). En este trabajo se alteró el contenido de las secciones de resultados y discusión de artículos en los que se investigaba la existencia de una asociación entre variables, por aquel entonces polémica. Sesenta y siete revisores de una revista se asignaron al azar a cinco grupos distribuidos del siguiente modo: grupo 1, resultados positivos y discusión omitida (a favor de la asociación); grupo 2, resultados negativos y discusión omitida (en contra de la asociación); grupo 3, sólo se les entregaron las secciones de introducción y métodos; grupo 4, resultados positivos y negativos con discusión positiva, y grupo 5, resultados positivos y negativos, y discusión negativa. Asimismo, se les pidió que evaluaran en una escala numérica la relevancia del estudio, métodos, presentación de resultados, discusión y contribución científica global de los artículos. Si bien los revisores no discreparon en su graduación de la relevancia del tema analizado, en los cuatro apartados restantes los estudios positivos recibieron sistemáticamente puntuaciones más altas que los negativos. Al recabar su opinión final, los revisores de los estudios positivos recomendaron aceptarlos en su mayoría, y los que revisaron los negativos, rechazarlos. Este estudio se acometió con aceptable rigor metodológico, fue el primero controlado sobre el tema y corroboró, con el respaldo que aportan los experimentos controlados bien realizados, la existencia del sesgo de confirmación. Además, destacó una deficiencia hasta cierto punto inmanente a la revisión editorial, puesta de manifiesto pocos años antes por editores de revistas de psicología (Scott

WA, 1974; Cichetti DV, Conn HO, 1976), que, posteriormente, se convirtió en medicina en un tema de hondo calado en el contexto de los estudios de la revisión editorial de artículos: la baja concordancia o correlación entre las evaluaciones de los revisores de artículos presentados para publicación.

A raíz de estas investigaciones, se desarrollaron variantes de modelos matemáticos para estimar dicha concordancia y se puso de relieve la subjetividad que embebe la revisión de artículos (Strayhorn J, McDermott JF, Tanguay P, 1993; Ector H, Aubert A, Strobandt R, 1995). Escapa a los objetivos de esta memoria pormenorizar los resultados de todos los estudios que reafirman estos hechos. Habida cuenta de la baja concordancia observada reiteradamente entre los revisores de numerosas revistas, sería un desatino negar que la subjetividad de los revisores se incorpora e influye, en mayor o menor medida, en todas y cada una de las decisiones editoriales relativas a la calidad, el rigor y la relevancia científica de los artículos, y, por ende, en la decisión final de publicarlos o rechazarlos.

1.5 El problema del archivador

El año 1979 marcó un hito en la investigación de los sesgos de publicación, pues su existencia se formuló en otros términos. Si la significación estadística constituye en realidad un criterio que rige la aceptación de estudios para publicación, entonces los estudios publicados integran una muestra sesgada de las investigaciones que se han llevado a cabo. La situación extrema en que puede desembocar este hecho es que las revistas den cabida a estudios que contienen errores de tipo I, mientras que los *archivadores* de los investigadores estén repletos de estudios negativos que nunca alcanzarán las páginas impresas. Para denominar este fenómeno, Rosenthal acuñó un término: *el problema del archivador* (en inglés, *The file drawer problem*) (Rosenthal, 1979).

Hasta ese momento, no había sido posible valorar el efecto neto o global estimado por los estudios que se habían realizado sobre un tema. Los archivadores podrían estar atestados de estudios --algunos de ellos originales, relevantes y metodológicamente rigurosos-- que jamás se difundirán por no haber rechazado la hipótesis nula principal de la investigación. Aunque no se había concebido ninguna medida para solucionarlo, Rosenthal ideó un método para estimar la magnitud del problema del archivador. Con este propósito, desarrolló dos fórmulas matemáticas sencillas a fin de calcular el número máximo de estudios negativos que debían contener los archivadores para mantener la significación estadística alcanzada por un conjunto de estudios publicados. Parafraseando, él trató de calcular el número necesario de estudios negativos no publicados para trocar el valor P significativo obtenido al combinar estudios publicados en un valor no significativo. Si esa cifra fuese alta, deberían existir muchos estudios negativos no publicados antes de que el estudio positivo publicado pudiese considerarse un error de tipo I o el resultado de la existencia de un sesgo de selección en la publicación. Por el contrario, si fuera baja, la magnitud del problema del archivador sería muy pequeña. El método de Rosenthal se aplicó posteriormente, como se comentará, en la estimación de sesgos de publicación en metanálisis realizados en medicina⁴.

Paralelamente a estos desarrollos matemáticos, se formuló la pregunta que poco tiempo después marcaría el rumbo de los estudios de los sesgos de publicación: ¿es mejor la calidad científica de los estudios publicados que la de los no publicados? Con el afán de disipar esta duda, Hemminki averiguó que un elevado número de ensayos clínicos

⁴ El método consiste en una modificación del método de Stouffer para combinar valores P en metanálisis (la adición de las desviaciones estándar (DE) asociadas con el valor P de cada estudio (i), dividido por la raíz cuadrada del número de estudios combinados (n)) (Stouffer SA, Suchman EA, Devinney LC, Star SA, Williams RA, 1949). Años más tarde, este método se criticó, pues, entre otras razones, ignora la posible variación del tamaño de los estudios no publicados y no permite derivar ningún estimador del efecto combinado ni una prueba de significación estadística de la magnitud del sesgo de publicación. Otra variante del método de Rosenthal es la de Fisher (Begg CB, Berlin JA, 1988).

realizados en Suecia entre 1965 y 1975 no se publicaron, es decir, permanecían en el *archivador*. Los ensayos no publicados de medicamentos psicotropos contenían más a menudo información importante relativa a los criterios de inclusión y exclusión de pacientes que los publicados. En términos generales, la calidad metodológica de buena parte de los ensayos publicados y no publicados, o cuando menos la de la presentación de los datos, corría pareja y en ambos casos era sustancialmente menor de la que cabía desear (Hemminki E, 1980). Asimismo, descubrió que la probabilidad de que se publicaran los estudios que demostraban efectos secundarios de los medicamentos cuya licencia de comercialización se había solicitado era menor que la de aquellos que no indicaban la existencia de esas reacciones. El estudio aportó pruebas fehacientes para pensar que la calidad metodológica de un ensayo clínico financiado por la empresa farmacéutica no constituye un factor decisivo de la decisión editorial de publicarlo.

Este y otros estudios similares pusieron de manifiesto deficiencias notables tanto en el diseño, análisis e interpretación de resultados de ensayos clínicos, como en la presentación de datos (Mosteller F, Gilbert JP, Bucknam Mcpeek, 1980; DerSimonian R, Charette LJ, McPeek BA, Mosteller F, 1982; Emerson JD, McPeek B, Mosteller F, 1983).

Otros estudios realizados con revistas de psicología arrojaron resultados congruentes con los mencionados. En uno de ellos se pidió a 101 editores asociados de dos revistas de psicología que evaluaran tres versiones preparadas de un manuscrito que sólo diferían en el grado de significación estadística de los resultados. La probabilidad de rechazar los negativos fue tres veces más alta que la de los positivos (Atkinson DR, Furlong MJ, Wampold BE, 1982), la correlación lineal entre variables tendía a ser más elevada en los estudios publicados que en los no publicados (White, 1982), y la magnitud de las diferencias observadas entre variables de estudios publicados en revistas fue mayor que la detectada en tesis doctorales (Smith, 1980). Al presentar versiones abreviadas de artículos

publicados a un grupo de psicólogos clínicos también se constató el mismo sesgo: la calidad de los trabajos publicados con resultados estadísticamente significativos se consideró menor que la de aquellos en los que no se había podido rechazar la hipótesis nula (Cohen, 1980).

1.6 Los estudios negativos

Un análisis somero de los estudios negativos suscita el siguiente interrogante: ¿qué problema de investigación plantea realmente un estudio negativo? (Maxwell C, 1981). Una posibilidad es que el estudio no haya detectado diferencias porque no existen, el caso ideal. Otra, no menos plausible, es que no la haya descubierto a pesar de su existencia real. En este caso, nos encontramos ante un error de tipo II: no rechazamos la hipótesis nula cuando en realidad es falsa⁵. Varios investigadores demostraron que en 67 de 71 ensayos clínicos negativos la probabilidad de no haber detectado diferencias de 25% entre las proporciones de respuesta al medicamento estudiado y al placebo ascendía a más de 10% (Freiman JA, Chalmers TC, Smith R, Keubler RR, 1978).

El problema que pueden plantear los estudios negativos se agrava si a la par se considera la existencia de sesgos de publicación. Inevitablemente, los ensayos clínicos publicados conforman una muestra no representativa de los realizados. Como solución primigenia Black había propuesto tiempo atrás la creación de una *revista de resultados negativos* (Black D, 1958). La pobre acogida que tuvo esta propuesta entre muchos investigadores condujo a plantear una variante más moderada: publicar el título de los trabajos de calidad aceptable (y que no parecieran corresponder a errores de tipo II), así como el nombre de sus autores y sus señas. Aunque esta medida no evitaría

⁵ Es preciso hacer hincapié en que la respuesta a este interrogante exige la concurrencia de cuatro factores matemáticamente relacionados: la diferencia entre los tratamientos --un elemento inalterable--, los riesgos de cometer errores de tipo I y tipo II, y los tamaños muestrales. Los tres últimos varían a medida que se modifica cada uno de ellos.

completamente los distintos riesgos que acarrea el sesgo de publicación, podría ser de cierta ayuda a tal efecto (Maxwell, 1981; Black DB, 1958). Hasta la fecha, no ha cristalizado.

En los años ochenta se dio un espaldarazo a los estudios destinados a evaluar tanto la existencia de sesgos de publicación en la literatura médica como la magnitud de sus efectos. En sus prolegómenos, a 296 autores de ensayos clínicos publicados, integrantes de una muestra aleatoria, se les preguntó si habían participado en ensayos clínicos no publicados. De los 921 ensayos en que habían participado, 196 no se habían publicado; de éstos, 76 porque el ensayo finalizó antes de lo previsto —muchos eran negativos—, otros 76 no se enviaron a publicación, 7 fueron rechazados por las revistas, y 37, por razones no especificadas. La principal limitación de este estudio es que en él no se comparó la calidad de los ensayos publicados con la de los no publicados. No obstante, confirmó de nuevo que la probabilidad de que los ensayos positivos se publiquen es más alta que la de los negativos. El estudio sustentaba la sospecha de que la literatura médica contuviera una proporción de resultados positivos falsos más alta de lo esperado.

En las revistas médicas se hacía cada vez más patente el problema de los estudios negativos⁶. Para documentarlo, diversos investigadores analizaron la potencia de los ensayos negativos para detectar diferencias consideradas clínicamente importantes. Buena parte de los tratamientos tildados de <<similares al control>> en ensayos clínicos con tamaños muestrales insuficientes no se habían revisado adecuadamente (Freiman JA, Chalmers TC, Smith H, Keubler RR, 1978). Como solución, se han elaborado diversos

⁶ El tamaño muestral que debe emplearse en un estudio depende de varios factores: la variable de resultado que se evalúa (vgr., dicotómica, continua...), los errores de tipo I y II escogidos, la magnitud de la diferencia que se pretende detectar (¿qué diferencia es clínicamente importante encontrar?), la variabilidad inherente de la medición de la variable de interés (cuanto mayor, mayor tamaño muestral se precisa), y la razón entre el número de sujetos en el grupo experimental y el control (una razón 1:1 minimiza el tamaño muestral necesario) (Levy PS, Lemeshow S, 1991; Lwanga SK, Lemeshow S, 1991; Silva LC, 1993).

algoritmos, tablas y guías que permiten determinar si el tamaño muestral de los estudios es suficiente para detectar las diferencias prefijadas en los estudios, especialmente en los negativos (Young MJ, Bresnitz EA, Strom BL, 1983; Detsky AS, Sackett DL; 1985).

La importancia o gravedad de este hecho estriba además en la índole de los estudios revisados. El ensayo clínico controlado es el tipo de estudio que permite ejercer el mayor grado de control sobre las variables analizadas. El control alcanzado explica su elevada validez interna, si bien a expensas de reducir a menudo su validez externa (Meinert CL, 1986). En los restantes tipos de estudios (cuasiexperimentales y observacionales) el control es sustancialmente menor o inexistente y, por ende, el rigor de las inferencias (Campbell DT, Stanley JC, 1963; Breslow NE, Day NE, 1980; Schlesselman JJ, 1982; Fleiss JL, 1986; Breslow NE, Day NE, 1987).

La importancia que revisten estos principios en el contexto del sesgo de publicación reside en que los ensayos clínicos se realizan con menor asiduidad y, por tanto, su publicación es menos frecuente. Hasta los años ochenta, el sesgo de publicación sólo se había investigado en ensayos clínicos. Restaba por comprobar si dicho sesgo también afectaba a la publicación de los restantes tipos de estudios, mucho más habituales en la literatura. De ser esto cierto, la gravedad de sus consecuencias aumentaría.

La polémica desatada en torno a los estudios negativos cristalizó en una recomendación cardinal para editores, revisores y autores: todos los ensayos realizados adecuadamente, conforme a principios metodológicos aceptados, constituyen contribuciones <<positivas>> al conocimiento científico, y los criterios de revisión editorial han de ceñirse a este principio (O'Brian Smith E, 1986).

Muchos proclamaron que el concepto de *ensayo negativo* derivaba de una acusada reverencia a diferencias entre valores de variables que sobrepasaban un dintel de significación estadística elegido, con harta frecuencia, sin justificarlo adecuadamente.

Incluso la editora jefe de la prestigiosa *N Engl J Med* concurrió más tarde al debate (Angell, 1989) sobre la idoneidad del término <<estudio negativo>>, a raíz de la publicación en esa revista de un estudio de casos y controles negativo (Mills JL, Rhoads GG, Simpson JL, et al, 1989). Angell señaló que la interpretación de los estudios negativos presenta escollos singulares. La ausencia de datos para respaldar, por ejemplo, la eficacia de un medicamento o la efectividad de una intervención no es sinónimo de disponer de pruebas sobre su ineficacia o inefectividad. A la pregunta ¿publicaría dicha revista un estudio negativo de calidad pareja a uno positivo? Angell respondió que todo depende del tema que se trate, y que el principal motivo que explica que la revista acepte más estudios positivos que negativos es que muchos estudios negativos abordan hipótesis poco relevantes. Por último, concluyó que la frecuencia relativa con que se publican estudios negativos en esa revista no es indicativa de la existencia de sesgo de publicación, sino que simplemente trasluce la importancia del tema estudiado.

Antes de tomar decisiones sobre este tema, deberían escrutarse a fondo los estudios que los investigadores han decidido arrinconar en el archivador. Dichos estudios no han atravesado el tamiz de la revisión editorial, y ello puede explicar, por ejemplo, que muchos depositen más confianza en los metanálisis que sólo incluyen estudios publicados. Los criterios de aceptación de la *N Engl J Med* son idénticos para los estudios positivos y negativos, y en la revista se percibe como obligación el publicar estudios negativos que refutan a otros positivos publicados siempre que su calidad sea similar o superior. Al juicio de Angell podría objetársele el exceso de credibilidad que atribuye a la revisión editorial, sobre todo a la luz de la subjetividad y baja concordancia que aqueja el proceso (Dickersin K, 1990 (2); Erill S, 1992).

En sendas respuestas a las aseveraciones de Angell se subrayó el sentido peyorativo y poco atinado del término <<estudio negativo>> (Dickersin K, 1990; Tunstall-Pedoe H,

1990). Al contrastar una hipótesis uno puede obtener cinco resultados distintos: una asociación positiva y estadísticamente significativa; una asociación positiva no significativa; ausencia de asociación (resultado que en inglés se denominó *nil* en lugar de *nul*); una asociación negativa o inversa significativa, y una asociación negativa no significativa. El uso del término <<estudio negativo>> hace referencia a las cuatro últimas alternativas, en tanto que la posición de Angell sólo aludía a la segunda y tercera. El problema terminológico responde en realidad al modo en que se presentan resultados a una revista, porque el hincapié recae más en el rechazo de hipótesis que en la estimación de, por ejemplo, diferencias o de la fuerza de una asociación. Otros detractores del término esgrimieron como argumento la paradoja de Simpson⁷, que a veces aparece en metanálisis.

Al ahondar en las consecuencias derivadas de los estudios negativos, Newcombe esbozó una nueva propuesta para minimizar los efectos del sesgo de publicación (Newcombe RG, 1987). Según su planteamiento, la interpretación de la potencia estadística sólo es satisfactoria cuando ésta se calcula con valores especificados en la etapa de diseño del estudio. La forma adecuada de valorar la idoneidad del tamaño muestral reside en revisar los valores especificados en el protocolo del estudio. En numerosas ocasiones se había comprobado –y se sigue constatando en la actualidad– que los manuscritos recibidos en las revistas a menudo contienen errores estadísticos crasos, a pesar de las numerosas guías, recomendaciones y listas de comprobación sobre métodos estadísticos y presentación de estos datos que se han publicado hasta la fecha (Mahon WA, Daniel EE, 1964; Schor S, Karten I, 1966; Lionel NDW, Herxheimer A, 1970; O'Fallon JR, Dubey SD, Salsburg DS, Edmonson JH, Soffer A, Colton T, 1974; Mosteller F, Gilbert JP, McPeck B, 1980; DerSimonian R, Charette LJ, McPeck B, Mosteller F, 1982; Zelen M, 1983; Altman DG,

⁷ Esta paradoja expresa la obtención de un resultado global de signo opuesto al de los estudios individuales que se combinan.

Gore SM, Gardner MJ, Pokock SJ, 1983; Emerson JD, McPeck B, Mosteller, 1984; Meinert CL, Tonascia S, Higgins K, 1984; Kroch CL, 1985; Lock S, 1985; Meinert CL, 1986; Gardner MJ, Machin D, Campbell MJ, 1986; Pokock SJ, Hughes MD, Lee RJ, 1987; Poynard T, 1988; Campillo C, Silva G, 1991; Wilson A, Henry DA, 1992; *Annals of Internal Medicine*, 1992; *American Journal of Epidemiology*, 1992; *American Journal of Epidemiology*, 1992; Audet N, Gagnon R, Ladouceur R, Marcil M, 1993; Programa de Publicaciones, Organización Panamericana de la Salud/Organización Mundial de la Salud, 1993; Campillo C, 1993; Concato J, Feinstein AR, Holdford TR, 1993; González JC, Pulido M, Sanz F, 1995)⁸.

La consecuencia más palmaria del tamaño muestral insuficiente es la obtención de resultados negativos falsos. La publicación de estos estudios puede dar la falsa impresión de que no existen diferencias entre los valores comparados. De este hecho dimanan dos corrientes de pensamiento en el seno de las decisiones editoriales. Primero, la que propone el requisito de presentar intervalos de confianza en lugar de valores P (Gardner MJ, Altman DG, 1986) y segundo, la que postula no publicar estudios negativos de baja potencia estadística. En este último caso, se puede proceder de tres maneras distintas: aceptar únicamente los estudios que alcancen un nivel determinado de significación estadística (como el valor P depende de la variabilidad muestral, el empleo de este criterio puede desembocar en sesgo de publicación); evaluar la potencia del estudio a partir de los datos

⁸ Sobre este particular cabe hacer una acotación. Diversas revisiones metodológicas de la avalancha de publicaciones sobre evaluaciones socioeconómicas de programas, servicios y tecnologías sanitarios realizadas en los últimos años también han puesto de relieve los numerosos errores de diseño, análisis y presentación de datos que éstas contienen (Udvarhelyi S, Colditz GA, Epstein AM, 1992; Ganiats TG, Wong AF, 1991; Drummond M, Brandt A, Luce B, Rovira J, 1993; Drummond MF, Stoddart GL, Torrance GW, 1987). En dichas revisiones se ha destacado la urgencia de disponer de guías para la realización, presentación de datos y revisión de esos estudios (Jefferson T, Demicheli V, 1995). Al revisarlos, se aprecia la tendencia a financiar sólo aquellos que pueden incluirse bajo el epígrafe de estudios positivos y se pone reiteradamente de manifiesto el acicate que supone la aceptación preferente de supuestos metodológicos favorables (Hillman AL, Eisenberg JM, Pauly MV, et al, 1991). Esto es indicativo de los errores metodológicos que aún se cometen en todas las disciplinas médicas.

notificados (una alternativa que no introduce ese sesgo), y exigir establecer *a priori* un valor beta adecuadamente bajo (lo cual protege de ese sesgo, habida cuenta de que dicho valor beta no está sujeto a variabilidad muestral).

Newcombe añadió que el sesgo de publicación continuaba produciéndose básicamente por dos motivos: la utilización de criterios de selección editoriales inadecuados (pruebas de significación estadística) y la profusión de estudios con bajo tamaño muestral (Newcombe RG, 1987). Una de las soluciones que propuso --el reemplazo de las pruebas de hipótesis por las de estimación-- también acarrea el riesgo de introducir sesgos de selección. Autores y editores podrían seguir mostrando una predilección por enviar a publicación y a aceptar, respectivamente, sólo aquellos intervalos de confianza que equivalen a diferencias estadísticamente significativas (es decir, que no contuvieran, según el caso, el valor 1 o el 0). En cuanto al segundo motivo, ya se han mencionado los peligros que encierra la inclusión exclusiva en metanálisis de estudios publicados. Si la alta prevalencia de estudios con bajo tamaño muestral deriva de la forma como se organizan las investigaciones, acaso la solución estribe en fomentar la realización de estudios colaborativos y multicéntricos, dotados de mayor número de participantes que los monocéntricos (Meinert CL, 1986).

A la luz de estas reflexiones, Newcombe propuso basar la publicación de estudios en su idoneidad *a priori*: ¿puede brindar información útil el diseño empleado? Si la respuesta es afirmativa, entonces el estudio habría de publicarse al margen de que hubiese alcanzado significación estadística. El único requisito adicional *a posteriori* radica en su adherencia a las premisas fijadas en su protocolo.

Sin embargo, esta propuesta abrigaba dos peligros: no publicar algunos estudios positivos verdaderos al tiempo que los estudios negativos verdaderos se publicarían con mayor frecuencia a expensas de publicar más estudios negativos falsos, si bien esta posibilidad se minimizaría si las betas elegidas fuesen suficientemente bajas. Por otra parte,

respaldaba indirectamente la postura de otros investigadores, por cuanto era congruente con la recomendación de registrar los estudios desde la etapa de aprobación de los protocolos (Dickersin K, 1987). (Los registros de estudios se abordarán en detalle en una sección posterior). Empero, ese registro debería ser independiente de las revistas, dada la sobrecarga que soportan normalmente sus sistemas de revisión editorial. Aludiendo a su recomendación de fomentar la realización adecuada de metanálisis, es preciso tomar en cuenta la dificultad de localizar estudios negativos. Este obstáculo podría salvarse creando registros de estudios en todas las especialidades, una iniciativa que ya se estaba consolidando a fines de la década de los ochenta. Como ejemplos cabe recordar el mencionado *Oxford Database of Perinatal Trials* (Chalmers I, Hetherington J, Newdick M, Mutch L, Grant A, Enkin M, et al, 1983), *el Registro Europeo de Ensayos Clínicos sobre Trombolíticos* (Verstraete M) y *The Physician Data Query*, dependiente de los Institutos Nacionales de la Salud de los Estados Unidos, que pone al alcance de investigadores información actualizada periódicamente sobre el tratamiento del cáncer (Molloy Hubbard S, Henney JE, DeVita VT, 1987). El cometido de difundir estudios negativos bien realizados también podría recaer en instituciones o entidades independientes del mundo editorial, tales como los comités éticos de investigación (Easterbrook P, 1987) o la Agencia Internacional de Investigación sobre el Cáncer (Higginson J, 1987), por citar algunos ejemplos.

1.7 Primeros modelos matemáticos para estimar el sesgo de publicación

A mediados de los ochenta, el sesgo de publicación ya se habían convertido en un hecho palmario aceptado por amplios sectores de las ciencias médicas y, especialmente, por editores y epidemiólogos anglosajones. Las primeras soluciones propuestas para evitarlo se habían circunscrito a formular pautas y guías para valorar la idoneidad de la

presentación de los elementos metodológicos capitales del diseño y el análisis del estudio, como la descripción de los métodos de muestreo, selección y asignación de participantes o métodos estadísticos. Un segundo grupo de soluciones trató de construir modelos conceptuales para analizar los resultados de las pruebas estadísticas de los ensayos clínicos como variables binarias: estadísticamente significativos y no significativos. Se demostró fehacientemente que la razón entre resultados positivos (estadísticamente significativos) verdaderos y positivos falsos sería relativamente baja si la frecuencia de diferencias clínicamente relevantes *verdaderas* fuese baja. Esto es cierto, sobre todo, cuando los tamaños muestrales de los estudios son pequeños (Staquet MJ, Rozenzweig M, Von Hoff DD, Maggia FM, 1979; Simon R, 1982; Zelen, 1982; Fleming TR, 1982). En este contexto debe enmarcarse el estudio que realizó tempranamente Rosenthal (Rosenthal, 1979). Asimismo, se hizo hincapié en los peligros que encerraban los análisis de subgrupos --la tendencia a sobrevalorar y sólo publicar los resultados de los subgrupos más positivos -- y su proclividad a engrosar el arsenal de estudios positivos falsos publicados (Donner A, 1979). A ellos se añadieron las alusiones a los estudios positivos resultantes de lo que se conoce bajo el epígrafe <<problema de las comparaciones múltiples>> (Donner A, 1982; Campillo C, 1996).

En el plano descriptivo restaba por satisfacer la necesidad de disponer de un método sencillo que ayudara al lector --sujeto a la premura por tomar decisiones clínicas prácticas-- a estimar específicamente su magnitud potencial. Al hilo de las investigaciones señaladas, se elaboró con ese objetivo una medida promisoria. La medida diseñada requería conceptualizar el estudio publicado objeto de análisis como perteneciente a un conjunto de investigaciones sobre el mismo tema llevadas a cabo de forma concurrente. Además, exigía conocer su tamaño muestral, así como estimar el número de sujetos participantes en los restantes estudios. El valor estandarizado de esa medida no cuantificaba la magnitud de la

sobrestimación de, digamos por caso, una diferencia o de la fuerza de una asociación, sino la del sesgo de publicación introducido en la estimación del estudio analizado en tantos por ciento (Begg CB, 1985). Combinando en una tabla el tamaño muestral de ese estudio y el número estimado de participantes en ensayos clínicos similares se obtiene la magnitud del sesgo de publicación introducido en la estimación del estudio individual. La medida se concibió para aplicarla a un solo ensayo no aleatorizado y, aunque permite estimar la magnitud del sesgo de publicación en las estimaciones de cada estudio individual, no proporciona una medida global y promediada de un estimador insesgado obtenido al combinar todos los estudios. Además, cualquier estimación del número de participantes en otros estudios concurrentes no publicados se encuentra al conjuro de la subjetividad del lector.

En esta misma línea se desarrollaron otras formulaciones matemáticas del sesgo de publicación basadas en modelos de muestreo truncados (Hedges, 1984; Hedges LV, Olkin I, 1985; Berlin JA, Begg CB, Louis TA, 1989) o en la teoría de la distribución ponderada (Iyengar S, Greenhouse JB, 1988). Hasta la fecha, ninguno de ellos ha recibido una calurosa acogida, probablemente debido a las múltiples limitaciones que presentan: complejidad teórica y computacional, la profunda comprensión que exigen de los modelos en que se fundamentan, y su carácter arcano para la inmensa mayoría de epidemiólogos y, sobre todo, para la mayor parte de los profesionales que se beneficiarían de su utilización (Poynard T, Conn HO, 1985; Dickersin K, Hewitt P, Mutch, Chalmers I, Chalmers TC, 1985). A esta crítica del modelo de Begg cabe añadir que su medida exige suponer que los tamaños muestrales de todos los estudios son iguales, un requisito difícil de satisfacer. No obstante, a pesar de que no representa una corrección formal del sesgo de publicación, ofrece una idea tosca de su magnitud y engrosa el arsenal de herramientas para acotar la exactitud de las estimaciones publicadas.

1. 8 El sesgo de publicación y la financiación de los estudios

La difusión de artículos sobre este tema sirvió de acicate para abordar el sesgo de publicación desde otros ángulos. Las reducciones de la fracción del gasto público destinada a investigación en muchos países está teniendo entre otras consecuencias un aumento sustancial de las subvenciones concedidas a esta actividad por la industria farmacéutica. Ello ha dado cabida a conflictos de interés de los que se han hecho eco los editores de diversas revistas, algunos de los cuales ya exigen a los autores revelar las fuentes de financiación de los estudios presentados a publicación (Relman AS, 1984A; Relman AS, 1984B).

El desconocimiento del efecto que podía ejercer la subvención de investigaciones en su publicación condujo a Davidson a evaluar, conforme a su fuente de financiamiento, los resultados de 107 ensayos clínicos con diseños concurrentes y entrecruzados (*cross-over trials*) publicados en cuatro prestigiosas revistas médicas (Davidson RA, 1986). Ignorando *a priori* la fuente de financiamiento, los artículos se categorizaron como positivos (71%) o negativos (29%) según mostraran o no la mayor eficacia de la nueva opción terapéutica. A continuación, los ensayos clínicos se clasificaron en subvencionados y no subvencionados por la industria farmacéutica. De los 76 que proclamaban la eficacia del nuevo tratamiento, 33 habían sido subvencionados, y de los 31 que no pudieron demostrar su superioridad, sólo 4 estaban subvencionados. Davidson concluyó: «Aunque parece poco probable que exista una conspiración para no publicar resultados no favorables de ensayos clínicos, se puede estar produciendo una exclusión *de facto* de resultados negativos». Si bien la diferencia observada fue estadísticamente significativa y arrojaba luz sobre el problema, el diseño del estudio no podía establecer inequívocamente una asociación causal. Diversos factores pueden actuar como variables de confusión en esa asociación. El no conocer con certeza la

fuelle de financiamiento de algunos ensayos puede introducir un sesgo de clasificación errónea (*misclassification bias*). Las empresas farmacéuticas disuaden a veces la publicación de los estudios negativos de baja potencia estadística que financian. También es legítimo pensar que, en ausencia de restricciones de fondos, subvencionen estudios con tamaños muestrales suficientes para detectar diferencias clínicamente relevantes. El estudio de Davidson, sin embargo, no reveló diferencias entre los tamaños muestrales de los estudios que recibieron subvenciones de la industria farmacéutica y los que las recibieron de otras fuentes. Además, la proporción de estudios positivos que recibieron fondos de otras fuentes fue menor que la de los subvencionados por compañías farmacéuticas. Es probable que la combinación de todos estos factores explique la mayor proporción de estudios positivos financiados por esa industria. Obviamente, tampoco puede descartarse la posibilidad de que no se envíen a publicación los estudios negativos o los positivos contrarios a sus intereses. Eso sólo lo pueden revelar los investigadores tanto de los departamentos de investigación de compañías farmacéuticas como de entidades independientes de esa industria.

La objetividad de las publicaciones depende en parte de la revelación de esta información. La preocupación que despertó esa posible <<exclusión conspirativa>> --no obstante los alegatos publicados por sus detractores (Lauritsen K, Havelund T, Laursen LS, Rask-Madsen J, 1987) y las defensas de sus apologetas (Nicholson PA, 1987)-- ha desembocado en que hoy día cada vez más revistas exijan revelar a los autores la fuente de financiación de sus estudios.

1.9 Nuevos diseños para estimar la existencia del sesgo de publicación. Los registros de estudios y el metanálisis

El interés mostrado en medicina por el estudio del sesgo de publicación se fue fraguando y aumentó a medida que las investigaciones realizadas en distintas especialidades arrojaban resultados. Un estudio que se convirtió en clásico de este tema fue el llevado a cabo por Dickersin y colaboradores (Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith Jr H, 1987). Remedando el estudio de Simes (Simes, 1986), estos investigadores encuestaron a 156 autores de artículos de ensayos clínicos publicados para averiguar en cuántos ensayos habían participado, la cifra de publicados y no publicados y los motivos, así como las características metodológicas más prominentes de los estudios. De los 1312 ensayos en que habían participado, 1041 se publicaron y 271 no. De estos últimos, 33 estaban en curso en el momento de la encuesta, y 238 se habían concluido. El tamaño muestral del 73% de los no publicados era menor de 50, y las razones esgrimidas más sobresalientes para no publicarlos (51%) fueron las siguientes: haber obtenido resultados negativos, falta de interés y tamaño muestral insuficiente. Al comparar los ensayos publicados y los no publicados observaron, respectivamente, que 55,1% frente a 14,6% favorecían el nuevo tratamiento; que en el 22,2% frente al 44,4% no se encontraron diferencias entre la eficacia de los tratamientos investigados, y que en 3,3% frente a 12,9% la eficacia del tratamiento estándar era superior a la de la nueva opción terapéutica.

Este estudio no estuvo exento de limitaciones derivadas del diseño. La muestra de estudios utilizada se obtuvo de una lista de más de 2400 ensayos clínicos proporcionada por uno de los autores. Por consiguiente, y a pesar de su envergadura, podría no ser representativa del universo de ensayos clínicos realizados que cubrió la investigación (1963-1981). Otra fuente de sesgo de selección podría ser el haber obtenido información de ensayos no publicados a partir de ensayos publicados. Las limitaciones de que adolece el

instrumento empleado para obtener los datos impide cerciorarse del grado de representatividad de las muestras analizadas de ensayos publicados y no publicados. Por otro lado, las peculiaridades de la publicación de estos estudios podría ser distinta en especialidades médicas no abarcadas por esa muestra; de ser esto cierto, reduciría la validez externa de los resultados obtenidos. El sesgo de no respuesta (33% no respondieron) también podría afectar los resultados obtenidos, así como el que sólo 49% de los ensayos seleccionados inicialmente pudieron analizarse para los propósitos del estudio. Las respuestas de los entrevistados podrían estar escoradas hacia la sobrerrepresentación, tanto las de aquellos que pensaban que la probabilidad de publicar un ensayo negativo es menor que la de uno positivo, como la de los convencidos de la inexistencia del sesgo de publicación. No obstante estas limitaciones, el estudio permitió confirmar la existencia del sesgo de publicación en más amplias parcelas de las especialidades médicas.

Los hallazgos obtenidos en este y otros estudios similares iban cobrando cada vez mayor importancia a medida que el metanálisis se consolidaba como herramienta de análisis en la investigación médica. De nuevo, las limitaciones intrínsecas a esas investigaciones constituían un escollo para estimar con precisión la verdadera magnitud del sesgo de publicación. Para solventar el problema, Dickersin propuso emplear el diseño de cohortes, por medio del cual se debía seguir el curso de los ensayos clínicos desde su etapa de protocolo hasta su publicación (Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith Jr H, 1987).

En 1986, Simes ideó una estrategia innovadora para indagar la existencia del sesgo de publicación y, casi de forma inopinada, para paliar sus efectos en la revisión de la literatura y, por ende, en la toma de decisiones terapéuticas (Simes, 1986; Simes, 1987). Genéricamente, la alternativa consistió en comparar los resultados de ensayos clínicos publicados con los de ensayos clínicos no publicados incluidos en un registro de ensayos

clínicos sobre el cáncer, *The International Cancer Research Data Bank*. El estudio se centró en el tratamiento de dos neoplasias. Primero, comparó los ensayos clínicos sobre la supervivencia de pacientes con cáncer de ovario avanzado tratadas con un alquilante y la de las que recibieron quimioterapia combinada. A continuación, hizo lo propio con los de la supervivencia de pacientes con mieloma múltiple tratados con una combinación de citotóxicos y la de los que recibieron un alquilante y prednisona. El estudio reveló datos que son sorprendentes a la luz de sus repercusiones en la toma de decisiones y la práctica clínicas. El análisis *combinado* de los ensayos clínicos publicados sobre el cáncer de ovario demostró que el tiempo de supervivencia de las pacientes tratadas con quimioterapia combinada fue significativamente mayor que el de las que recibieron un alquilante (razón de las medianas del tiempo de supervivencia de las primeras en relación con las segundas = 1,16; $P = 0,02$). No obstante, no se encontraron diferencias al combinar los resultados de los ensayos clínicos no publicados incluidos en el registro (razón de las medianas del tiempo de supervivencia = 1,05; $P = 0,25$). Por su parte, al combinar los ensayos sobre el mieloma múltiple también se puso de manifiesto la mayor eficacia de la quimioterapia combinada (razón de las medianas del tiempo de supervivencia = 1,26; $P = 0,04$). Sin embargo, los resultados combinados de los ensayos no publicados registrados revelaron que la magnitud del beneficio de la quimioterapia combinada se redujo notablemente (razón de medianas = 1,11; $P = 0,06$).

A pesar de la reducida validez externa de estos resultados (el análisis se limitó a estudios de dos enfermedades y a un sólo registro), el estudio de Simes ilustró la disparidad de las conclusiones que pueden extraerse al incluir en una revisión de la literatura estudios no publicados. Su principal valor estriba en que innovó la investigación del sesgo de publicación. Primero, por ofrecer una nueva opción para estimar su magnitud, recurriendo a los registros de estudios. Los registros ofrecen una muestra de estudios que no está influida

por el sesgo de publicación, ya que los estudios se incluyen en la etapa de diseño antes de obtener resultado alguno. Segundo, porque esbozó la aplicación del diseño de cohortes a la estimación de la existencia del sesgo de publicación. Conforme a este diseño, los ensayos se seguirían desde el momento de su protocolo o registro hasta una fecha razonable, a fin de dar tiempo para publicarlos. Como exposición se considera la obtención de resultados positivos, y como efecto, su publicación. Posteriormente, podría cuantificarse la fuerza de la asociación entre exposición al factor de riesgo y la publicación o no del estudio. En presencia de sesgo de publicación cabe esperar que el riesgo de los estudios expuestos (positivos) de ser publicados sea mayor que el de los no expuestos (negativos)⁹. Tercero, porque entreabrió las puertas a una solución parcial para contrarrestar sus efectos, que ulteriormente ganaría numerosos adeptos: el registro de los estudios desde el momento de protocolizarlos. En otro estudio similar en que se compararon las eficacias del 5-fluoracilo y del tratamiento coadyuvante se extrajeron las mismas conclusiones que en el de Simes (Moertel CG, 1984).

Como se indicó al principio de esta introducción, los estudios en el área de la psicología llevaron durante años la avanzadilla en la investigación del sesgo de publicación. También en esta disciplina la realización de metanálisis precedió a su introducción en medicina. La revisión crítica de un metanálisis efectuado en 1977 en la esfera de la investigación de resultados psicoterapéuticos dio fiel cuenta tempranamente de las consecuencias que el sesgo de publicación puede acarrear en los metanálisis (Coursol A, Wagner EE, 1986). Casi una década más tarde —a consecuencia del ritmo vertiginoso con que se vienen realizando y publicando metanálisis en medicina— comenzamos a ver

⁹ Sin embargo, el análisis riguroso de esta asociación ha de tener en cuenta el efecto de posibles factores de confusión tales como la calidad metodológica de ambos tipos de estudios.

publicadas en revistas médicas cada vez más llamadas de alerta contra nuevos peligros que encierra ese sesgo.

Conscientes de ello, en la revisión apuntada los mismos autores llamaron la atención sobre la cautela que debía regir la evaluación de los resultados de metanálisis, al comprobar que en el que revisaron sólo el 3% de los estudios incluidos no se habían publicado. Los problemas que los estudios negativos suponen para el sesgo de publicación alentaron a los autores a encuestar a miembros de la *American Psychological Association*. La encuesta corroboró resultados que estudios precedentes habían adelantado: la decisión de enviar estudios a publicación dependía de sus hallazgos (de un total de 129 estudios positivos, 106 se enviaron a revistas). Paralelamente, la encuesta puso de manifiesto que la aceptación editorial también dependía de los resultados: se publicó el 66% de los estudios positivos y sólo el 22% de los negativos.

El contrapunto a este estudio en medicina lo constituye la revisión de 91 metanálisis publicados hasta aquella fecha, que fue llevada a cabo por un grupo de epidemiólogos y matemáticos (Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R, 1987). Un hecho sustancial los alentó a realizar el estudio: a pesar de que entre 1984 y 1986 se publicaron 6000 ensayos clínicos aleatorizados, a juzgar por el número de ensayos indizados en MEDLINE bajo el acápite *asignación aleatoria (random allocation)*, la información sobre su calidad era prácticamente inexistente. Por añadidura, en los escasos estudios realizados tampoco se había considerado el sesgo de publicación. La moderada replicabilidad de los resultados de metanálisis sobre un mismo tema que habían observado (más marcada en la magnitud que en la dirección de los resultados) podría ser atribuible, entre otros factores, a la aplicación de diferentes criterios de inclusión y exclusión en los estudios; variabilidad en los métodos para evaluar, sintetizar y combinar la información procedente de los estudios individuales; empleo de distintos métodos de análisis estadístico

en la combinación de resultados; discrepancias entre las variables de resultado empleadas en cada metanálisis, y a criterios dispares en cuanto a la inclusión de ensayos no aleatorizados y ensayos negativos.

Partiendo de estos hallazgos, los autores citados hicieron hincapié en la pertinencia de elaborar y diseminar guías metodológicas destinadas a homogeneizar los métodos de obtención de datos y análisis de estos estudios. Además, para obtener resultados combinados no sesgados, barajaron la posibilidad de incluir en los metanálisis estudios no publicados, resúmenes y la información recabada directamente de autores de estudios.

Era preciso escrutar cautelosamente esta solución potencial. Si los estudios publicados difieren sistemáticamente de los no publicados en algún aspecto relacionado con su calidad, se incurriría en un error al realizar metanálisis con todos los datos disponibles. Las perspectivas que abría el metanálisis también podrían verse eclipsadas si en ellos sólo se incluyeran estudios positivos y no se obtuvieran estimaciones de la magnitud del sesgo de publicación (Mosteller F, Gilbert JP, McPeck B, 1980; Minerva, 1983; Chalmers I, 1985). La constatación en estudios precedentes de que la dirección de los resultados era una de las fuentes potenciales de las diferencias entre los estudios publicados y los no publicados se convierte en el argumento suficiente y necesario para recomendar la inclusión en metanálisis de sólo aquellos estudios que aporten datos metodológicamente rigurosos al margen de que se hayan publicado.

Estos resultados condujeron a diversos investigadores a dejar constancia en la literatura de la perentoria necesidad de incluir en los registros correspondientes ensayos clínicos no publicados, así como de crear registros prospectivos de ensayos planificados y concurrentes, con objeto de minimizar los efectos del sesgo de publicación en las conclusiones extraídas por medio de revisiones de ensayos clínicos. La *Oxford Data Base of Perinatal Trials* fue presentada en la literatura como el paradigma de un registro de

estudios a estos efectos (Chalmers I, Hetherington J, Newdick, et al., 1986). En algunos registros de investigaciones de otras especialidades se comenzó a buscar ensayos clínicos activamente y a registrar los no publicados conforme a determinados criterios de inclusión de índole metodológica (UK Cancer Trials Register, 1982; Verstraete M, 1984).

En 1988, las propuestas de registrar los protocolos de los ensayos clínicos recibieron el espaldarazo necesario por medio de artículos, notas de los editores y debates mantenidos en reuniones organizadas por diversas las sociedades científicas. En varios artículos publicados se insistió en las ventajas del registro, así como en su capacidad de contrarrestar el efecto del sesgo de publicación. También se pormenorizó la información que debía incluirse en los estudios registrados y los tipos de registros que podían crearse (por enfermedades, especialidades, estudios mono y multicéntricos). Además, se instó a los editores de revistas a promover el registro y a exigir a los autores la inclusión del número de registro de sus estudios en los artículos que publicasen. Dicha citación permitiría a los lectores acceder fácilmente a las bases de datos de los registros y, por ende, a la información completa del estudio. Ello, aunado a las mejoras en los sistemas de indización de revistas (vgr., las incluidas en la *National Library of Medicine* de los Estados Unidos, *Current Contents*, *Excerpta Medica* y otras bases bibliográficas), y al refinamiento de la terminología específica utilizada, facilitaría, en suma, el acceso al ingente número de estudios publicados y de los registrados que no se publican (Meinert CL, 1987; Dickersin K, 1987; Piantadosi S; Byar DP, 1988). También se analizaron los beneficios que podría rendir la publicación de los protocolos de los estudios (Piantadosi S, Byar DP, 1987): su revisión podría disuadir a los investigadores de proseguir la realización de estudios irrelevantes o mal diseñados e incurrir en repeticiones innecesarias.

Además del efecto disuasorio de la publicación sesgada de estudios positivos, el registro ofrece ventajas adicionales: podría promover la colaboración entre investigadores;

facilitar la realización de metanálisis *no sesgados* en lo referente al sesgo de publicación; aportar una fuente sustantiva de datos metodológicos y, finalmente, permitir discernir a los comités de aprobación de ensayos clínicos entre replicación de estudios (un criterio de causalidad) y su duplicación (una dilapidación de recursos cuya prevalencia es inaceptablemente elevada, sobre todo habida cuenta de la limitación de recursos destinados a la investigación).

Aunque no todos acogieron con buenos ojos la idea de crear revistas especializadas en la publicación de protocolos (y de hecho no ha cristalizado hasta la fecha), algunos investigadores justificaron --acertadamente-- su validez y factibilidad por los beneficios señalados. Los primeros resultados de los estudios en curso sobre la efectividad y eficiencia de los registros citados, junto con la amplia difusión de argumentos esgrimidos por sus paladines, concretaron la propuesta, que fue aceptada y se difundió con celeridad entre ciertos sectores de la comunidad científica.

Es preciso hacer hincapié en que el mero conocimiento de las amenazas que el sesgo de publicación plantea para la toma de decisiones médicas y la posibilidad de acceder a registros que contienen estudios rigurosos no publicados no constituyen un garante de la validez de las conclusiones de cualquier revisión, cualitativa o cuantitativa de la literatura (Collins R, Gray R, Godwin J, Peto R, 1987). De hecho, diversos estudios realizados en la última década han permitido comprobar, por un lado, que muchas revisiones refuerzan los efectos del sesgo de publicación (pueden hacerlo todas aquellas basadas exclusivamente en estudios publicados que sean consecuencia de este sesgo). Por otro, muchas revisiones son poco confiables por motivos tales como la baja representatividad de los estudios incluidos en la revisión; sesgos introducidos en la revisión al preponderar las opiniones del revisor sobre los resultados de los estudios revisados --a veces contrarias a los hechos; ausencia de análisis del rigor científico de los estudios revisados y falta de ponderación consiguiente de

las conclusiones de la revisión; baja repetibilidad y resultados contradictorios de revisiones llevadas a cabo sobre un mismo tema.

Frente a estos hallazgos, numerosos investigadores y editores de revistas han tomado medidas de muy diversa estirpe. Valga citar a modo de ejemplo los nuevos requisitos que exigen cada vez más revistas para elaborar artículos de revisión, la inclusión de resúmenes estructurados de este tipo especial de artículos, la exigencia en ellos y en los metanálisis de realizar revisiones conforme a criterios estrictos destinados a pautar las búsquedas bibliográficas, la aplicación de criterios estrictos de inclusión y exclusión de estudios publicados o registrados, la extracción de información y su síntesis sistemáticas, y la formulación de conclusiones ponderadas teniendo en cuenta factores tales como la calidad de los estudios revisados. Todo ello ha desembocado, al cabo, en modelos normativos actualmente muy elaborados, cuya aplicación garantiza en gran medida la obtención de revisiones de la literatura razonablemente confiables. Estos modelos se conocen como revisiones sistemáticas (Light RJ, Pillemer DB, 1984; Cooper H, Hedges LV, 1994; Powe NR, Turner JA, Maklan CW, Ersek M, 1994; Slavin RE, 1995; Cook DJ, Sackett DL, Spitzer WO, 1995).

1. 10 Nuevos diseños para investigar el sesgo de publicación

En 1987, la psicología aportó datos innovadores para el diseño de los estudios del sesgo de publicación, como los de una encuesta realizada a 140 miembros de la Society for Menstrual Cycle Research residentes en Estados Unidos y Canadá. De los 143 estudios analizados, 73 se habían publicado y 70 no (Sommer B, 1987). De estos últimos, 42 eran muy recientes y sus responsables todavía no habían decidido enviarlos a publicación en el momento de la encuesta. El 73% (22 de 30) y el 54% (14 de 26) de los estudios positivos y negativos, respectivamente, se habían publicado. No se observaron diferencias en cuanto a

la posibilidad de que los estudios negativos tendieran a ser publicados con mayor asiduidad en libros que en revistas. Tampoco se observaron diferencias en las proporciones de publicación al considerar el tipo de estudio (experimental, informes, revisiones de la literatura, clínicos) ni según el aspecto considerado del ciclo menstrual. Sin embargo, los investigadores que habían efectuado más de una investigación tendían a enviar a publicación más artículos que los investigadores novicios. Acaso este último hecho pueda ser tributario de la menor experiencia de estos últimos en investigación y en publicación.

Las limitaciones del estudio pueden resumirse diciendo que buena parte de los que investigaban el ciclo menstrual no eran miembros de la sociedad mencionada. Por tanto, los resultados sólo pueden extrapolarse a los residentes en Estados Unidos y Canadá. La pertenencia a esa sociedad podría traducir una actividad investigadora más intensa entre sus miembros respecto a los no miembros. Es probable que la encuesta no captara a aquellos investigadores que, si bien finalizaron algunos estudios, no publicaron los negativos por inexperiencia o desconocimiento. El sesgo de recuerdo también pudo haber desempeñado algún papel: algunos encuestados pudieron haber olvidado estudios realizados con bastante anterioridad cuyos resultados fueron negativos o poco concluyentes. Además, se descubrió que la afiliación a la sociedad mencionada y el grado de desarrollo profesional de los investigadores fueron mejores factores predictivos de la publicación final de un estudio que el grado de significación estadística de los resultados. Los investigadores más jóvenes pueden desalentarse más fácilmente y no publicar estudios negativos; los que han acumulado experiencia en investigación probablemente sean más avezados en los entresijos del proceso de publicación¹⁰.

¹⁰ A este respecto cabe mencionar el denominado *efecto Mathew*, un sesgo que consiste en la mayor posibilidad de que se publique un artículo escrito por un investigador prominente, con reputación o que proceda de una institución de renombre que uno elaborado por un desconocido (Lock S, 1985).

No obstante sus limitaciones, este estudio introdujo una innovación decisiva en el diseño de los estudios del sesgo de publicación: analizar, además del grado de significación estadística, el potencial que pueden tener otros factores de introducir sesgos de publicación. Así, Sommer, al tiempo que analizó la asociación entre significación estadística de estudios sobre el ciclo menstrual y su publicación, también tuvo en cuenta la posible asociación entre el estado de publicación y el tipo de diseño utilizado, el tema específico sobre el que versaba la investigación y la afiliación y experiencia de los autores.

Otro estudio innovador --relacionado con el anterior por el tipo de variables analizadas-- se efectuó para averiguar si existía asociación entre la dirección de los resultados (positivos o negativos) y variables del investigador y metodológicas (Swaen GMH, Meijers JMM, 1989). La investigación se circunscribió al ámbito de los estudios retrospectivos. Para realizarlo, se seleccionaron seis revistas de renombre internacional en el campo de la epidemiología ocupacional. Entre las variables del investigador se observó que en los estudios realizados por epidemiólogos pertenecientes a instituciones públicas (gobierno y universidades) el porcentaje de estudios positivos sobrepasó al de los financiados por la industria. Asimismo, el de aquellos que encontraron asociaciones entre ciertas variables y la exposición a agentes químicos también fue más elevado que en los que no las detectaron. Dicho porcentaje también fue más alto en los estudios llevados a cabo en los Países Bajos que en los efectuados en los Estados Unidos. En cuanto a las variables metodológicas, no se apreció asociación alguna entre el resultado del estudio y el tamaño muestral, el haber utilizado como criterio de selección un período de exposición breve ni la duración del seguimiento. Sin embargo, en los que utilizaron algún grupo de comparación y en los que tuvieron pérdidas de participantes menores de 1% el porcentaje de estudios positivos fue más elevado que en los que no incorporaron grupo control o las

proporciones de pérdidas en el seguimiento excedieron 1% (razón de posibilidades, en inglés *odds ratio*, o $RP = 2,66$; IC95%: 1,96 - 3,36).

Es de destacar la ausencia de asociación entre tamaño muestral y dirección de los resultados. Según los autores, esto podría indicar que los cálculos de los tamaños muestrales no debieron de ser el único elemento en que se fundamentó la decisión de proseguir las investigaciones y enviarlas a publicación, especialmente en virtud de las restantes asociaciones encontradas. En una carta al editor (McNamee, 1989) se adujo que ese hallazgo podría ser el resultado del sesgo de publicación introducido por los autores o los editores. Tanto unos como otros pudieron mostrarse renuentes a enviar o a aceptar para publicación los de bajo tamaño muestral, a menos que detectaran diferencias estadísticamente significativas. Los estudios negativos con muestras grandes podrían considerarse más publicables. Por consiguiente, la mayor selección de estudios con muestras pequeñas en virtud de los resultados tendería a ocultar dicha asociación.

Este estudio constituye, al igual que el de Sommer, otro avance metodológico indirecto en la investigación del sesgo de publicación, en la medida en que, como se indicó, también se tomaron en consideración otras variables (del investigador y metodológicas del estudio) que pudiesen estar asociadas con el sesgo.

Acaso la significación estadística no explique por sí sola el sesgo de publicación, aserto en cierto modo respaldado por las otras asociaciones detectadas. En cualquier caso, no es posible descartar la existencia de ese sesgo en la serie de trabajos escrutada a causa del diseño transversal empleado en la investigación –todos ellos habían sido publicados con o sin la influencia de dicho sesgo. Para ello, sería preciso discernir la exposición del desenlace (publicación o no de un artículo), saber cuántos estudios positivos y negativos (expuestos y no expuestos) quedaron en el archivador, así como las causas por las que no se publicaron.

La estimación del sesgo de publicación se fue ampliando a otras especialidades y a diseños distintos de los ensayos clínicos. En esta línea se encuentran los hallazgos de Vanderbroucke, (Vanderbroucke JP, 1988) quien interpretó los resultados de un estudio sobre el riesgo de los fumadores pasivos de desarrollar cáncer de pulmón (Wald NJ, Nanchackal K, Thompson SG, et al 1986; Clarke MJ, Stewart LA, 1994). En el mismo se estimó que el riesgo de este cáncer en los fumadores pasivos era 30% más elevado que el de los no expuestos al humo del tabaco. Mantel indicó que ese dato podría atribuirse al sesgo de publicación (Mantel, 1987) y Vanderbroucke intentó corroborarlo aplicando una técnica novedosa en medicina procedente de la psicología: el *gráfico en embudo* (en inglés, *funnel display*), descrita originalmente por Light y Pillemer (Light RJ, Pillemer DB, 1984)¹¹.

¹¹ El *gráfico en embudo* reposa en un principio estadístico notorio: a medida que aumenta el tamaño muestral, los estadísticos muestrales tienden a estimar el parámetro poblacional con mayor precisión. O, parafraseando, cuanto mayor es el tamaño muestral, menor resulta la variabilidad atribuible al error muestral. La presencia o ausencia de sesgo de publicación en un conjunto de estudios puede sospecharse por la forma que adopta el gráfico, que se construye del siguiente modo. En el eje vertical se representan los tamaños muestrales de cada uno de los estudios considerados y en el horizontal, las estimaciones que brindan de la magnitud del efecto, o sea, del parámetro que estiman (bien sean diferencias, por ejemplo, entre proporciones o medias). Si el conjunto de estudios proporciona estimaciones insesgadas del efecto real, es decir, en ausencia de sesgo de publicación, se espera observar una imagen en embudo. Esto es, los diversos puntos del gráfico tienden a distribuirse alrededor del valor del efecto real más o menos simétricamente, aproximándose a dicho valor en la zona de tamaños muestrales grandes (zona superior del gráfico) y alejándose de él a medida que se reduce el tamaño muestral (mitad inferior del gráfico). El grado de dispersión de los puntos refleja el error muestral aleatorio de los estudios. Cuando existe sesgo de publicación, los estudios en que no se han detectado diferencias o efectos estadísticamente significativos no se publican y, por tanto, no aparecen en el gráfico. El tamaño muestral y, por ende, la potencia estadística de estos estudios suelen ser bajos. Consiguientemente, en el gráfico el sesgo (la ausencia de esos estudios) se manifiesta como una zona sin puntos en la zona inferior (de bajo tamaño muestral).

Posteriormente, el gráfico se amplió para detectar la presencia del sesgo cuando el parámetro estimado en los estudios es una medida de la fuerza de asociación (riesgos relativos, razones de productos cruzados, tasas de riesgos instantáneos (*hazard rates*), o cualquier medida que se ajuste a una distribución asintótica normal). En este caso, en el eje vertical se representa en papel milimétrico el *ln* de la medida empleada (a fin de normalizar su distribución) o su valor no transformado (en cuyo caso debe emplearse papel semilogarítmico). En el eje horizontal, se representa el error estándar del *ln* de la medida empleada, por los mismos motivos señalados. El empleo del error estándar en sustitución del tamaño muestral se justifica porque el primero suele ser inversamente proporcional al segundo y expresa igualmente la variabilidad muestral de las estimaciones. En ausencia de sesgo, los puntos se sitúan alrededor del verdadero valor de la medida del efecto. En su presencia, por el contrario, los estudios en que no se demuestre la existencia de asociación (las medidas del efecto se aproximan al valor 1,0 y sus intervalos de confianza contienen este valor) no se publican y no aparecen en el gráfico. Asimismo, estos estudios suelen tener bajo tamaño muestral y, por tanto, errores estándar elevados. Esto explica que en el gráfico observemos igualmente un área sin puntos en la zona de errores estándar elevados situada alrededor del valor 0,0 del estimador del efecto.

A pesar de que el gráfico puede ser considerado por algunos como un apero burdo de estimación, la solidez de los principios en que descansa, la utilidad que ha demostrado tener en numerosos estudios y su fácil aplicación, lo convierten en una técnica útil para los interesados en realizar revisiones y metanálisis más rigurosos. No obstante, debe precisarse que no estima con certeza la existencia del sesgo de publicación, sino que solamente la sugiere.

Aplicándolo, Vanderbroucke corroboró las sospechas de Mantel, dado que el perfil del gráfico era congruente con la existencia de sesgo de publicación en los estudios realizados con hombres. En los estudios efectuados con mujeres dicho perfil no la sugería. Por ello, recomendó que las revisiones de estas investigaciones se estratificaran según el sexo, a fin de poder estimar su existencia en los estudios llevados a cabo con hombres. Finalmente, propuso realizar un nuevo estudio en los hombres con una muestra grande y urgió a los investigadores a revelar los resultados de sus estudios no publicados y con poca potencia estadística con objeto de disipar las dudas relativas a ese sesgo.

El gráfico en embudo, junto con los modelos de estimación presentados anteriormente, sirvieron de punto de anclaje para la elaboración de otro modelo, también desarrollado para estimar la magnitud del sesgo de publicación (Berlin JA, Begg CB, Louis TA, 1989). Este modelo es prácticamente una versión cuantitativa del gráfico en embudo, pues utiliza el tamaño muestral de los estudios revisados para descubrir el sesgo (de hecho estima la magnitud del sesgo en función del tamaño muestral). La presunción de esta relación es lógica, ya que abundan razones para pensar que el sesgo de publicación es más marcado en los estudios con muestras pequeñas que en los de tamaño muestral elevado. La suposición en que se fundamenta es que la magnitud del efecto estimado en los estudios es independiente de las características de su diseño y, específicamente, de su tamaño muestral.

Expresado en otras palabras, las magnitudes del efecto estimado son similares sea cual fuere el tamaño muestral. Por consiguiente, cualquier asociación que se encuentre entre la magnitud del efecto y el tamaño muestral es atribuible al sesgo de publicación. Como el modelo predice que el sesgo de publicación está fuerte e inversamente relacionado con el tamaño muestral, cabe esperar que se observen efectos pronunciados en los estudios con bajo tamaño muestral y pequeños en los de tamaño muestral grande. Las diferencias entre las estimaciones de la magnitud media de esos efectos que calcula el modelo equivalen a la estimación de la magnitud del sesgo de publicación. El modelo obliga a analizar el sesgo exclusivamente con estudios publicados, con lo cual elude el escollo que representa la búsqueda de estudios no publicados. Una ventaja adicional es que permite considerar simultáneamente otras variables —diversas características del diseño supuestamente independientes de las diferencias o del efecto observado— que podrían actuar como factores de confusión de la asociación entre sesgo de publicación y tamaño muestral; es decir, el modelo permite asimismo analizar el efecto que ejerce cada una de esas variables en el sesgo y ajustar la asociación entre sesgo de publicación y tamaño muestral según dichas variables.

Al aplicarlo a una muestra de 246 ensayos clínicos publicados, se obtuvieron los siguientes resultados. El sesgo de publicación afectó con mayor frecuencia a los estudios con pequeño tamaño muestral. La asociación detectada entre tamaño muestral y magnitud de los efectos notificados en los estudios fue más marcada en los ensayos no aleatorizados que en los aleatorizados. Asimismo, los estudios monocéntricos eran más sensibles a la influencia del sesgo que los multicéntricos. Esto implica que los ensayos pequeños que demuestran grandes efectos tienden a publicarse con preferencia, en tanto que los ensayos con muestras grandes tienden a publicarse independientemente del resultado, lo cual

confirma resultados de estudios consignados anteriormente. Por último, la revista en que se publicaron los estudios no pareció estar relacionada con el sesgo de publicación.

Como principales limitaciones del modelo sobresalen las siguientes. En primer lugar, encierra varios supuestos, obligados por razones de conveniencia analítica, que deberían validarse. Desde un punto de vista analítico, su formulación formal es extremadamente compleja, lo cual reduce su utilización a profesionales con alto nivel de conocimientos matemáticos. Un lector promedio no podría aplicarlo en sus revisiones periódicas de la literatura. Tampoco puede contemplarse como una herramienta que permita ajustar las estimaciones de los efectos en virtud de la magnitud del sesgo de publicación. No obstante, podría ser útil —como el gráfico en embudo, su análogo menos formal— para detectar simplemente la presencia del sesgo al realizar una revisión sistemática de la literatura o antes de llevar a cabo un metanálisis.

Recientemente, Begg y Mazumdar han incorporado al arsenal de métodos estadísticos de estimación del sesgo una variante que marca otro contrapunto estadístico del gráfico en embudo y ofrece la ventaja de ser menos complejo que el anterior (Begg CB, Mazumdar M, 1994). Como se indicó, el perfil del gráfico en embudo se modifica en presencia de sesgo de publicación, es decir, cuando se viola el supuesto de independencia entre el tamaño muestral (o el error estándar del estimador del efecto) y el estimador del efecto. Esos autores lo expresaron de otro modo. Por definición, en ausencia de sesgo los valores del estimador del efecto y el tamaño muestral no están correlacionados y, por el contrario, cuando existe, puede expresarse como una correlación entre esas dos variables. Partiendo de ese principio, formularon una variante de la prueba de correlación de rango (*rank correlation test*). Dicha prueba estima concretamente la existencia de correlación entre el estimador del efecto de los distintos estudios (diferencias entre proporciones de respuestas, riesgos relativos, razones de productos cruzados, tasas de riesgos instantáneos

(*hazard rates*) o cualquier medida que se ajuste a una distribución asintótica normal) y su varianza. La construcción de una prueba válida de esta estirpe (basada en la *tau* de Kendall y en el supuesto de que la probabilidad de publicación es una función del valor *P* observado) exige estabilizar las varianzas, estandarizando los estimadores de los efectos (es decir, mediante su conversión en valores *z*). En este método se emplean las varianzas de los estimadores del efecto, porque en muchos casos éstas son inversamente proporcionales a los tamaños muestrales de los estudios. Por consiguiente, la prueba equivale a correlacionar los estimadores de los efectos de los distintos estudios con sus tamaños muestrales. En síntesis, si existe sesgo de publicación, el resultado de la prueba de correlación será estadísticamente significativo, y viceversa.

La principal ventaja de este método es su sencillez. Como desventajas cabe citar su variable potencia estadística, que depende del número de estudios revisados y de la amplitud del recorrido de sus varianzas, el mecanismo de selección de los estudios y la verdadera magnitud del efecto que pretenden estimar. En los metanálisis y revisiones sistemáticas que habitualmente se realizan en medicina su potencia es moderada, ya que el número de estudios que se incluyen en ellos suele oscilar entre 25 y 75.¹² No permite ajustar el valor combinado de la estimación del efecto en función de la magnitud del sesgo. Tampoco orienta sobre el modo de proceder cuando los resultados dan pruebas palmarias de su existencia.¹³ Estas limitaciones no son óbice para emplear el método en la detección *grosso modo* del sesgo. Por consiguiente, debe circunscribirse al ámbito de aplicación del

¹² De esta limitación dimana la imposición de no descartar el sesgo cuando el resultado de la prueba no es estadísticamente significativo. Para superarla, se han propuesto diversos desarrollos matemáticos que exceden el alcance de esta introducción (Iyengar S, Greenhouse JB, 1988; Dear KGB, Begg CB, 1992; Hedges LV, 1992).

¹³ Ante esta eventualidad, uno puede tratar de calcular un estimador global de la magnitud del efecto ajustado según la magnitud del sesgo de publicación, aplicando el método que se juzgue apropiado entre los mencionados (a los que en esta instancia debe añadirse el de Rubin DB, 1990). Otra opción es concluir que la marcada verosimilitud del sesgo y su notable magnitud invalidan cualquier combinación o análisis postrero.

gráfico en embudo y aplicarse en las etapas de comprobación de supuestos que preceden a la combinación cuantitativa de resultados en el metanálisis.

Posteriormente, se presentó otro método estadístico para estimar la magnitud del sesgo de publicación, si bien con aplicaciones muy acotadas (Sugita M, Kanamori M, Izuno T, Miyakawa M, 1992). El modelo permite ajustar la *RP* global estimada en un metanálisis teniendo en cuenta tanto las *RP* de los estudios publicados como las de los no publicados. Las *RP* de estudios no publicados y sus intervalos de confianza del 95% se obtienen indirectamente de las *RP* de los estudios publicados y sus intervalos de confianza de 95%, mediante complejos desarrollos matemáticos. El modelo puede utilizarse cuando las *RP* de los estudios en un área determinada de investigación son homogéneos (DerSimonian R, Laird N, 1986).

Como corolario es posible afirmar que en diversas especialidades médicas la estimación del sesgo de publicación transita por un terreno abonado por la especulación que subyace a los modelos descritos. Por ello, su utilidad y aplicabilidad están en buena parte atenazadas por la validez de sus suposiciones y la formación estadística del profesional que los utiliza.

Otros estudios aportaron métodos novedosos no matemáticos para estimar el sesgo de publicación. En un sondeo realizado desde el *Oxford Database of Perinatal Trials* (Hetherington J, Dickersin K, Chalmers I, Meinert CL, 1989) se encuestó por correo a 42 000 obstetras y pediatras de 18 países a fin de identificar *retrospectivamente* ensayos clínicos en perinatología no publicados y a establecer las bases para determinar la existencia del sesgo de publicación. El trabajo se limitó a documentar los métodos empleados para llevar a cabo la encuesta, notificar la proporción de respuesta y extraer conclusiones sobre medidas parciales encaminadas a minimizar el sesgo. Obstetras y ginecólogos de 17 de los 18 países seleccionados estuvieron dispuestos a proporcionar

información sobre 481 ensayos clínicos no publicados. De ellos, 453 devolvieron la información solicitada. El 12% de los ensayos no publicados notificados no cumplían los criterios de selección. De los 395 ensayos restantes, 143 habían finalizado el reclutamiento de participantes y todavía no habían enviado a publicación los resultados. De estos últimos, 125 lo habían suspendido en los dos años precedentes. De los 453, 193 todavía estaban incorporando pacientes al estudio en el momento de la encuesta y 59 aun no habían comenzado a hacerlo.

El sondeo reveló que, incluso entre los obstetras y ginecólogos que habían participado en investigaciones clínicas, sólo una baja proporción también lo habían hecho en ensayos aleatorizados. Las proporciones de respuestas de otra encuesta parecida fueron muy similares (Shadish W, Doherty M, Montgomery LM, 1989). La razón entre el número de estudios no publicados y publicados fue 1:128 (Chalmers I, Hetherington J, Newdick M, Mutch L, Grant A, Enkin M, et al., 1986). Por consiguiente, y conforme a encuestas mucho más restringidas que la descrita (Simes RJ, 1986; Simes RJ, 1987), esta razón oscila entre 1:128 (esta última cifra fue considerada por los autores una subestimación, dado que entre 1940 y 1984 se habían publicado al menos 2300 informes de ensayos clínicos) y 1:1 (Simes RJ, 1986). Empero, la mayor parte de los estudios en que se ha intentado estimar han arrojado cifras comprendidas entre 10:1 y 1:1 (Smith LM, 1980; Shadish W, Doherty M, Montgomery LM, 1989; Hetherington J, Dickersin K, Chalmers I, Meinert CL, 1989; Simes RJ, 1987; Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith Jr H, 1987; Sommer B, 1987). También se comprobó que muchos ensayos de alta calidad metodológica que no se habían publicado tampoco fueron notificados en la encuesta por sus responsables. Por algún motivo, estas personas no consideraron interesante o conveniente revelar esos datos.

Los resultados de esta importante encuesta son sumamente útiles a la hora de escoger los métodos más adecuados para investigar el sesgo de publicación. A juzgar por

ellos, se sugiere que la identificación *retrospectiva* de ensayos no publicados (y probablemente de estudios observacionales) es poco útil para escrutar el sesgo de publicación. Es probable que si los autores de estudios no publicados reparan en la importancia que revisten sus resultados, serán menos renuentes a facilitarlos.

Por otro lado, la encuesta respalda las propuestas ya expuestas de registrar *prospectivamente* los ensayos clínicos desde su protocolización. Según sus autores, la investigación rigurosa de este sesgo no exigiría registrar la mayor parte de los ensayos que se lleven a cabo. Lo que se precisa es contar con una cohorte de ensayos registrados de una especialidad médica concreta antes de conocer sus resultados. Si las conclusiones extraídas antes de revisarlos difieren de las obtenidas después de analizar todos los ensayos asequibles de esa especialidad, la existencia del sesgo de publicación puede convertirse en una explicación plausible de la discrepancia.

Otra forma propuesta para indagar la existencia del sesgo consistió en investigar la proporción de estudios publicados al inicio en forma de resúmenes que posteriormente se publicaron como artículos completos (Dickersin, 1990; Dudley HAF, 1978; Goldman L, Loscalzo AF, 1980; Meranze J, Ellison N, Greenhow DE, 1982). En conjunto, los datos revelaron con bastante congruencia que sólo entre 30 y 60% de los resultados de estudios pediátricos publicados en forma de resúmenes aparecen en revistas como artículos de estudios finalizados. También se observó que mientras que entre 13 y 22% de los resúmenes de estudios pediátricos enviados a congresos y rechazados se publican, de 49 a 54% de los aceptados para presentación se publican (McCormick MC, Holmes JH, 1985). Si bien puede presumirse que la calidad de los estudios aceptados en esos congresos es más alta que la de los rechazados, también es legítimo sospechar que ello está dejando traslucir el mayor interés editorial en los resultados de los aceptados.

En la misma línea de investigación, un estudio de seguimiento constató que ni la calidad ni los resultados de los estudios juzgados en los resúmenes indizados en la *Oxford Database of Perinatal Trials* (37% de los cuales se publicaron entre 1940 y 1984) se asociaron con su publicación subsiguiente como artículos completos (Chalmers I, Adams M, Dickersin K, et al., 1990). Sólo el tamaño muestral se asoció con la aceptación o el rechazo editorial. Como este estudio no incluye información sobre el proceso que transcurre entre el análisis de sus resultados y la decisión de publicar, en él no es posible excluir la influencia del sesgo de publicación.

En otra investigación de resúmenes acometida para averiguar si los estudios que no demuestran la existencia de efectos adversos a la cocaína durante el embarazo se publican más que aquellos que revelan esos efectos, se observó que de 49 resúmenes de estudios positivos y de 10 negativos se publicaron, respectivamente, 28 y 1 (Koren G, Shear H, Graham K, Einarson T, 1989). Los negativos tendían a verificar con mayor frecuencia el consumo de cocaína, la variable independiente más importante en ese tipo de estudios. La calidad metodológica de los negativos fue superior, lo cual refuerza la idea de que los negativos no se rechazaron por deficiencias metodológicas, sino probablemente por el sesgo que subvalora la falta de detección de reacciones adversas.

1.11 Factores de riesgo del sesgo de publicación

El reconocimiento internacional de la importancia que iba cobrando este tema, aunado al consenso sobre la necesidad de buscar soluciones a los problemas que afronta la revisión editorial, desembocaron en el Primer Congreso Internacional sobre Revisión Editorial, que se celebró en Chicago del 10 al 12 de mayo de 1989 (Sharp DW, 1990; Chalmers TC, Frank CS, Reitman D, 1990; Chalmers I, 1990; Chalmers I, Adams M, Dickersin, et al, 1990; Agency for Health Care Policy and Research, 1990). Aunque

numerosos investigadores y editores habían considerado el sesgo de publicación como fenómeno notorio y de etiología conocida, en realidad sólo se disponía de pruebas parciales de su existencia y la ejecución de medidas encaminadas a impedir su influencia estaba prácticamente en ciernes. Como se demostraría más adelante, las ideas barajadas en dicho congreso contribuyeron a estimular la adopción de medidas sistemáticas, consensuadas y efectivas para minimizar los efectos del sesgo.

Los corolarios que se derivan de los debates que se mantuvieron en dicho congreso son importantes y pueden resumirse del siguiente modo. El intrincado proceso de notificación de resultados parece depender del concurso de numerosos factores. Las preferencias de autores, revisores de revistas y editores someten a los estudios a la influencia del sesgo de publicación. Este error sistemático influye en la publicación de investigaciones de distintas especialidades médicas. Aunque su magnitud no se conoce en todos los casos, la comunidad médica dispone cada vez de más herramientas para estimarla y conocer sus efectos y causas. Tal parece que existen otros elementos –además de la significación estadística– que podrían explicarlo (Begg CB, Berlin JA, 1989).

En dicho congreso se ampliaron notablemente las perspectivas del análisis del sesgo de publicación al presentarse un estudio que lo categorizó en tres tipos: *sesgo de prepublicación* –que incluye aquellos factores (ignorancia u omisión de otras investigaciones, intereses creados, avaricia o pereza por publicar) que pueden influir en el diseño y la realización de una investigación y, por tanto, en su subsiguiente envío a publicación; el *sesgo de publicación* propiamente dicho –el que introducen revisores y editores al no utilizar la calidad de un estudio, en sentido lato, como criterio principal de las decisiones editoriales–, y el *sesgo de postpublicación* –que se basa en la exclusión sistemática de revisiones y metanálisis de estudios no publicados de alta calidad

metodológica y pertinencia, fenómeno cuya prevalencia ha aumentado con el tiempo. Dicha categorización permitió armonizar las soluciones barajadas en décadas anteriores.

Para minimizar el sesgo de prepublicación, es preciso hacer hincapié en las normas que deben gobernar las investigaciones que pretendan publicarse en revistas científicas, incluida la revisión íntegra de los estudios relacionados, al margen de que se hayan publicado o no. Debe corregirse, asimismo, la dualidad de normas que se aplican en la valoración de los ensayos controlados y los no controlados. La estimación de la efectividad o eficacia de tecnologías médicas debe realizarse antes de su *adopción* (Rogers EM, 1983), mediante ensayos controlados y bajo consentimiento informado tan estrictos como los que rigen la realización de ensayos clínicos. Además, y acaso la medida más importante, los investigadores no deben decidir enviar estudios para publicación exclusivamente sobre la base de la significación estadística de los resultados.

Como medidas destinadas a minimizar el sesgo de publicación propiamente dicho destacan el reconocimiento por parte de los revisores de sus intereses creados, la necesidad de mejorar la revisión editorial, y la adopción de la calidad como criterio cardinal de publicación; esto supone abandonar criterios prevalecientes hasta la fecha, como la fuerza, la dirección o la significación estadística de los resultados.

Por último, el sesgo de postpublicación sólo puede evitarse efectuando revisiones y metanálisis basados en evaluaciones íntegras y rigurosas de los estudios pertinentes y obviando la significación estadística como criterio de inclusión cardinal en la combinación de estudios (Light RJ, Pillemer DB, 1984; Cooper H, Hedges LV, 1994; Powe NR, Turner JA, Maklan CW, Ersek M, 1994; Slavin RE, 1995; Cook DJ, Sackett DL, Spitzer WO, 1995).

En el congreso se señaló, además, que la <<mala conducta>> en ciencia se ha conceptualizado tradicionalmente como falsificación deliberada de datos, cuando su

omisión, por su mayor prevalencia, puede acarrear mayores peligros. Como parangón de esa <<mala conducta>> cabe mencionar la omisión de resultados estadísticamente no significativos o considerados "poco interesantes" y la baja proporción de estudios inicialmente preparados en forma de resumen que terminan publicándose como estudios completos.

Los estudios realizados desde la clausura de dicho congreso pueden utilizarse como un indicador orientativo de los progresos realizados para solventar el problema. Dicha solución exige, entre otras acciones, aumentar la concienciación de las partes afectadas (investigadores-autores, revisores y editores), mejorar la difusión de información pertinente, sistematizar las normas de revisión y de decisión editorial, utilizar métodos científicos para llevar a cabo revisiones íntegras de la literatura y metanálisis, galvanizar la creación y el mantenimiento de registros prospectivos actualizados de estudios, así como dirimir las responsabilidades de los comités de ética en los ensayos clínicos (Muir Gray JA, 1995).

1.12 El enfoque de riesgo

Desde el punto de vista metodológico, una innovación cardinal contribuyó a perfilar los métodos actuales de estimación del sesgo de publicación: la identificación, análisis y categorización de sus factores de riesgo putativos. Entre ellos sobresalen el diseño de los estudios (tamaño muestral, tipo de diseño, número de centros participantes, calidad y rigor científicos de estudios), la fuente de financiación, el papel desempeñado por revisores y editores y el prestigio de los autores. Esta aplicación derivada del conocido en epidemiología como *enfoque de riesgo* (Dever GEA, 1984; Rose G, 1993), sentaría las bases de estudios posteriores, en los que se enmarca la presente memoria de investigación, y del diseño de futuras medidas preventivas del sesgo.

Un aplicación analítica del enfoque de riesgo es la siguiente. Como se ha señalado, la publicación o no de un estudio se considera la variable de resultado (el efecto) o la variable dependiente de un modelo lineal generalizado (Kleinbaum DG, Kupper LL, Muller Keith E, 1988) y la significación estadística, el principal factor de riesgo (o la causa), como la variable independiente. Si existe sesgo de publicación, ambas variables están asociadas. Asimismo, existen otros factores de riesgo de la publicación o el rechazo de un estudio (variable que en adelante se denominará con fines analíticos *estado de publicación*) – además de la significación estadística–, tales como las características del diseño del estudio (inclusión de grupos de comparación, tamaño muestral, etc.), su fuente de financiamiento, etc., cuya posible asociación con la publicación también deben tenerse en cuenta. Estas otras variables intervienen en el modelo como independientes. Las supuestas asociaciones entre el destino final del artículo (publicación o rechazo) y el grado de significación estadística –asociación que definió originalmente el sesgo de publicación– podrían ser espurias y explicarse, en realidad, por la intervención de factores de confusión (Schlesselman J, 1982), que en este caso podrían corresponder a algunos de los factores de riesgo mencionados. Por ello, cualquier análisis de esa asociación debe eliminar (mantener constante o ajustar) el efecto de los posibles confusores.

En rigor, cabe esperar que sólo se publiquen los estudios de buena calidad y que los de inferior calidad se rechacen. Ello obliga a despejar dudas: si bien se ha observado reiteradamente que la posibilidad de que un estudio positivo se publique es más alta que la de uno negativo, ¿acaso la calidad de los positivos es más elevada que la de los negativos? La respuesta afirmativa a este interrogante descarta, por definición, la existencia del sesgo de publicación. Por contra, si la calidad de los estudios positivos es menor o igual que la de los negativos, en la práctica el criterio de enviar artículos a publicación o su aceptación editorial no reside en la calidad sino en la significación estadística o en otros factores. Este

caso define la existencia del sesgo. Por ello, al intentar explicar qué factores determinan el destino final de un estudio debe mantenerse constante (o eliminar) la influencia de las variables independientes mencionadas, algunas de las cuales podrían actuar como factores de confusión.

Una de las primeras investigaciones posteriores al congreso mencionado en que se aplicó el enfoque de riesgo fue un estudio de seguimiento llevado a cabo con 574 estudios aprobados por el Comité de Ética de la Universidad Johns Hopkins (Dickersin K, Meinert CL, 1990). En él se detectaron asociaciones directas entre la publicación o el rechazo de estudios y factores de riesgo putativos del sesgo, como notificar resultados estadísticamente significativos ($RP = 2,7$; IC95%: 1,67 - 4,36), haber sido financiado por los Institutos Nacionales de la Salud de los Estados Unidos (frente a los que lo estuvieron por otras instituciones), ser multicéntrico (frente a los monocéntricos), incluir un grupo de comparación (frente a no incluirlo), el sexo y rango de los investigadores, y tener un tamaño muestral mayor de 100. El diseño del estudio (observacional o ensayo clínico) no se asoció con la publicación.

En otro estudio similar en el cual se analizaron 293 ensayos clínicos se obtuvieron resultados parejos y congruentes con los del anterior: también se detectó una asociación entre el estado de publicación y la notificación de resultados estadísticamente significativos ($RP = 6,34$; IC95%: 1,62 - 24,97), así como con el rango del investigador. Un hallazgo sobresaliente fueron las razones esgrimidas por los autores para no publicar los ensayos: 45,4% los consideraron incompletos y 22,7% pensaban que no revelaban resultados "interesantes" (Dickersin K, Min YI, Meinert CL, 1991).

Mediante la construcción de varios modelos de regresión logística para investigar la existencia de asociaciones entre la publicación de un estudio y diversos factores "de riesgo" putativos del sesgo de publicación, otros autores también encontraron asociaciones entre la

publicación –incluso en revistas con factor impacto elevado– y la significación estadística ($RP_{ajustada} = 2,32$; IC95% 1,25 - 4,28) (Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR, 1991). La probabilidad de publicación también fue más alta en los estudios realizados por investigadores de rango elevado. La tendencia hacia el sesgo de publicación fue mayor en los estudios observacionales y de laboratorio ($RP = 3,79$; IC95%: 1,47 - 9,76) que en los ensayos clínicos aleatorizados ($RP = 0,84$; IC95%: 0,34 - 2,09). Por otro lado, se observó una tendencia más marcada a presentar resultados estadísticamente significativos en los estudios de baja calidad metodológica. La importancia de los estudios percibida por los investigadores también se asoció con la significación estadística de los resultados. La razón más frecuentemente aducida para no publicar los estudios fue la ausencia de significación estadística. La calidad de los ensayos no se asoció con la posibilidad de publicación. No obstante, el sesgo de publicación fue más prominente en los ensayos clínicos de menor tamaño muestral, en los que incluían un grupo control no concurrente y en aquellos cuya asignación no había sido aleatoria.

Parece entonces verosímil que los estudios más susceptibles a dicho sesgo son los más proclives a presentar otros sesgos y factores de confusión que amenazan marcadamente la validez de sus conclusiones. Así lo confirma el hallazgo de Easterbrook y colaboradores: la RP del sesgo de publicación de los estudios observacionales fue más elevada que la de los ensayos clínicos aleatorizados. Además, la RP de envío para publicación de un estudio positivo en relación con uno negativo fue más alta (2,94; IC95%: 1,43 - 6,01) que la de publicación (2,32). Por el contrario, ni el diseño, ni la inclusión de grupos de comparación, ni el tamaño muestral ni la fuente de financiamiento se asociaron con la publicación.

En otra investigación en la que se utilizó un modelo de análisis similar tampoco se detectó asociación alguna entre la publicación y la calidad del estudio (Chalmers I, Adams M, Dickersin K, et al., 1990), lo cual confirmó de nuevo la existencia del sesgo.

En otra aplicación de un diseño similar y de modelos lineales generalizados en el análisis para detectar asociaciones entre la publicación de estudios y sus posibles determinantes, no se encontró asociación entre publicación y tamaño muestral, la inclusión de grupos de comparación ni el tipo de estudio (observacional frente a ensayo clínico) (Dickersin K, Min YI, Meinert CL, 1992). No obstante, dicho estudio sugirió la existencia del sesgo, ya que la *RP* de la asociación entre estado de publicación y la obtención de resultados estadísticamente significativos fue 2,54 (IC95%: 1,63 - 3,94). Además, se confirmó el hallazgo de otros estudios: el sesgo es introducido fundamentalmente por los autores y no por los editores, por cuanto sólo 6 de los 124 estudios no publicados fueron rechazados por revistas. Los restantes no se enviaron a publicación por motivos parejos a los notificados en los estudios descritos con anterioridad. Ni la fuente de financiamiento, ni el número de centros participantes y de grupos en el estudio, ni el sexo ni el rango del investigador principal se asociaron con la publicación. Aunque en todos esos estudios se abogó por la adopción de las mismas medidas postuladas con anterioridad para minimizar o evitar la influencia del sesgo (Kahtan S, 1991; Charny M, 1991; Johnson JD, Breimer LH, 1991; Wilmshurst P, 1991; Rennie D, Flanagin A, 1992; Kaufman JL, 1992; Liberati A, 1992; Dickersin K, Meinert CL, Min YI, 1992; Rennie, Flanagin A, 1992), es menester subrayar el avance que supuso la publicación de un directorio de ensayos clínicos (Easterbrook PJ, 1992). Ello da cierta cuenta del grado de aceptación de dichas medidas a la luz de los beneficios que se comenzaron a grangear. Esta medida se vio acompañada por la propuesta de crear una sección sobre dichos registros en la revista *Controlled Clinical Trials* (Dickersin, 1992). Es preciso hacer hincapié en que los resultados de estos estudios señalan a los autores como principales responsables de la introducción del sesgo de publicación.

1. 13 Fuente de financiación como factor de riesgo y conflictos de intereses

La difusión editorial de la asociación encontrada entre la fuente de financiación de los estudios y el sesgo de publicación (especialmente los que involucraban a compañías farmacéuticas, cuyo parangón podría ser el estudio de Davidson) (Davidson RA, 1986), aunada a los debates que apelaban al papel y a la atribución de funciones de los comités de ensayos clínicos en relación con el sesgo, tuvo eco en el campo de la farmacología clínica (Levy G, 1992). Entre otras soluciones, se propuso que dichos comités deberían exigir la presentación de informes completos de los hallazgos de todas las investigaciones aprobadas por ellos no más de tres años después de finalizarlas. Los informes de progreso deberían presentarse, a más tardar, a los cinco años del inicio de los estudios que hubiesen estado realizándose durante más de tres años. Sólo por razones científicas, los comités podrían permitir el aplazamiento de la notificación de resultados. Asimismo, deberían enviar los informes a un registro central, que por su parte habría de publicar cuando menos sus resúmenes.

Esta propuesta suscitó respuestas heterogéneas, que traslucían distintos intereses y giraban en torno al derecho de los comités a arrogarse funciones normativas en política pública y a la perentoria necesidad de discernir entre sus atribuciones académicas y reguladoras¹⁴ (Mindel JS, 1992; Burris JF, 1993). Además, se reiteró la necesidad de aumentar la proporción de ensayos clínicos que se publican después de ser aprobados por entidades públicas reguladoras. Algunos atribuyeron esa baja proporción al nutrido número de estudios que las revistas reciben para publicación y al elevado porcentaje de rechazos editoriales. Otros, por el contrario, proclamaron que, desde la perspectiva de las compañías farmacéuticas, el diseño, la conducción de ensayos clínicos y la preparación de los documentos requeridos por las agencias reguladoras constituyen las principales prioridades,

¹⁴ Algunos académicos esgrimieron diversas razones intentando exonerarse de esas funciones.

mientras que su publicación se relega a un segundo plano, acaso por su interés –al menos en determinadas instancias– de no revelar al público ciertos datos (Novack GD, 1993).

En el contexto de los ensayos multicéntricos se ha puesto en tela de juicio si procede publicar resultados positivos obtenidos en un centro participante antes de disponer de los resultados de todos los centros, y si está justificada la publicación múltiple de los resultados de ensayos multicéntricos, es decir, la publicación independiente de los estudios procedentes de los centros participantes más la publicación de los resultados derivados del estudio multicéntrico (Leizorovicz A, Haugh MC, Boissel JP, 1992). En defensa de estos postulados algunos apelan al derecho de revelar y no postergar la publicación de cualquier resultado susceptible de rendir beneficios a los pacientes, sin esperar los hallazgos de los restantes centros (Stiller MJ, Mehrel T, 1992). Miembros de algunas sociedades científicas, como la American Academy of Dermatology y la Society of Investigative Dermatology, han propuesto, incluso, que dichas organizaciones creen comités encargados de revisar los estudios financiados por empresas farmacéuticas y de definir y fiscalizar la aplicación de requisitos para notificar sus resultados (Shear SL, 1993). También se ha aludido a la renuencia de ciertos editores a publicar estudios cuya autoría corresponde a empresas farmacéuticas, y se ha criticado con vehemencia la aplicación por los revisores de criterios más estrictos cuando revisan artículos de esta índole (Goldstein DJ, 1993; Mindel JS, 1993).

1. 14 Avances recientes en la investigación del sesgo de publicación

En 1993, la creciente búsqueda de factores de riesgo adicionales del sesgo de publicación rindió nuevos frutos a la metodología de investigación en este campo. En el *Second International Congress on Peer Review in Biomedical Publication* se presentaron dos estudios en los que se indagó si el prestigio de los investigadores y el sexo de autores y

revisores podrían sesgar la decisión editorial relativa a la publicación de artículos. El primer estudio no detectó pruebas que respaldaran la hipótesis según la cual revisores y editores de la *Journal of Pediatrics* emitían mejores juicios y recomendaciones y aceptaban con más frecuencia, respectivamente, los artículos cuyos autores pertenecían a instituciones prestigiosas que los pertenecientes a entidades de mucho menor renombre (Garfunkel JM, Hamrik HJ, Lawson EE, Ulshen MH, 1993). En el segundo estudio no se encontraron diferencias según el sexo de los autores, y de los editores a los que se asignaron, entre la proporción de artículos publicados y rechazados por la *JAMA* en 1991 (Gilberg J, Williams E, Lundberg GD, 1993).

Al margen de este congreso, en un tercer estudio se intentó dirimir una sospecha según la cual la revista alemana *Angewandte Chemie* estaba sesgada hacia la publicación preponderante de estudios de química orgánica en detrimento de los de otras especialidades de la química (Daniel HD, 1993). Por medio de un análisis comparativo de la distribución temática de los artículos aceptados y publicados, aquella denuncia se refutó. El perfil temático de la revista resultó remedar el de los artículos recibidos para publicación y, por consiguiente, no pudo atribuirse a decisiones sesgadas de los editores que escoraran el ámbito temático de la revista hacia alguna subespecialidad concreta de la química. Este estudio introdujo el *objeto de estudio* como nuevo factor de riesgo de los sesgos de publicación.

El mismo año se realizó otra investigación con objeto de responder a una crítica según la cual las revistas *Boletín de la Oficina Sanitaria Panamericana de la Salud* y *Bulletin of the Pan American Health Organization* –las dos revistas que publicaba la Oficina Regional de la OMS en América– aceptaban sesgadamente para publicación artículos de ciertos países miembros a expensas de rechazar inaceptablemente los de otros países (Campillo C, Alexander D, Ibrahim I, 1991). En dicha investigación, se acuñó el

término *sesgo de país* con el propósito de denotar la selección sesgada de artículos de determinados países en detrimento de los procedentes de otros países. El análisis, basado en un diseño de casos y controles y en un modelo logístico de la covarianza, constató que el país de procedencia no estaba asociado con la publicación o el rechazo de los artículos, al mantener constante el efecto de otras variables de confusión putativas.

Estos estudios, al incorporar nuevos factores de riesgo del sesgo de publicación, amplían el ámbito semántico del término y, por ende, la comprensión del intrincado entramado de factores causales y las posibilidades de hacerle frente desde más flancos. La significación estadística o la dirección o magnitud de los resultados de un estudio es solamente uno de los factores de riesgo de la introducción de sesgos de publicación. Acotar su significado a esa definición --y a los factores de naturaleza metodológica barajados en los estudios mencionados--, soslayando otros factores como el sexo y el prestigio de autores o revisores, o la procedencia de los artículos y la lengua materna de los autores, puede errar los análisis. Más aún, cuando por determinados motivos interesa investigar estos sesgos y elaborar medidas para paliarlos en lo que atañe exclusivamente a los editores, la omisión de esos factores puede desembocar indefectiblemente en análisis parciales y soluciones vanas. Por estos motivos, en adelante se hablará de *sesgos de publicación* en lugar de sesgo de publicación, a fin de reflejar que la significación estadística no es la única variable que puede sesgar la publicación de estudios.

La revisión de la literatura sobre este tema aparecida en los años más recientes permite confirmar que los sesgos de publicación han sido objeto de creciente consideración en medicina. Ello es en buena parte tributario del auge de los metanálisis y de la incorporación a un tiempo en ellos de evaluaciones y estimaciones de estos sesgos. Por ejemplo, distintos estudios epidemiológicos indicaron que el riesgo de padecer enfermedades de la vesícula biliar estaba asociado con el uso de anticonceptivos orales. Sin

embargo, investigaciones posteriores arrojaron resultados que contravenían este supuesto. Con objeto de dirimir esta disyuntiva, se realizó un metanálisis en el cual se intentó responder al siguiente interrogante: ¿existía en los años setenta una tendencia a publicar exclusivamente estudios positivos por considerarlos como los únicos "interesantes"? (Thijs C, Knipschild P, 1993). En definitiva, los autores no trataban sino de inquirir la existencia de sesgos de publicación que pudieran explicar los resultados contradictorios apreciados. Con este propósito emplearon el gráfico en embudo, mediante el cual representaron los *In RP* estimados en los estudios incluidos en el metanálisis frente a sus errores estándar, como medida de incertidumbre o inestabilidad de las estimaciones. El punto de intersección entre los valores de ambas variables se representó con un icono distinto según el diseño del estudio. Esta modificación de dicho gráfico sugiere la existencia del sesgo de publicación al tiempo que ilustra el papel del diseño como fuente del sesgo. El gráfico demostró la posible presencia de sesgo de publicación en esos estudios.

En varios debates suscitados por estudios realizados para evaluar la eficacia o la efectividad de diversas tecnologías médicas también se ha apelado a los sesgos de publicación como una de las posibles explicaciones de los resultados contrapuestos o no decisivos obtenidos. Tal es el caso de atribuciones infundadas de la eficacia de la vincristina para tratar la púrpura trombótica trombocitopénica en ausencia de resultados procedentes de ensayos clínicos controlados y aleatorizados (Makris M, Greaves M, 1991) o de las voces de alerta lanzadas sobre los peligros que puede acarrear la publicación de opiniones sesgadas a favor o en contra de las ventajas de la cirugía ginecológica laparoscópica (Kadar N, 1995; Smith DC, Donohue LR, Waszak SJ, 1994; Smith DR, 1995).

Los sesgos de publicación también se han sacado a colación en los recientes debates mantenidos entre académicos, investigadores y miembros de compañías tabacaleras sobre

los efectos nocivos del humo del tabaco en los fumadores pasivos. El estudio de Bero y colaboradores (Bero LA, Glantz SA, Rennie D, 1994) y la respuesta emitida por Levois y colaboradores (Levois ME, Layard MW, 1995) representan dos posiciones que ilustran sin parangón el peso específico que pueden tener estos sesgos a la hora de dilucidar relaciones causales en estudios epidemiológicos. Estos dos estudios constituyen, por añadidura, una prueba palmaria de la importancia que reviste el principio expuesto en los primeros pasajes de esta introducción: los estudios epidemiológicos pueden considerarse observaciones más o menos aleatorias destinadas a contrastar hipótesis. Por tanto, al obtener una estimación promediada, la verdadera fuerza de una asociación depende en parte de la disponibilidad de una muestra *representativa*, o sea no sesgada, de los estudios realizados. En este punto radica el debate mencionado sobre el tabaco.

En 1993, varias compañías tabacaleras lanzaron un alegato contra las afirmaciones de la Agencia de Protección del Medio Ambiente de los Estados Unidos, publicadas en 1992, sobre los riesgos asociados con el humo del tabaco, en las cuales se adujo que el sesgo de publicación las invalidaba (Levois ME, Layard MW, 1995). Para evaluar el fundamento de dicha protesta, se revisaron los estudios positivos y negativos publicados y no publicados que pretendían contrastar la hipótesis sobre la asociación referida (Bero LA, Glantz SA, Rennie D, 1994; Bero LA, Rennie D, 1993). Los autores no detectaron sesgos de publicación en los estudios publicados en revistas que emplean sistemas de revisión por expertos.

La industria tabacalera no tardó en responder y desplegó gran parte del arsenal metodológico disponible en un intento por demostrar la veracidad de su protesta (Levois ME, Layard MW, 1995). Para ello, revisaron estudios positivos y negativos publicados y no publicados sobre el particular. El gráfico en embudo obtenido fue asimétrico: todos los resultados publicados se situaban por encima del riesgo relativo (*RR*) global estimado al

combinar todos los estudios incluidos en esta investigación (1,29), lo cual sugiere la existencia del sesgo. La comparación del (*RR*) global obtenido al combinar los 14 estudios publicados (*RR* = 1,29; IC 95%: 1,18 - 1,41) con el de los 3 estudios no publicados (*RR* = 1,00; IC95%: 0,97 - 1,04) indica que los estudios publicados sobreestiman ligeramente la fuerza de la asociación ($\chi^2 = 25,1$; $P = 0,0001$).

Finalmente, utilizaron un modelo de regresión de Poisson¹⁵ para estimar la presencia del sesgo en esos estudios. Tanto el método desarrollado por Berlin (Berlin JA, Begg CB, Louis TA, 1989) como el gráfico en embudo y otros métodos mencionados se basan en el supuesto de que en todos los estudios, publicados y no publicados, la magnitud estimada de los efectos y los tamaños muestrales son independientes. Este supuesto no es más que otra forma de formular una hipótesis nula, que obviamente puede ser contrastada. En el modelo de Poisson, los *ln RR* estimados en los estudios se modelaron como una función de los correspondientes errores estándar de los *ln RR*. El coeficiente de regresión obtenido (que ofrece una prueba de la significación estadística del sesgo de publicación) fue significativo ($P < 0,01$), indicando la existencia de una asociación entre la magnitud de los efectos y los tamaños muestrales, es decir, de sesgo de publicación.

Cabe señalar que estos estudios ponen de manifiesto, además, el potencial que tienen los sesgos de publicación para influir no sólo en los debates sobre políticas de salud, sino también en su aplicación y en la vida cotidiana de los individuos¹⁶. Incluso hay quien

¹⁵ La regresión de Poisson es un modelo lineal generalizado basado en la distribución de Poisson, a diferencia de la regresión lineal múltiple, que se basa en la distribución normal. La distribución de la variable dependiente (*Y*) es Poisson. *Y* también puede ser una variable aleatoria binomial que pueda aproximarse con la distribución de Poisson. Esta variable se expresa como un recuento de acontecimientos muy poco frecuentes. Las variables independientes del modelo pueden ser de cualquier tipo, y sus coeficientes expresan razones de riesgo (vgr., riesgos relativos) ajustadas según las restantes variables independientes.

¹⁶ Valga mencionar a modo de ejemplo la reciente aplicación en el condado de Friendship Highs del Estado de Maryland, EEUU, de una ley que prohíbe fumar en la calle. El ciudadano que la infringe debe pagar una multa de \$US 100.

proclama que el asesoramiento normativo para elaborar políticas que dimana de algunos metanálisis --como los que se realizaron para dirimir la polémica sobre el efecto protector de la reducción de la ingesta de sal en la cardiopatía isquémica y los accidentes vasculares cerebrales-- puede ser extremadamente equívoco (Le Fanu J, 1995).

En esta misma línea también se han publicado diversos artículos y cartas al editor en los que se insiste en la necesidad de evaluar la existencia de sesgos de publicación --y se critica el no haberlo hecho, aunque cada vez se publican más metanálisis que incluyen evaluaciones de este tipo-- antes de aceptar la validez de los resultados procedentes de revisiones sistemáticas y metanálisis (Chalmers I, Collins RE, Dickersin K, 1992 K; Ketherndahl DA, 1992; Eaton CB, 1992; Earleywine M, 1993; Parmley WW, 1994; Morris RD, 1994). Las opiniones vertidas en ellas se basan en los resultados obtenidos al construir gráficos en embudo con los datos aportados por esos estudios y en la aplicación del método de Rosenthal para estimar la magnitud de dichos sesgos.

El amplio abanico temático que abarcan estas críticas (estudios sobre trastornos obsesivo-compulsivos, tratamiento de arritmias, alcoholismo, coronariopatías, tratamiento quirúrgico de fracturas del peroné, guías para realizar metanálisis en epidemiología del cáncer) dan buena cuenta de la prominencia del tema entre los profesionales de distintas disciplinas médicas.

En lo que respecta a la relación entre metanálisis y sesgos de publicación, los últimos años han sido testigos del aumento de artículos y libros publicados sobre metanálisis en los cuales se hace hincapié en la necesidad de incluir en todo metanálisis *análisis de la sensibilidad* destinados a evaluar el efecto potencial de los sesgos de publicación en los resultados; esto es, estimar el número de estudios negativos necesarios para trocar el resultado estadísticamente significativo obtenido en un metanálisis en otro no

significativo (valga recordar aquí con fines ilustrativos el método que Rosenthal propuso para estimar la magnitud del *problema del archivador*). Los estudios descritos anteriormente muestran la diversidad de métodos y modelos desarrollados con este propósito. Unos han sido más aceptados que otros y algunos se consideran heteróclitos y muy poco confiables por el carácter extremadamente especulativo de los modelos y supuestos en que asientan.

El estudio de Cook y colaboradores arroja luz sobre este asunto (Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, et al., 1993). A fin de identificar hasta qué punto los metanálisis efectuados entre 1989 y 1991 incluían estudios no publicados y recabar las opiniones de los investigadores sobre la conveniencia de incluirlos, se llevó a cabo una revisión de metanálisis publicados y una encuesta a autores de metanálisis, autores de artículos sobre metanálisis y editores de revistas que publican estos estudios. Su investigación reveló que de 150 metanálisis publicados, 46 (30,7%) incluyeron estudios no publicados. El 77,7% de los encuestados y 46,9% de los editores insistieron en la necesidad de incluirlos en los metanálisis y en las revisiones sistemáticas. Los autores concluyeron acertadamente que, si bien la inclusión de estudios no publicados en metanálisis sigue siendo objeto de controversia, buena parte de los que participan en ellos opinan que no debe incurrirse en el error de omitirlos sistemáticamente. En ausencia de datos confirmatorios, su conclusión debe aceptarse como razonable: la síntesis (cualitativa o cuantitativa) más válida de información parece ser aquella que somete a los estudios publicados y no publicados al mismo escrutinio metodológico riguroso y presenta los resultados de los estudios publicados y no publicados combinados y por separado.

Sobre este particular se ha descrito una reciente variante del sesgo de publicación, conocido como *sesgo de la torre de Babel*, que denota el efecto de la exclusión de estudios publicados en idiomas distintos del inglés en los resultados de los metanálisis (Gregoire G,

Derderian F, Le Iorier J, 1995; Sackett DL, Spitzer WO, 1994). En ausencia de más información sobre este particular, y teniendo en cuenta las reiteradas denuncias de discriminación lingüística de autores hispanoparlantes y las amenazas descritas en esta introducción a la validez de los estudios publicados, se justifican plenamente la postura avizora y la cautela con que deben revisarse los estudios individuales, los metanálisis y las revisiones publicadas. De ello depende parte del beneficio o perjuicio que pueden derivarse de las decisiones y la práctica médicas (Warren KS, Mosteller F, 1993).

1.15 Recapitulación

La calidad de un estudio no siempre es el criterio que rige su publicación. Las investigaciones revisadas revelan que la significación estadística alcanzada en un estudio, y no su calidad, determina con alta frecuencia si un estudio se publicará o no. Con el correr de los años, distintos estudios han puesto de manifiesto que otros factores, además de la significación estadística, pueden sesgar la selección de estudios para publicación.

Si se pretende que los estudios publicados sobre una especialidad determinada constituyan una muestra representativa de los conocimientos científicos rigurosos acumulados, es preciso garantizar que la calidad científica gobierne la investigación y la publicación de resultados. Ello exige conocer, entre otras cosas, cómo llevan a cabo las revistas médicas la revisión y la selección de las investigaciones que reciben para publicación o, en otras palabras, averiguar si la publicación de estudios depende de su calidad o, por el contrario, de otros factores como los analizados en esta introducción.

La utilidad de las investigaciones publicadas y, por tanto, la idoneidad de la práctica médica dependen en parte de la validez de los procesos editoriales de revisión y selección de estudios. Por consiguiente, para garantizar la validez de ambos procesos, las revistas médicas deben analizarlos y monitorizarlos de forma permanente.

CAPÍTULO 2. OBJETIVOS E HIPÓTESIS

2.1 Objetivos

A continuación, se definen los objetivos de esta memoria de investigación.

- Objetivo 1. Analizar los artículos originales recibidos para publicación por las revistas *Atención Primaria* y *Medicina Clínica* de Barcelona en función de las siguientes variables: estado de publicación, región española de procedencia del estudio, tamaño muestral, objeto del estudio, tipo de estudio o diseño empleado, nivel de complejidad del diseño y el análisis estadístico utilizados, grado de significación estadística de los resultados obtenidos, y calidad científica global del artículo.
- Objetivo 2. Estimar la calidad científica de los artículos recibidos por ambas revistas, empleando una escala construida y validada previamente con este propósito.
- Objetivo 3. Delimitar el perfil editorial de los artículos originales que publican y rechazan ambas revistas en relación con las variables enumeradas en el objetivo 1.
- Objetivo 4. Averiguar si alguna de las variables mencionadas en el objetivo 1 está asociada con la publicación y el rechazo de los artículos originales que reciben para publicación las dos revistas.

- Objetivo 5. Investigar la existencia de sesgos de publicación en las dos revistas y, en caso afirmativo, estimar si son introducidos por autores o editores.
- Objetivo 6. Evaluar la validez de los procesos de revisión editorial y selección de los artículos originales que las dos revistas reciben para publicación.

2.2 Hipótesis

En esta memoria de investigación se formulan, además, las siguientes hipótesis. Por consiguiente, se trata de un estudio de contrastación, no de generación de hipótesis.

- Hipótesis 1. Los perfiles editoriales de *Atención Primaria y Medicina Clínica* difieren en cuanto a la cobertura temática y a las distribuciones del tipo y el objeto de los estudios presentados en los artículos originales que reciben para publicación.
- Hipótesis 2. En las dos revistas, la calidad científica de los artículos publicados es más alta que la de los rechazados.
- Hipótesis 3. Sólo la calidad de los estudios se asocia con el estado de publicación de los artículos originales que reciben ambas revistas.
- Hipótesis 4. *Atención Primaria y Medicina Clínica* no introducen sesgo de publicación en los artículos originales que reciben. Si este sesgo existe, es responsabilidad exclusiva de los autores.

- Hipótesis 5. La publicación de artículos originales en *Atención Primaria* y *Medicina Clínica* no está influida por el sesgo de región de procedencia.

CAPÍTULO 3. MATERIALES Y MÉTODOS

3.1 Materiales

3.1.1 Búsqueda bibliográfica

La búsqueda de las referencias bibliográficas necesarias para realizar esta memoria de investigación se efectuó del siguiente modo. En primer lugar, se procedió a obtener referencias pertinentes, seleccionando palabras clave del *Medical Subject Headings* y realizando una búsqueda computadorizada en *MEDLINE* por medio del programa de búsqueda computadorizada (*off line*) *Grateful Med Version 6.6* (US National Library of Medicine, Department of Health and Human Services, 1995). A continuación, se realizó una segunda búsqueda con objeto de obtener las referencias de interés incluidas en la bibliografía de los artículos y capítulos de libros revisados. Asimismo, se efectuaron consultas personales con especialistas en el tema a fin de indagar y obtener referencias no encontradas en las dos búsquedas anteriores (Dickersin K, Goodman S, Meinert CL, Rhodie C, Szklo M, comunicaciones personales, 1992-1996). Cuando se dudó de la pertinencia de una referencia, ésta se obtuvo y revisó.

3.1.2 Criterios de inclusión y exclusión

Para realizar este estudio, se decidió evaluar una muestra de 200 artículos, 100 de *Atención Primaria* y 100 de *Medicina Clínica*. Como criterios de inclusión se adoptaron los siguientes. Primero, de los 100 artículos de cada revista, 50 debían estar aprobados para publicación y 50, rechazados. Segundo, todos debían ser artículos científicos originales y, por tanto, tener el formato IMRYD (Introducción, Materiales y métodos, Resultados Y Discusión). Tercero, en todos ellos se había de presentar uno de los siguientes grandes tipos

de estudios epidemiológicos: observacionales, cuasi-experimentales y experimentales. Cuarto, todos los artículos debían haber sido aprobados para publicación o rechazados durante un lapso que no se hubiese visto interrumpido por cambios en las políticas editoriales que pudiesen modificar de algún modo los procesos de revisión y selección de los artículos. Los artículos que no cumplieran alguno de estos criterios debían excluirse del estudio.

3. 1. 3 Selección de la muestra

El objetivo que se persiguió en el muestreo fue extraer una muestra representativa de los artículos aprobados para publicación y rechazados de cada revista durante un período exento de cambios en las políticas editoriales de revisión y selección de artículos. Este criterio garantiza, por un lado, que los resultados obtenidos en esta memoria de investigación corresponden o son atribuibles, según los casos, a las políticas editoriales mantenidas durante el período editorial que cubre este estudio. Por otro lado, permite que los resultados obtenidos sirvan para llevar a cabo evaluaciones posteriores de la efectividad de cambios en dichas políticas, ya que pueden servir como mediciones iniciales o previas a dichos cambios.

Luego de solicitar a cada revista 100 artículos que cumplieran los criterios de selección enumerados previamente, los responsables de cada una notificaron haber configurado las muestras del siguiente modo. En *Atención Primaria*, se incluyeron los 50 artículos aceptados entre mayo de 1993 y mayo de 1994, luego de ser sometidos a revisión editorial. Los 50 artículos rechazados corresponden a 50 artículos que se rechazaron entre julio de 1993 y marzo de 1994. En este período se rechazaron, tras revisión, 76 artículos, 15 de ellos porque no se adaptaban a la cobertura temática de la revista y los 61 restantes, por

recomendación de los revisores. Los 50 artículos incluidos en la muestra fueron los 50 últimos recibidos de esos 61.

En *Medicina Clínica*, se procedió del siguiente modo. Cada día que se reunió el comité editorial de la revista para seleccionar artículos, se hicieron copias del mismo número de artículos aprobados y rechazados para incluirlos en la muestra. Cuando el número de artículos rechazados y aprobados difirió, se incluyeron los menos numerosos --por ejemplo, los aprobados-- y la misma cifra de los restantes --por ejemplo, los rechazados-- por orden cronológico de rechazo.

Por consiguiente, las dos muestras son consecutivas, y los responsables de ambas revistas las consideran representativas de la población de referencia de este estudio, o sea, la de artículos recibidos, aprobados y rechazados durante un período exento de cambios en las políticas editoriales de revisión y selección de manuscritos, así como de las variables consideradas en esta memoria de investigación. En otras palabras, los artículos que integran la muestra son representativos de la población de la cual proceden en cuanto a las variables consideradas en este estudio. Es decir, en la muestra no se observan cambios seculares de estas variables en relación con períodos precedentes. Debe dejarse constancia de que, pocos meses después de seleccionar la muestra, en *Atención Primaria* se modificaron estas políticas. Por tanto, los resultados de esta revista sólo son extrapolables al período que antecede a la introducción de dichos cambios.

3.2 Métodos

3.2.1 Diseño

Para realizar este estudio, se utilizó el diseño de casos y controles (Breslow NE, Day NE, 1980; Schlesselman JJ, 1982). De acuerdo con este diseño, se consideraron casos todos los estudios publicados, y controles, todos los estudios rechazados. El desenlace o

variable de resultado o de punto final (*outcome* o *end point*, en inglés) fue el estado de publicación, una variable integrada por dos categorías: publicado o rechazado. La exposición se definió como el conjunto de variables que presuntamente podrían explicar o ser la determinante de la publicación o el rechazo de un artículo. Estas son: región de procedencia de los autores, tamaño muestral del estudio, objeto de estudio, tipo de diseño o de estudio, complejidad de los análisis, significación estadística de los resultados y calidad global del estudio. Desde el punto de vista de este diseño, se parte del principio de que cada una de esas variables o sus posibles acciones combinadas pueden determinar la publicación o el rechazo de un estudio, el desenlace.

Las ventajas que ofrece la utilización de este diseño son múltiples. Dados los objetivos de este estudio, el diseño de casos y controles permite medir en este caso tanto la exposición como el desenlace, así como detectar las posibles asociaciones --causales o no-- que puedan existir entre ellos. La naturaleza de los datos y el objeto del estudio (artículos científicos) ofrecen asimismo la posibilidad de medir las exposiciones y el desenlace minimizando los errores de medición, habida cuenta de que se conocen los valores de las variables que se desea analizar. Por añadidura, este diseño permite detectar la existencia de sesgos de procedencia y de publicación, tal como se han definido en el capítulo dedicado a los antecedentes científicos de esta memoria de investigación. Esto es, la detección de una asociación real entre procedencia y estado de publicación o entre esta última variable y la significación estadística correspondería, respectivamente, a la presencia de sesgo de procedencia y de publicación, siempre y cuando se descarten otros sesgos que pueda explicarla y los resultados de las pruebas de asociación se hayan ajustado teniendo en cuenta el efecto de factores de confusión. Más aún, la aplicación del diseño de casos y controles en este estudio en particular ofrece la ventaja de poder ajustar cualquier asociación observada y, con ello, eliminar el efecto de factores de confusión. Asimismo, y

también como consecuencia de la naturaleza de las variables de exposición y de desenlace analizadas, este diseño, si bien no impide que se introduzca algún tipo de sesgo de selección, reduce notablemente la posibilidad de que la validez interna del estudio se vea amenazada por otros sesgos equivalentes que le son característicos, tales como los de valoración (*ascertainment*), diagnóstico, referencia (*referral*), no respuesta, recuerdo o memoria, o clasificación errónea (*misclassification*) y sus variantes (Breslow NE, Day NE, 1980; Schlesselman JJ, 1982).

Finalmente, el diseño de casos y controles no obliga a esperar a la obtención de resultados sobre el desenlace, es decir, el lapso equivalente al período de latencia de una enfermedad, como ocurriría si se emplease un diseño de cohortes concurrente.

3.2.2 Variables estudiadas

Para realizar el estudio se definieron y analizaron las variables que se describen a continuación.

Estado de publicación. Esta variable se definió como el destino final de un artículo recibido para publicación por una revista, después de ser revisado, y se dividió y codificó en dos categorías: publicado = 1 ó rechazado = 0.

Procedencia del artículo. La procedencia hace referencia a la región donde trabajaban los autores del artículo. Como uno de los objetivos de esta memoria de investigación es estimar la existencia de sesgo de procedencia, esta variable se dividió y codificó en dos categorías: procedente de Cataluña = 1 ó de otra región del territorio español = 0.

Tamaño muestral. Éste se definió como el número de observaciones analizadas en el estudio, no el tamaño muestral estimado en la fase de diseño. Las categorías de esta variable fueron menor o igual a 50 = 0 y mayor de 50 = 1.

Objeto de estudio. El objeto de estudio hace referencia a la naturaleza de la unidad de análisis del estudio. Por ello, esta variable se estratificó y codificó en las siguientes cinco categorías: poblaciones, comunidades o personas = 1; servicios, programas o proyectos = 2; desarrollo, utilización o evaluación de tecnología = 3; análisis y síntesis de políticas de salud = 4, y otro = 5.

Tipo de estudio. Esta variable traduce el tipo de diseño empleado para llevar a cabo el estudio. Con objeto de categorizarla, en primer lugar se revisaron distintas clasificaciones publicadas de los tipos de estudios o diseños utilizados en epidemiología (Campbell DT, Stanley JC, 1963; Barker DJP, 1976; Fletcher RH, Fletcher SW, 1979; Lilienfeld AM, Lilienfeld DE; 1980; Riegelman RK, 1981; Mausner JS, Kramer S, 1985; Fleiss JL, 1986; Rothman KJ, 1986; San Martín H, Martín AC, Carrasco JL, 1986; Beeler MF, Sappenfield RW, 1986; Fleiss JL, 1986; Last JM, 1988). A continuación, se elaboró una clasificación, combinando las diferentes clasificaciones originales encontradas, a fin de dar cabida a algunos diseños no contemplados en ellas y facilitar la clasificación de los artículos evaluados. La clasificación modificada para esta memoria de investigación incluye, a diferencia de las restantes, la categoría estudios *comparativos* (estudios sin diseño claro y explícito que se limitan a comparar variables entre grupos¹⁷), *otros estudios experimentales* (que abarcan diseños experimentales distintos del ensayo clínico controlado, (Fleiss JL, 1996)), y *estudios cuasi-experimentales* (en los cuales se incluyen todos los diseños contemplados en la clasificación de Campbell DT y Stanley JC (Campbell DT y Stanley JC, 1963). La exhaustividad y especificidad de esta clasificación fue evaluada por tres personas en el estudio piloto de validación de un cuestionario para evaluar informes finales de investigación (Campillo C, de los Ríos R, Flores R, 1996). En dicho estudio se

¹⁷ Téngase en cuenta que en esta memoria la mitad de los estudios que se revisaron habían sido rechazados, muy probablemente, como consecuencia de su baja calidad metodológica. Por ello, había que dar cabida a estudios carentes de diseño claro y explícito que se limitaban a comparar arbitrariamente valores de variables entre grupos.

comprobó que ningún diseño deja de clasificarse y que, una vez definidos cada uno de los diseños, es fácil clasificar a la inmensa mayoría de los estudios aceptables en esta memoria de investigación en virtud de los criterios de inclusión adoptados.

La variable tipo de estudio original se dividió en tres grandes categorías: estudios sin intervención, estudios con intervención y otros. Los estudios sin intervención se codificaron como se detalla a continuación: 1 = descriptivos; 2 = comparativos; 3 = de prevalencia o transversales; 4 = casos y controles, y 5 = cohortes. Por su parte, los estudios con intervención fueron: ensayos clínicos = 6; otros estudios experimentales = 7, y estudios cuasi-experimentales = 8. Por último, los estudios no catalogables en las dos grandes categorías anteriores se clasificaron como *otros* y se codificaron como = 9.

Nivel de complejidad global del análisis, o de forma abreviada, *complejidad*. Esta variable traduce el nivel de complejidad de los análisis estadísticos realizados en el estudio. De acuerdo con esta definición, los estudios con análisis estadísticos exclusivamente univariantes se clasificaron como de complejidad introductoria = 1; aquellos cuyos análisis más complejos fueron bivariantes se clasificaron como de complejidad intermedia = 2, y los que utilizaron análisis multivariantes, como de complejidad avanzada = 3.

Significación estadística. Esta variable hace referencia al grado de significación estadística alcanzada en las pruebas estadísticas más importantes aplicadas en los estudios revisados, es decir, en aquellas con que se intenta responder a los principales objetivos o hipótesis del estudio. Según el nivel de significación acordado y notificado en cada estudio particular, esta variable se categorizó y codificó en dos niveles: estadísticamente significativo = 1, y estadísticamente no significativo = 0.

Calidad. La calidad es un concepto harto complejo, pero comprensible a nivel intuitivo. Asimismo, es multidimensional, por cuanto engloba numerosos atributos de muy

diversa naturaleza. Por ello, en esta memoria de investigación se empleó una definición que incluye dichos atributos o dimensiones y es simultáneamente comprensible y fácilmente aplicable. Así, la calidad se definió como la pertinencia, oportunidad, originalidad, rigor metodológico, idoneidad de la presentación de datos y contribución al conocimiento de un estudio. Esta definición fue consensuada por los tres investigadores que llevaron a cabo el estudio piloto de validación de un cuestionario para evaluar informes de investigación (Campillo C, de los Ríos R, Flores R, 1996), quienes comprobaron la validez de contenido, criterio y constructo de la escala que se construyó para medirla. En esta memoria se consideró válido, fiable, útil y sencillo utilizar una escala de calidad de 1 a 5, en la cual el valor 1 corresponde a la calidad más baja y 5, a la más alta. De dicha escala se excluyó el valor 0, por considerarse que cualquier estudio realizado y descrito es merecedor de cierto crédito, en contraposición tanto a la no realización de una investigación como a la ausencia de su notificación¹⁸. Por añadidura, la exclusión de dicho valor reduce la escala a 5 valores y, por tanto, la variabilidad de la variable calidad. Ello aumenta en principio la probabilidad de alcanzar niveles de concordancia interobservador más altos que los que podrían obtenerse si se empleara una escala con más de 5 posibles valores¹⁹.

3. 2. 3 Recategorización de variables

En la presente memoria de investigación se recategorizaron las siguientes.

Objeto de estudio. Para realizar algunos análisis, el objeto se recategorizó como una variable con tres categorías en lugar de cinco, eliminando de la variable inicial las

¹⁸ En el capítulo 1, Antecedentes científicos, se consignó que la ausencia de notificación de los resultados de los estudios se contrapone al avance del conocimiento.

¹⁹ Es preciso hacer hincapié en que hasta la fecha no se han encontrado en la literatura estudios que aporten definiciones operacionales, parsimoniosas y cuantificables de este concepto. En ello estriba parte de la originalidad de esta memoria de investigación.

categorías 4 (análisis o síntesis de políticas) y 5 (*otros*). La variable recodificada fue la siguiente: poblaciones, comunidades o personas = 1; servicios, programas o proyectos = 2, y desarrollo, utilización o evaluación de tecnología = 3.

Tipo de estudio. Esta variable sufrió dos recategorizaciones. En la primera, se convirtió en una variable dicotómica, combinando los tipos de estudio menos complejos – descriptivos y comparativos– en una sola categoría (recodificada como = 1) y los restantes, en otra (recodificada como = 2). En la segunda, la categoría *estudios de prevalencia* se combinó con la de los estudios con diseños menos complejos que la primera recategorización (y se recodificó como = 1; por tanto, esta categoría agrupa los estudios descriptivos, comparativos y de prevalencia). Los restantes tipos de estudio continuaron integrando la categoría 2.

Complejidad de los análisis. Al realizar algunos análisis bivariantes, las categorías intermedia y alta se combinaron en una sola con objeto de aumentar el tamaño muestral en las estratificaciones.

Calidad. Para realizar análisis bivariantes, la calidad –definida inicialmente como variable cuantitativa continua– se transformó en una variable con dos categorías: calidad menor de 2,5 = baja = 0, e igual o mayor de 2,5 = intermedia = 1. En el análisis multivariante, la calidad se consideró como variable cuantitativa continua.

3.2.4 Medición de la calidad de un estudio

Con objeto de evaluar los artículos revisados y medir las variables analizadas en este estudio, se elaboró un cuestionario. Para ello, se realizó una búsqueda bibliográfica (tal como se describió en el apartado 2. 1. 1), a fin de encontrar cuestionarios o guías empleados para tal fin. En la misma se encontraron 29 cuestionarios (Mahon WA, Daniel EE, 1964; Schor S, Karten I, 1966; Lionel NDW, Herxheimer A, 1970; O'Fallon JR, Dubey

SD, Salsburg DS, Edmonson JH, Soffer A, Colton T, 1974; Mosteller F, Gilbert JP, McPeck B, 1980; DerSimonian R, Charette LJ, McPeck B, Mosteller F, 1982; Zelen M, 1983; Altman DG, Gore SM, Gardner MJ, Pocock SJ, 1983; Emerson JD, McPeck B, Mosteller, 1984; Lock S, 1985; Meinert CL, Tonascia S, Higgins K, 1984; Kroch CL, 1985; Lock S, 1985; Meinert CL, 1986; Gardner MJ, Machin D, Campbell MJ, 1986; Pocock SJ, Hughes MD, Lee RJ, 1987; Lichtenstein MJ, Murlow CD, Elwood PC, 1987; Poynard T, 1988; Campillo C, Silva G, 1991; Wilson A, Henry DA, 1992; *Annals of Internal Medicine*, 1992; *American Journal of Epidemiology*, 1992; *American Journal of Epidemiology*, 1992; Audet N, Gagnon R, Ladouceur R, Marcil M, 1993; Programa de Publicaciones, Organización Panamericana de la Salud/Organización Mundial de la Salud, 1993; Campillo C, 1993).

Los cuestionarios o guías encontrados difieren entre sí en cuanto al tipo de estudios para cuya evaluación están destinados y al nivel de detalle con que permiten hacerlo. Algunas sólo evalúan atributos generales de los artículos. Otras, por el contrario, son específicas de artículos que notifican ensayos clínicos, evaluaciones de programas y servicios de salud o estudios observacionales (como la de estudios de casos y controles (Lichtenstein MJ, Murlow CD, Elwood PC, 1987)). Asimismo, algunas guías son exclusivamente cualitativas, mientras que otras contienen escalas para asignar una puntuación o valores de ponderación a los distintos atributos y características que miden.

Para construir un cuestionario exhaustivo, se procedió a combinar los apartados de las diferentes guías encontradas que hacían alusión a los mismos atributos o variables y a sintetizar la información que contenían. Seguidamente, dichos apartados se clasificaron según hiciesen referencia a una u otra de las secciones de un artículo original: resumen, introducción, materiales y métodos, resultados, discusión, referencias y presentación de datos (Day R, 1995). A continuación, se formularon las preguntas relativas a la información

básica que debe contener cada sección de todo artículo en el que se presente un estudio observacional, experimental, cuasiexperimental y de evaluación, teniendo en cuenta la información pormenorizada que contenían los apartados de las guías revisadas.

Por añadidura, en el cuestionario se introdujo un subapartado sobre los datos de identificación del artículo (número, revista, estado de publicación, y región de procedencia del artículo) y sendas subsecciones correspondientes a las restantes variables que se pretendían analizar: utilización de métodos estadísticos y epidemiológicos, tamaño muestral, objeto de estudio, tipo de estudio, nivel de complejidad global del diseño y el análisis, significación estadística, y un subapartado para comentarios.

La versión preliminar del cuestionario se evaluó en una prueba piloto en la cual tres evaluadores revisaron a ciegas los mismos cinco artículos. A la luz de los resultados de esta prueba, se introdujeron modificaciones en el cuestionario con objeto de eliminar incongruencias, aclarar dudas en la redacción de algunas preguntas y estandarizar la evaluación de los artículos. Seguidamente, se procedió a validar la nueva versión del cuestionario y adaptarla a los propósitos específicos de esta memoria de investigación, ya que el cuestionario inicial estaba destinado a evaluar informes de investigación, no artículos científicos, si bien en la práctica las diferencias entre ambos tipos de documentos son mínimas, sobre todo las relativas a la información concreta que se pretende evaluar. Las diferencias entre el cuestionario inicial y el utilizado en esta memoria no afectan a los resultados de la prueba de validación, ya que los apartados que los distinguen sólo hacen referencia a los datos de identificación de los estudios. A continuación, se realizó la prueba de validación de la versión final del cuestionario (Campillo C, de los Ríos R, Flores R, 1996), así como el estudio de confiabilidad de la escala para medir la calidad .

Los resultados de la validación del cuestionario fueron congruentes con los del cuestionario que en la actualidad se emplea en la revista *Ann Intern Med* para evaluar los

artículos que reciben para publicación (Goodman S, encargado de la revisión estadística y metodológica de la revista *Ann Intern Med*, comunicación personal, Chicago, septiembre de 1993; Campillo C, de los Ríos R, Flores R, 1996). La concordancia interobservador del cuestionario de esta memoria de investigación --estimada por medio del estadístico *kappa*-- osciló entre 45 y 60%.

Respecto a la validación del cuestionario es menester indicar que se compone de dos tipos de subapartados. En el primero, la influencia de la subjetividad del evaluador en la valoración de una variable es mínima (como ocurre con los relativos al estado de publicación, procedencia del estudio, utilización de métodos estadísticos y epidemiológicos, tamaño muestral, objeto de estudio y tipo de estudio, nivel de complejidad del diseño y el análisis, y significación estadística). En el segundo, relativo a la medición de la calidad, la subjetividad del evaluador desempeña un papel importante en la valoración, por cuanto éste ha de asignar una puntuación de calidad de 1 a 5 a cada pregunta del cuestionario.

Como se ha descrito en el subapartado relativo a la elaboración del cuestionario de evaluación de artículos, dicho cuestionario está integrado por ocho grandes secciones, una referente a los datos de identificación del artículo, otra para comentarios generales, y la tercera, compuesta a su vez por seis subsecciones correspondientes a los apartados que integran un artículo científico original. Estas seis subsecciones contienen diversas preguntas destinadas a comprobar si el artículo contiene la información específica a cada una de ellas y a medir la calidad de dicha información y, por ende, la del artículo en conjunto.

A cada una de dichas preguntas se asignó la escala de calidad de 1 a 5 descrita anteriormente. Por consiguiente, la puntuación de calidad de cada pregunta corresponde exclusivamente a la del atributo que ésta traduce. Además, para cada artículo evaluado se calculó la media aritmética simple de las puntuaciones de calidad de cada subapartado:

introducción, materiales y métodos, resultados, discusión y presentación de datos. Por ello, la media de calidad calculada refleja únicamente la calidad de cada subapartado del artículo. Finalmente, para cada artículo se calculó la media aritmética simple de todas las puntuaciones individuales de calidad mencionadas. Esta media global traduce directamente la calidad del artículo entendida en virtud de la definición adoptada de este concepto en la presente memoria de investigación.

Para validar la escala de calidad construida con los resultados de la presente investigación, se realizaron análisis de componentes principales y análisis factoriales. La confiabilidad final de esta escala se calculó mediante el coeficiente de *theta* de Armor (). Su valor fue 0,97, lo cual indica que la confiabilidad de la escala es muy alta.

Teniendo en cuenta estos datos, el hecho de que en la literatura revisada no se hayan encontrado cuestionarios validados, así como los principios metodológicos de la construcción de cuestionarios y escalas (Carmines EG, Zeller RA, 1981; McDowell I, Newell C, 1987; Feinstein AR, 1987; Streiner DL, Norman GR, 1989), el cuestionario y la escala de calidad se consideran válidas para los objetivos de esta memoria de investigación.

3. 2. 5 Evaluación de los artículos

Para evaluar los artículos y obtener la información de las variables de interés, se procedió del siguiente modo. En primer lugar, una persona distinta del evaluador convirtió con un escáner los 50 artículos publicados de *Atención Primaria* a formato de WP 5.1, ya que se recibieron los números de la revista, no los manuscritos mecanografiados. De este modo, la versión impresa que se evaluó posteriormente de estos artículos no se distinguió de la de los restantes manuscritos. A continuación, esta misma persona barajó 200 artículos y a cada uno de ellos le asignó un código numérico de identificación de 1 a 200, carente por completo de significado para el evaluador. Luego, anotó el código de cada artículo en hojas

aparte, junto con la información relativa a la revista de la que procedía, la región de procedencia de los autores, así como el estado de publicación. El evaluador no tuvo acceso a esa información hasta la etapa de análisis de resultados. Seguidamente, aquella persona eliminó estos datos de cada artículo --incluidos todos aquellos que pudiesen revelar la procedencia en el título o en el texto del manuscrito--, suprimiendo la primera página, cuando ésta figuraba, o tachando con tinta la información. Por consiguiente, el evaluador revisó cada artículo a ciegas, es decir, desconociendo la revista a la que pertenecía, su estado de publicación, la procedencia y otras señas de los autores.

El siguiente paso consistió en la evaluación y cumplimentación a ciegas del cuestionario o guía de evaluación de cada uno de los 200 artículos. Dicha evaluación incluyó dos etapas. En la primera, se leyó el artículo y se fueron anotando todos los datos en el cuestionario que no correspondían a la medición de la calidad (utilización de métodos estadísticos y epidemiológicos, tamaño muestral, objeto de estudio, tipo de estudio, complejidad y significación estadística). En la segunda --una vez leído el artículo--, a cada pregunta de cada subapartado del cuestionario se asignó una puntuación de calidad comprendida entre 1 y 5. En esta asignación sólo se asignaron números enteros.

Sólamente después de cumplimentar estos datos, es decir, sólo tras finalizar la evaluación de los 200 artículos, se procedió a anotar en el cuestionario la revista de la que procedía cada artículo, su estado de publicación y su procedencia. De este modo se minimizó cualquier posibilidad de introducir sesgos en la evaluación de los artículos y en la valoración de su calidad y, por ende, en las estimaciones obtenidas en los análisis. Además, la evaluación de los 200 artículos se realizó sin interrupciones que pudiesen modificar el modo en que se llevó a cabo. De este modo, se sistematizó la evaluación, se redujo la posibilidad de introducir tendencias en la forma de evaluarlos y, en el peor de los casos, se homogeneizó el sesgo de evaluación que pudiese introducirse en esta fase del estudio.

3. 2. 6. Tabulación de datos

El siguiente paso consistió en tabular en una matriz de datos los valores de las variables de cada artículo. Para tabular la calidad, en la matriz de datos se anotó, primero, el valor de la calidad de cada pregunta del cuestionario, la media de cada uno de los seis subapartados de cada artículo y, por último, la media global de la calidad de cada artículo. (Estos datos están a disposición de las personas que los soliciten.) Los valores de las restantes variables se tabularon directamente para cada artículo sin efectuar transformación numérica alguna.

Posteriormente, todos los valores de las variables codificadas tal como se describió anteriormente, se tabularon en un archivo con el procesador de palabras *WP 5.1* y con la hoja de cálculo *Excel*. Así, se creó un fichero con una columna para cada variable y una fila para cada artículo. Para corregir incongruencias y eliminar los errores de transcripción, antes de convertir este fichero para analizar los datos con paquetes estadísticos, todos los valores de las variables de cada artículo anotados en dicho fichero fueron cotejados por el evaluador y una segunda persona contra los anotados en el cuestionario. Los datos corregidos de este fichero fueron los utilizados directamente para realizar los análisis con paquetes estadísticos.

3. 2. 7 Análisis estadísticos

3. 2. 7. 1 Análisis univariantes

Para describir cuantitativamente las variables analizadas, se calcularon las distribuciones de frecuencias absolutas y relativas de cada una de ellas. Esto se llevó a cabo para los datos de ambas revistas en conjunto y para cada revista por separado. Cuando se trató de estimar medidas de tendencia central (como medias) o porcentajes, éstas medidas

se calcularon junto con sus intervalos de confianza del 95% a fin de presentar los errores asociados con su estimación.

3. 2. 7. 2 **Análisis bivariantes**

Con objeto de averiguar la existencia y estimar la fuerza de asociaciones entre pares de variables, se estimó la razón de posibilidades (*RP*) junto con su intervalo de confianza del 95%, siguiendo el método de Woolf o el de Cornfield (Kahn HA, Sempos CT, 1989). El uso de uno u otro método sólo responde a que cada programa estadístico utiliza preferentemente uno de ellos. Las *RP* se estimaron tanto para datos procedentes de las dos revistas, como en los análisis estratificados por revistas o según alguna variable en particular. Asimismo, cuando el número de observaciones fue bajo e impidió estimar con exactitud las *RP* a partir de tablas 2×2 , los límites de los intervalos de confianza se calcularon por medio del método de límites exactos de Mehta CR et al. (Mehta CR, Patel NR, Gray R, 1985). La estimación de *RP* ajustadas en función de terceras variables se llevó a cabo mediante la construcción de modelos de regresión logística (ver subapartado 3. 2. 7. 3).

Para comparar medias y porcentajes se realizaron pruebas de estimación, calculando y comparando los respectivos intervalos de confianza. En ninguna de estas comparaciones se realizaron pruebas de hipótesis.

3. 2. 7. 3 **Análisis multivariantes**

Para analizar qué variables están asociadas con el estado de publicación (la variable de resultado o dependiente) se llevaron a cabo dos tipos de análisis. En primer lugar, se averiguó qué variables se asociaron con ella a nivel bivariante, por medio de la estimación de la *RP* en tablas 2×2 o mediante la construcción de modelos de regresión logística,

compuestos por la constante del modelo y la variable independiente de interés. Cuando se sospechó la existencia de algún factor de confusión putativo, la estimación de la *RP* se estratificó en función de dicho factor o bien se construyó otro modelo logístico con la constante, la variable de interés y la variable potencialmente confusora como términos independientes. En caso de confirmarse el efecto confusor de dicho factor, las estimaciones de la *RP* se presentan ajustadas.

Para detectar asociaciones entre el estado de publicación y las restantes variables consideradas simultáneamente, se constuyó, primero, un modelo logístico de la covarianza (*logistic analysis of covariance*) (Kleinbaum DG, Kupper LL, Muller KE, 1988). En dicho modelo, la variable dependiente, estado de publicación se codificó en dos categorías: publicado o rechazado. Como términos independientes del modelo saturado se introdujeron todas las restantes variables, de las cuales la calidad se codificó como cuantitativa continua y las restantes, de acuerdo con la codificación descrita anteriormente. El modelo final se seleccionó siguiendo el método de eliminación de variables paso a paso (*stepwise regression*), según el cual las variables cuyo coeficiente de regresión no es estadísticamente significativo en la prueba de Wald se elimina del modelo (Kleinbaum DG, Kupper LL, Muller KE, 1988).

El principal supuesto del modelo de análisis logístico de la covarianza es el del paralelismo (*parallelism*), según el cual la relación entre cada variable independiente y la dependiente ha de ser la misma (es decir, sus rectas de regresión han de ser paralelas) para cada una de las categorías que integran cada variable independiente (Kleinbaum DG, Kupper LL, Muller KE, 1988). Para comprobar el cumplimiento de este supuesto básico (que también puede definirse como la ausencia de interacción), en cada modelo construido se incluyeron, además de las variables independientes, sus respectivos términos de interacción. La ausencia de interacción (o, en otras palabras, la falta de significación

estadística de los estadísticos F parciales múltiples de los coeficientes de regresión de los términos de interacción del modelo) confirma que en el modelo se respeta dicho supuesto.

Por último, y a fin de conocer la bondad de ajuste (*goodness of fit*) de los modelos finales, se realizó la prueba de la razón de verosimilitud (*likelihood ratio test*) (Kleinbaum DG, Kupper LL, Muller KE, 1988).

3. 2. 7. 4 Programas estadísticos

Los análisis estadísticos se realizaron con los programas EPI INFO (Centers for Disease Control, World Health Organization, 1990) y SAS (SAS Institute, 1992).

CAPÍTULO 4. RESULTADOS

4.1 Análisis uni y bivariantes

De la muestra diseñada inicialmente de 200 artículos, se revisaron 200. De éstos, se excluyeron 4, porque no correspondían al tipo de estudios o diseños seleccionados para realizar esta memoria de investigación. De los excluidos, 3 (1,5%) eran de *Atención Primaria* y 1 (0,5%), de *Medicina Clínica*. De este modo, la muestra final estuvo integrada por 196 artículos, 97 de *Atención Primaria* (48,5% de la muestra completa) y 99 de *Medicina Clínica* (49,5%).

4.1.1 Estado de publicación

Como se aprecia en la tabla 1, de los 196 artículos evaluados, 100 se publicaron --50 en cada revista-- y 96 fueron rechazados --47 por *Atención Primaria* y 49 por *Medicina Clínica*--. Tres de los 4 artículos excluidos de este estudio fueron rechazados por *Atención Primaria* y 1, por *Medicina Clínica*. Los cuatro abordaban los siguientes temas: 2 eran estudios de un caso, 1, un resumen del protocolo de un ensayo multicéntrico, y el último, un ensayo sobre medicina comunitaria.

TABLA 1. Estado de publicación de los 196 artículos evaluados según la revista.

Estado de publicación	Revista		Total
	<i>Atención Primaria</i>	<i>Medicina Clínica</i>	
Publicado	50	50	100
Rechazado	47	49	96
Total	97	99	196

4.1.2 Calidad de los artículos

Los valores mínimo y máximo de calidad asignados a los 196 artículos revisados fueron 1,0 y 4,8. La mediana de la calidad fue 2,0. Es decir, la calidad del 50% de los 196 artículos revisados fue igual o menor de 2,0 y la del 50% restante, mayor de 2,0.

Como muestra la tabla 2, las medias de la calidad total de los artículos publicados y rechazados por *Atención Primaria* y *Medicina Clínica* oscilan, con una

confianza del 95%, entre 1,48 y 2,92. En las dos revistas, las medias de la calidad total de los artículos publicados fueron más altas que las de los rechazados y distintas desde el punto de vista estadístico ([Atención Primaria: X publicados = 2,34; IC95%: 2,20 - 2,47, y X rechazados = 1,60; IC95%: 1,48 - 1,72]. [*Medicina Clínica*: X publicados = 2,70; IC95%: 2,55 - 2,92, y X rechazados = 1,80; IC95%: 1,68 - 1,89]).

TABLA 2. Medias de la calidad estimada, y sus correspondientes intervalos de confianza del 95% (IC95%), de las secciones de los artículos evaluados según la revista y su estado de publicación.

Sección	Atención Primaria		Medicina Clínica	
	Publicados	Rechazados	Publicados	Rechazados
Total	2,34 (2,20 - 2,47)	1,60 (1,48 - 1,72)	2,70 (2,55 - 2,92)	1,80 (1,68 - 1,89)
Introducción	2,58 (2,40 - 2,76)	1,90 (1,69 - 2,08)	3,20 (3,01 - 3,37)	2,10 (1,98 - 2,30)
Materiales y métodos	2,35 (2,18 - 2,53)	1,50 (1,39 - 1,70)	2,70 (2,49 - 2,94)	1,70 (1,56 - 1,86)
Resultados	2,44 (2,29 - 2,59)	1,60 (1,42 - 1,73)	2,40 (2,19 - 2,65)	1,80 (1,60 - 1,92)
Discusión	2,15 (2,00 - 2,30)	1,40 (1,27 - 1,47)	2,60 (2,35 - 2,79)	1,50 (1,39 - 1,62)
Referencias	2,08 (1,86 - 2,30)	1,90 (1,71 - 2,16)	3,30 (3,11 - 3,53)	2,50 (2,27 - 2,71)
Presentación	2,51 (2,35 - 2,67)	1,90 (1,73 - 2,09)	2,60 (2,41 - 2,81)	2,10 (1,91 - 2,25)

4. 1. 2. 1 **Calidad de las secciones**

Tanto en *Medicina Clínica* como en *Atención Primaria* la calidad estimada de todas las secciones de los artículos publicados también fue más alta que la de los rechazados (tabla 2). Las medias estimadas de las seis secciones de los artículos publicados por ambas revistas fueron distintas de las de los rechazados, con una confianza del 95%, excepto la de las Referencias de *Atención Primaria*, cuya diferencia puede ser atribuible al azar.

La clasificación de las cuatro secciones principales del artículo científico por orden decreciente de calidad fue la siguiente: Introducción, Resultados, Materiales y Métodos, y Discusión, exceptuando los artículos aprobados en *Medicina Clínica*. En estos últimos, la calidad más alta también correspondió a la Introducción, pero estuvo seguida por las secciones de Materiales y Métodos, Discusión y Resultados. La calidad de los apartados restantes --Referencias y Presentación-- fue asimismo mayor en los artículos publicados que en los rechazados de las dos revistas.

4. 1. 2. 2 **Calidad de los subapartados**

El siguiente nivel de análisis correspondió al de los subapartados que integran cada una de las seis secciones en que se subdividió la valoración de la calidad total. En la tabla 3 aparecen las medias de la calidad de los cinco subapartados de la Introducción, junto con sus correspondientes intervalos de confianza del 95%. Como ilustra la tabla, la calidad de todos los subapartados de la Introducción de los artículos publicados en ambas revistas fue más elevada que la de los rechazados. Las medias de la calidad de los primeros fueron estadísticamente más altas que las de los segundos, ya que ningún intervalo de confianza se solapa.

En todos los artículos, la descripción del marco general del estudio fue el subapartado de mayor calidad, la formulación de objetivos e hipótesis, el de menor, y los restantes, de calidad intermedia.

En cuanto a la sección de Materiales y Métodos (tabla 4), las medias de los seis subapartados de los artículos publicados también fueron más altas --y estadísticamente distintas con una confianza del 95%--, que las de los rechazados. En todos los artículos,

el subapartado de calidad más alta fue el relativo a la idoneidad del diseño empleado respecto a los objetivos e hipótesis del estudio. La calidad de los restantes subapartados varió según la revista y el estado de publicación del artículo, si bien, en promedio, la descripción de los métodos utilizados fue el subapartado de menor calidad.

Todas las medias de la calidad de los tres subapartados de la sección de Resultados (tabla 5) de los artículos publicados en las dos revistas fueron, al igual que las de las dos secciones precedentes, mayores que las de los rechazados. Además, fueron distintas desde el punto de vista estadístico con una confianza del 95%. De modo congruente, en todos los artículos el subapartado de calidad más alta (aunque baja e intermedia en la escala empleada) fue el relativo a la construcción de los modelos y la corrección de los análisis, seguido por el de la corrección del muestreo y, finalmente, por el de la adecuación de la presentación de los resultados.

En los nueve subapartados que componen la Discusión (tabla 6) se observó el mismo patrón. Todas las medias de los subapartados de los artículos publicados, aunque en general fueron bajas e intermedias en ambas revistas, excedieron a las de los rechazados y difirieron estadísticamente de ellas con una confianza del 95%. Aunque no se aprecia una jerarquía clara de los subapartados en virtud de la calidad, en las dos revistas los subapartados de mayor y menor calidad fueron, en promedio y respectivamente, la congruencia de la interpretación de los resultados con los objetivos e hipótesis del estudio y la notificación de sus limitaciones.

La calidad de las referencias (tabla 7) de los artículos publicados también fue más alta que la de los rechazados en las dos revistas, si bien no puede afirmarse que en *Atención Primaria* las medias fuesen estadísticamente distintas.

Por último, la calidad de la presentación de la información de los artículos publicados en las dos revistas fue mayor que la de los rechazados (tabla 8). Todas las medias de este apartado fueron estadísticamente distintas, con la excepción de las del subapartado relativo a la calidad del resumen de los artículos de *Medicina Clínica*. En esta revista no puede concluirse que la calidad de los resúmenes de los artículos aprobados difiriera de la de los rechazados, dado que los dos intervalos de confianza se solapan. Por otro lado, la calidad del subapartado sobre la concisión, sintaxis, estructura

de la información y adhesión a las normas internacionales de edición fue más elevada que la del resumen en todos los artículos, salvo la de los rechazados por *Medicina Clínica*, en los cuales la calidad del resumen excedió la del otro subapartado de la Presentación (tabla 8).

TABLA 3. Medias de la calidad estimada, y sus correspondientes intervalos de confianza del 95% (IC95%), de los subapartados de la sección de Introducción de los artículos evaluados según la revista y su estado de publicación.

Subpartado	Atención Primaria		Medicina Clínica	
	Publicados	Rechazados	Publicados	Rechazados
¿Se enmarca adecuadamente el estudio?	2,78 (2,58 - 2,98)	2,11 (1,89 - 2,33)	3,54 (3,36 - 3,72)	2,51 (2,34 - 2,68)
¿Se formulan con precisión el problema y la pregunta de investigación?	2,42 (2,22 - 2,62)	1,87 (1,65 - 2,09)	3,32 (3,11 - 3,53)	2,06 (1,86 - 2,26)
¿Se sustentan con argumentos sólidos y referencias idóneas?	2,68 (2,47 - 2,89)	1,85 (1,63 - 2,07)	3,14 (2,91 - 3,37)	2,16 (1,95 - 2,37)
¿Se indican claramente la utilidad y la aplicabilidad del estudio?	2,62 (2,40 - 2,84)	1,85 (1,63 - 2,07)	3,00 (2,74 - 3,26)	2,10 (1,91 - 2,30)
¿Se formulan correctamente los objetivos y las hipótesis?	2,38 (2,17 - 2,59)	1,76 (1,57 - 1,96)	2,96 (2,74 - 3,18)	1,88 (1,70 - 2,06)

TABLA 4. Medias de la calidad estimada, y sus correspondientes intervalos de confianza del 95% (IC95%), de los subapartados de la sección de Materiales y Métodos de los artículos evaluados según la revista y su estado de publicación.

Subapartado	Atención Primaria		Medicina Clínica	
	Publicados	Rechazados	Publicados	Rechazados
¿Es suficiente la descripción de los procedimientos de recolección, control de calidad, y el análisis de los datos?	2,46 (2,28 - 2,64)	1,83 (1,67 - 1,99)	2,78 (2,55 - 3,01)	1,90 (1,72 - 2,07)
¿Es apropiado el diseño utilizado para los objetivos y las hipótesis del estudio?	2,62 (2,42 - 2,82)	1,85 (1,61 - 2,08)	3,16 (2,91 - 3,41)	2,14 (1,93 - 2,35)
¿Se describe con suficiente detalle?	2,18 (1,98 - 2,38)	1,35 (1,19 - 1,51)	2,60 (2,33 - 2,87)	1,49 (1,33 - 1,65)
¿Se identifica claramente la unidad de análisis?	2,30 (2,07 - 2,53)	1,48 (1,28 - 1,67)	2,82 (2,55 - 3,09)	1,55 (1,38 - 1,72)
¿Son adecuadas la descripción y las referencias de los métodos de análisis utilizados?	2,32 (2,11 - 2,53)	1,33 (1,17 - 1,49)	2,70 (2,43 - 2,97)	1,51 (1,33 - 1,69)
¿Es adecuada la descripción del método empleado?	2,24 (2,03 - 2,45)	1,43 (1,23 - 1,64)	2,22 (1,88 - 2,56)	1,67 (1,46 - 1,88)

TABLA 5. Medias de la calidad estimada, y sus correspondientes intervalos de confianza del 95% (IC95%), de los subpartados de la sección de Resultados de los artículos evaluados según la revista y su estado de publicación.

Subpartado	Atención Primaria		Medicina Clínica	
	Publicados	Rechazados	Publicados	Rechazados
¿Se construyeron adecuadamente los modelos y fueron correctos los análisis realizados?	2,60 (2,43 - 2,77)	1,96 (1,73 - 1,18)	2,72 (2,47 - 2,97)	1,96 (1,78 - 2,14)
¿Fue correcto el muestreo?	2,46 (2,27 - 2,65)	1,41 (1,24 - 1,59)	2,14 (1,81 - 2,47)	1,78 (1,56 - 1,99)
¿Se presentan adecuadamente los resultados?	2,26 (2,08 - 2,44)	1,35 (1,20 - 1,50)	2,40 (2,16 - 2,64)	1,55 (1,38 - 1,72)

TABLA 6. Medias de la calidad estimada, y sus correspondientes intervalos de confianza del 95% (IC95%), de los subapartados de la sección de Discusión de los artículos evaluados según la revista y su estado de publicación.

Subapartado	Atención Primaria		Medicina Clínica	
	Publicados	Rechazados	Publicados	Rechazados
¿Las conclusiones obtenidas pueden obtenerse a partir de los métodos y resultados del estudio?	2,68 (2,53 - 2,83)	1,87 (1,69 - 2,05)	2,94 (2,69 - 3,19)	2,04 (1,84 - 2,24)
¿Es congruente la interpretación de los resultados con los objetivos e hipótesis del estudio?	2,96 (2,81 - 3,11)	2,17 (1,94 - 2,41)	3,24 (2,99 - 3,49)	2,43 (2,21 - 2,64)
¿Se comparan e interpretan adecuadamente los resultados a la luz de otros estudios pertinentes?	2,00 (1,84 - 2,16)	1,24 (1,11 - 1,37)	2,66 (2,40 - 2,92)	1,51 (1,35 - 1,67)
¿Se revisan críticamente los estudios relacionados?	1,32 (1,17 - 1,47)	1,04 (0,99 - 1,10)	2,06 (1,83 - 2,29)	1,02 (0,98 - 1,06)
¿Se señalan las fallas de asociación y se delimitan aspectos no resueltos?	1,84 (1,61 - 2,07)	1,07 (0,98 - 1,15)	2,42 (2,16 - 2,68)	1,33 (1,16 - 1,49)
¿Genera el estudio preguntas genuinas para futuras investigaciones?	2,08 (1,83 - 2,33)	1,07 (0,98 - 1,15)	2,40 (2,13 - 2,67)	1,33 (1,15 - 1,50)
¿Se exponen las consecuencias teóricas y las posibles aplicaciones prácticas del estudio?	1,90 (1,65 - 2,15)	1,07 (0,98 - 1,15)	2,50 (2,25 - 2,75)	1,27 (1,10 - 1,43)
¿Se indican claramente las limitaciones del estudio?	1,42 (1,21 - 1,63)	1,02 (0,98 - 1,06)	2,00 (1,73 - 2,27)	1,10 (0,98 - 1,22)
¿Aporta el estudio datos novedosos y necesarios?	3,16 (2,90 - 3,42)	1,78 (1,53 - 2,03)	2,90 (2,65 - 3,15)	1,53 (1,31 - 1,75)

TABLA 7. Medias de la calidad estimada, y sus correspondientes intervalos de confianza del 95% (IC95%), del subpartado de la sección de Referencias de los artículos evaluados según la revista y su estado de publicación.

Subpartado	Atención Primaria		Medicina Clínica	
	Publicados	Rechazados	Publicados	Rechazados
¿Sugieren una revisión pertinente, actualizada e íntegra de la literatura?	2,08 (1,86 - 2,30)	1,93 (1,71 - 2,16)	3,32 (3,11 - 3,53)	2,49 (2,27 - 2,71)

TABLA 8. Medias de la calidad estimada, y sus correspondientes intervalos de confianza del 95% (IC95%), de los subpartados de la Presentación de los artículos evaluados según la revista y su estado de publicación.

Subpartado	Atención Primaria		Medicina Clínica	
	Publicados	Rechazados	Publicados	Rechazados
¿Se lee con facilidad, es conciso y está bien estructurada la información según el formato IMRYD y las normas internacionales de edición?	2,56 (2,39 - 2,73)	1,96 (1,67 - 2,15)	2,80 (2,56 - 3,04)	2,04 (1,87 - 2,21)
¿Contiene el resumen la información necesaria?	2,46 (2,25 - 2,67)	1,87 (1,65 - 2,09)	2,42 (2,22 - 2,62)	2,12 (1,93 - 2,32)

4. 1. 2. 3 Calidad y estado de publicación

De modo coherente con estos resultados también se detectó una fuerte asociación entre la calidad y el estado de publicación. En el análisis combinado de las dos revistas dicha asociación fue fuerte ($RP = 21,94$; $IC95\%: 9,84 - 49,83$) (tabla 9). Estos datos indican que, con una confianza del 95%, la posibilidad de que un artículo de calidad intermedia ($\geq 2,50$) sea aceptado en alguna de las dos revistas es de 9,48 a 49,83 veces más alta que la de un artículo de calidad baja.

TABLA 9. Relación entre la calidad de los 196 artículos evaluados y su estado de publicación.

Calidad	Estado de publicación		Total
	Publicado	Rechazado	
Intermedia	86	21	107
Baja	14	75	89
Total	100	96	196

$RP = 21,94$; ($IC95\%: 9,84 - 49,83$).

Los hallazgos del análisis estratificado por revista confirman este hallazgo. Como se ilustra en las tablas 10 y 11, en *Atención Primaria* y en *Medicina Clínica* la posibilidad de que un artículo de calidad intermedia se publique es, respectivamente y con una confianza del 95%, de 5,68 a 55,07 y de 10,60 a 241,59 veces más elevada que la de un artículo de calidad baja.

TABLA 10. Relación entre la calidad de los artículos evaluados y su estado de publicación en *Atención Primaria*.

Calidad	Estado de publicación		Total
	Publicado	Rechazado	
Intermedia	39	8	47
Baja	11	39	50
Total	50	47	97

$RP = 17,28$; ($IC95\%: 5,68 - 55,07$).

TABLA 11. Relación entre la calidad de los artículos evaluados y su estado de publicación en *Medicina Clínica*.

Calidad	Estado de publicación		Total
	Publicado	Rechazado	
Intermedia	47	13	60
Baja	3	36	39
Total	50	49	99

$RP = 43,38$; (IC95% (límites exactos): 10,60 - 241,59).

4. 1. 3 Procedencia de los artículos

De los 196 artículos evaluados, se pudo conocer la procedencia de 173 (88,26%) y no pudo averiguarse la de los 23 restantes. De los 173, 129 (74,6%) procedían de Cataluña y 44 (25,4%) de otras regiones del territorio español. La tabla 12 muestra la procedencia de los artículos según la revista.

TABLA 12. Procedencia por revista de los 173 artículos en los que ésta se pudo averiguar.

Procedencia	Revista		Total
	<i>Atención Primaria</i>	<i>Medicina Clínica</i>	
Cataluña	78	51	129
Otro	16	28	44
Total	94	79	173

$RP = 2,68$ (IC 95%: 1,25 - 5,78).

Como puede observarse, en la muestra estudiada existe una asociación moderada entre la procedencia de los artículos y la revista a la que se enviaron. La posibilidad de que un artículo recibido por *Atención Primaria* procediera de Cataluña fue 2,68 veces más alta que la un artículo recibido por *Medicina Clínica*. No obstante, en la muestra la procedencia no se asoció con el estado de publicación (tabla 13).

TABLA 13. Relación entre el estado de publicación y la procedencia de los 173 artículos en los que ésta se pudo averiguar.

Estado de publicación	Procedencia		Total
	Cataluña	Otra	
Publicado	64	29	93
Rechazado	65	15	80
Total	129	44	173

$RP = 0,51$ (IC 95%: 0,23 - 1,10).

En otras palabras, al considerarlos en conjunto y por revista, los artículos publicados no se distinguen de los rechazados en cuanto a su procedencia. Asimismo, cuando el estado de publicación se estratificó por revista, tampoco se detectó ninguna asociación entre ambas variables, con una confianza del 95% (tabla 14).

TABLA 14. Relación entre el estado de publicación y la procedencia de 173 artículos en los que ésta se pudo averiguar según la revista.

Atención Primaria

Procedencia	Estado de publicación		Total
	Publicado	Rechazado	
Cataluña	39	39	78
Otro	9	7	16
Total	48	46	94

$RP = 0,78$ (IC95%: 0,23 - 2,58).

Medicina Clínica

Procedencia	Estado de publicación		Total
	Publicado	Rechazado	
Cataluña	25	26	51
Otro	20	8	28
Total	45	34	79

$RP = 0,38$ (IC95%: 0,13 - 1,14).

Es decir, ni en *Atención Primaria* ni en *Medicina Clínica* es más probable que se publique un artículo que proceda de Cataluña que uno que provenga de otra región.

4. 1. 3. 1 Procedencia y calidad

Por otra parte, la calidad se asoció con la región de procedencia del artículo al combinar los de ambas revistas (tabla 15). Como aprecia en esta tabla, los datos obtenidos indican con una confianza del 95%, que la probabilidad de que la calidad de un artículo de otra región sea intermedia (frente a baja) es tres veces más alta que la de un artículo proveniente de Cataluña. Es decir, es tres veces más probable que la calidad de los artículos de otra región sea más alta que la de los artículos procedentes de Cataluña.

TABLA 15. Relación entre el origen de los 173 artículos en los que se pudo averiguar la procedencia y su calidad.

Calidad	Procedencia		Total
	Cataluña	Otra	
Baja	65	11	76
Intermedia	64	33	97
Total	129	44	173

$RP = 0,33$ (IC 95%: 0,14 - 0,75).

Cuando estos resultados se estratificaron por revista, en *Atención Primaria* no se detectó asociación alguna, con una confianza del 95%, entre calidad y procedencia (tabla 16). Por el contrario, en *Medicina Clínica* los resultados son compatibles, con una confianza del 95%, con una asociación de moderada a fuerte entre calidad y procedencia. Así, en esta revista la probabilidad de que la calidad de un artículo proveniente de Cataluña sea baja (frente a intermedia) es unas siete veces más elevada que la de un artículo que proceda de otra región. En otras palabras, en *Medicina Clínica* es más probable que la calidad de los artículos de Cataluña sea más baja que la de los de otra región.

TABLA 16. Relación entre el origen de los 173 artículos en los que pudo averiguar la procedencia y su calidad según la revista.

Atención Primaria

Calidad	Procedencia		Total
	Cataluña	Otra	
Baja	41	8	49
Intermedia	37	8	45
Total	78	16	94

$RP = 1,11$ (IC 95%: 0,33 - 3,67).

Medicina Clínica

Calidad	Procedencia		Total
	Cataluña	Otro	
Baja	24	3	27
Intermedia	27	25	52
Total	51	28	79

$RP = 7,41$ (IC 95% (límites exactos): 1,85 - 42,12).

4. 1. 4 Tamaño muestral

De los 196 artículos revisados, en 182 (92,86%; IC95%: 89,26% - 96,46%) se presentaron estudios en los que se analizaron muestras. Los 14 (7,14%) restantes (8 de *Atención Primaria* y 6 de *Medicina Clínica*) se basaron en poblaciones. En 28 (15,4%) de esos 182 el tamaño muestral fue 50 o menor y en los 154 restantes (84,6%), mayor de 50.

En la muestra estudiada en esta memoria de investigación, el tamaño muestral no se asoció, con una confianza del 95%, con el estado de publicación al considerar conjuntamente las dos revistas (tabla 17) ni al estratificar esta asociación según la revista (tabla 18).

TABLA 17. Relación entre el tamaño muestral y el estado de publicación de los artículos evaluados en que se analizaron muestras.

Tamaño muestral	Estado de publicación		
	Publicado	Rechazado	Total
≤ 50	15	13	28
> 50	81	73	154
Total	96	86	182

$RP = 1,04$ (IC 95%: 0,43 - 2,51).

TABLA 18. Relación entre el tamaño muestral y el estado de publicación de los artículos evaluados en que se analizaron muestras según la revista.

Atención Primaria

Tamaño muestral	Estado de publicación		
	Publicado	Rechazado	Total
≤ 50	5	2	7
> 50	44	38	82
Total	49	40	89

$RP = 2,16$ (IC 95% (límites exactos): 0,33 - 23,71).

Medicina Clínica

Tamaño muestral	Estado de publicación		
	Publicado	Rechazado	Total
≤ 50	10	11	21
> 50	37	35	72
Total	47	46	93

$RP = 0,86$ (IC 95%: 0,29 - 2,52).

4. 1. 4. 1 Tamaño muestral y calidad

Los resultados indican, asimismo, que, con una confianza del 95%, el tamaño muestral no se asoció con la calidad de los estudios al analizar los datos de ambas revistas en conjunto (tabla 19) ni al estratificar dicha asociación según las revistas (tabla 20).

TABLA 19. Relación entre el tamaño muestral y la calidad de los artículos evaluados en que se analizaron muestras.

Calidad	Tamaño muestral		Total
	≤ 50	> 50	
Baja	14	67	81
Intermedia	14	87	101
Total	28	154	182

$RP = 1,30$ (IC 95%: 0,54 - 3,13).

TABLA 20. Relación entre la calidad y el tamaño muestral de los artículos evaluados en que se analizaron muestras según la revista.

Atención Primaria

Calidad	Tamaño muestral		Total
	≤ 50	> 50	
Baja	3	41	44
Intermedia	4	41	45
Total	7	82	89

$RP = 0,75$ (IC 95% (límites exactos): 0,10 - 4,75).

Medicina Clínica

Calidad	Tamaño muestral		Total
	≤ 50	> 50	
Baja	11	26	37
Intermedia	10	46	56
Total	21	72	93

$RP = 1,95$ (IC 95%: 0,66 - 5,81).

4. 1. 5 Objeto de estudio

Considerando en conjunto los artículos de *Atención Primaria* y *Medicina Clínica*, el objeto de estudio más frecuente de los artículos revisados fue poblaciones, comunidades o personas (109 artículos o 55,6%), seguido por servicios, programas o proyectos (47 o 24,0%), desarrollo, utilización o evaluación de tecnologías médicas (37 o 18,9%), y otros (3 o 1,%). En ningún artículo se presentó análisis o síntesis de políticas alguno. La categoría <<otro>> comprendió los siguientes objetos de estudio: 1 validación de un cuestionario, 1 estudio de producción científica y bibliométrica, y 1 estudio de validación de una pregunta de un cuestionario (tabla 21).

TABLA 21. Distribución del objeto de estudio de los 196 artículos evaluados según la revista y su estado de publicación.

Objeto	Revista					
	<i>Atención Primaria</i>			<i>Medicina Clínica</i>		
	Publicado	Rechazado	Total	Publicado	Rechazado	Total
Poblaciones, comunidades o personas	22	20	42	31	36	67
Servicios, programas o proyectos	26	18	44	2	1	3
Desarrollo, utilización o evaluación de tecnología	2	8	10	16	11	27
Análisis, síntesis de políticas	0	0	0	0	0	0
Otro	0	1	1	1	1	2
Total	50	47	97	50	49	99

El análisis estratificado de la distribución de frecuencias del objeto de estudio por revista (tabla 22) revela que en *Atención Primaria* el objeto de estudio más frecuente fue servicios, programas o proyectos, seguido por poblaciones, comunidades y personas, y por desarrollo, utilización y evaluación de tecnología. El menos frecuente fue <<otro>>. En *Medicina Clínica*, poblaciones, comunidades y personas ocupó el primer lugar, seguido por desarrollo, utilización y evaluación de tecnología, y por servicios, programas o proyectos. Para propósitos de estimación, en el tabla 22 se presentan los datos de la tabla 21 expresados como proporciones junto con sus correspondientes intervalos de confianza del 95%. La distribución de frecuencias del objeto de estudio de los artículos publicados por las dos revistas fue muy similar a la de los artículos rechazados (tabla 22). Desde un punto de vista inferencial, sólo se encontró una diferencia: en *Medicina Clínica* el porcentaje de artículos publicados cuyo objeto de estudio fue servicios, programas o proyectos fue mayor que el de rechazados, con una confianza del 95%.

TABLA 22. Porcentajes del objeto de estudio de los 196 artículos evaluados según la revista y su estado de publicación, junto con sus correspondientes intervalos de confianza del 95% (IC95%).

Objeto	Atención Primaria						Medicina Clínica					
	Publicado		Rechazado		Total		Publicado		Rechazado		Total	
	%	IC 95%	%	IC 95%	%	IC 95%	%	IC 95%	%	IC 95%	%	IC 95%
Poblaciones, comunidades o personas	22,7	14,7 - 30,7	20,6	12,6 - 28,6	43,3	32,3 - 54,3	31,3	22,9 - 40,3	36,40	27,4 - 45,4	67,7	58,7 - 76,7
Servicios, programas o proyectos	26,8	17,8 - 35,8	18,5	11,5 - 25,5	45,4	35,4 - 55,4	2,02	0 - 5,02	1,01	0 - 3,01	3,0	0 - 6,0
Desarrollo, utilización o evaluación de tecnología	2,1	0 - 5,1	8,2	3,2 - 13,2	10,3	4,3 - 16,3	16,2	9,2 - 23,2	11,1	5,1 - 17,1	27,3	19,3 - 35,3
Análisis, síntesis de políticas	0	0	0	0	0	0	0	0	0	0	0	0
Otro	0	0	1,03	0 - 3,03	1,00	0 - 3,0	1,01	0 - 3,0.	1,01	0 - 3,0	2,0	0 - 4,0

4. 1. 5. 1 Objeto del estudio y calidad

Para averiguar si el objeto de estudio se asoció con la calidad, el objeto se redefinió como una variable con tres categorías, eliminando de la variable inicial las categorías 4 (análisis o síntesis de políticas --pues no hubo ningún estudio con este objeto--) y 5 (<<otros>> --en la cual sólo se contabilizaron 3 de 196). En la muestra estudiada el objeto de estudio no se asoció con la calidad ni en el análisis combinado de ambas revistas (tabla 23) ni al estratificar el análisis según la revista (tabla 24).

TABLA 23. Relación entre los tres objetos de estudio más frecuentes y su calidad.

Calidad	Objeto de estudio			Total
	Poblaciones, comunidades o personas	Servicios, programas o proyectos	Desarrollo, utilización o evaluación de tecnología	
Baja	54	20	14	88
Intermedia	55	27	23	105
Total	109	47	37	193

RP (Poblaciones frente a servicios) = 1,33 (IC95%: 0,63 - 2,80).

RP (Poblaciones frente a desarrollo) = 1,61 (IC95%: 0,71 - 3,71).

TABLA 24. Relación entre los tres objetos de estudio más frecuentes y su calidad según la revista.

Atención Primaria

Calidad	Objeto de estudio			Total
	Poblaciones, comunidades o personas	Servicios, programas o proyectos	Desarrollo, utilización o evaluación de tecnología	
Baja	23	20	6	49
Intermedia	19	24	4	47
Total	42	44	10	96

RP (Poblaciones frente a servicios) = 1,45 (IC95%: 0,57 - 3,72).

RP (Poblaciones frente a desarrollo) = 0,81 (IC95% (límites exactos): 0,15 - 4,01).

Medicina Clínica

Calidad	Objeto de estudio			Total
	Poblaciones, comunidades o personas	Servicios, programas o proyectos	Desarrollo, utilización o evaluación de tecnología	
Baja	31	0	8	39
Intermedia	36	3	19	58
Total	67	3	27	97

RP (Poblaciones frente a desarrollo) = 2,05 (IC95%: 0,72 - 5,95).

4. 1. 6 Tipo de estudio

En la tabla 25 se presenta la distribución de frecuencias del tipo de estudio o diseño de los 196 artículos revisados. Los tres tipos de estudios más empleados (contabilizaron el 83,2%), por orden decreciente de frecuencia, fueron los comparativos, los de prevalencia y los descriptivos. La frecuencia de los estudios comparativos, aunque mayor, no difirió estadísticamente de la de los de prevalencia, con una confianza del 95%. La de los descriptivos fue menor que la de estos dos últimos. En cuarto, quinto y sexto lugar se clasificaron, respectivamente y en orden decreciente, los estudios cuasi-experimentales, otro tipo de estudio y los de cohortes. Dado el pequeño tamaño de las muestras de estas tres categorías, los intervalos de confianza del 95% son relativamente amplios y algunos de sus valores se solapan.

TABLA 25. Número, porcentaje e intervalos de confianza del 95% correspondientes (IC95%) de los tipos de estudio de los 196 artículos revisados.

Tipo de estudio	No.	%	IC 95%
Descriptivo	25	12,8	7,8 - 17,8
Comparativo	78	39,8	32,8 - 46,8
Prevalencia	60	30,6	24,6 - 36,6
Casos y controles	1	0,5	0,45 - 0,55
Cohortes	5	2,6	0,60 - 4,60
Ensayos clínicos	0	0	
Otros estudios experimentales	4	2,0	1,17 - 2,30
Cuasi-experimentales	12	6,1	3,10 - 9,10
Otro	11	5,6	2,6 - 8,6
Total	196	100	

Consiguientemente, las frecuencias de estos tres tipos de estudios no difieren, con una confianza del 95%. En los dos últimos lugares se situaron otros estudios experimentales y los de casos y controles. Sus frecuencias absolutas respectivas fueron muy bajas.

En ningún artículo se presentó un ensayo clínico. La muestra estudiada sólo incluyó 1 estudio de casos y controles y 5 de cohortes. Los 4 estudios experimentales distintos del ensayo clínico fueron investigaciones de laboratorio llevadas a cabo en condiciones experimentales. Los 11 estudios en los que se utilizaron diseños cuasi-experimentales fueron evaluaciones de servicios y programas de salud. El apartado integrado por otro tipo de estudios consistió, por revistas y estado de publicación, en las siguientes investigaciones: en *Atención Primaria*, 1 estudio de costes (publicado), y 2 estudios de validez de pruebas diagnósticas (1 publicado y 1 rechazado); y en *Medicina Clínica*, 3 estudios de validez de pruebas diagnósticas (1 publicado y 2 rechazados), 1 análisis de eficiencia de una tecnología médica (publicado), 1 estudio de seguimiento (publicado), 1 revisión de la literatura (publicado), y 2 validaciones de cuestionarios (1 publicado y 1 rechazado).

En la tabla 26 aparecen los tipos de estudios clasificados, de mayor a menor frecuencia, por revista. En *Atención Primaria* los tres tipos de estudios más frecuentes fueron los de prevalencia, los comparativos y los descriptivos. En *Medicina Clínica* estos tres tipos fueron el comparativo, el de prevalencia y el descriptivo.

TABLA 26. Orden de frecuencia de los tipos de estudio presentados en los 196 artículos evaluados en conjunto y según la revista.

Tipo de estudio	Total	Revista	
		<i>Atención Primaria</i>	<i>Medicina Clínica</i>
Descriptivo	3	3	3
Comparativo	1	2	1
Prevalencia	2	1	2
Casos y controles	8	6	
Cohortes	6	5	7
Ensayos clínicos			
Otros estudios experimentales	7		5
Cuasi-experimentales	4	4	6
Otro	5	5	4

No obstante estas diferencias, el porcentaje que representaron estos tres tipos de estudios en conjunto en cada revista no difirieron, como indican sus correspondientes IC95%: en la muestra de *Atención Primaria* integraron el 83,51% (IC95%: 76% - 90%) de todos los artículos, y en la de *Medicina Clínica*, el 82,82% (IC95%: 75% - 89%). Como ambos IC95% son prácticamente idénticos, la pequeñísima diferencia puede atribuirse a la variabilidad muestral.

Al comparar los porcentajes de los tipos de estudios de los artículos publicados y rechazados en cada revista sólo se encontraron diferencias entre los estudios descriptivos y los de prevalencia, y esa diferencia sólo se observó en *Atención Primaria* (tablas 27 y 28). Por tanto, desde un punto de vista inferencial, en esta revista, se rechazan más estudios descriptivos de los que se publican y se publican más estudios de prevalencia de los que se rechazan.

TABLA 27. Distribución de los tipos de estudio de los 196 artículos revisados según la revista y su estado de publicación.

Tipo de estudio	Revista					
	Atención Primaria			Medicina Clínica		
	Publicado	Rechazado	Total	Publicado	Rechazado	Total
Descriptivo	2	14	16	4	5	9
Comparativo	11	17	28	26	24	50
Prevalencia	29	8	37	12	11	23
Casos y controles	0	1	1	0	0	0
Cohortes	0	3	3	0	2	2
Ensayos clínicos	0	0	0	0	0	0
Otros estudios experimentales	0	0	0	2	2	4
Cuasi-experimentales	6	3	9	1	2	3
Otro	2	1	3	5	3	8
Total	50	47	97	50	49	99

Asimismo, también desde la perspectiva de la inferencia estadística, se deduce lo siguiente. Primero, a excepción de las dos diferencias mencionadas, en *Atención Primaria* no existen diferencias en cuanto a la frecuencia de los restantes tipos de estudios entre los artículos publicados y rechazados. Y, segundo, en *Medicina Clínica* las frecuencias de los tipos de estudios de los artículos publicados no difieren de las de los rechazados.

En síntesis, en ninguna de las dos revistas se ha encontrado asociación alguna entre el tipo de estudio y el estado de publicación, con la salvedad mencionada de *Atención Primaria*. La existencia en la tabla 27 de varios valores observados iguales a 0 y menores de 5 no permite efectuar pruebas de hipótesis habituales para contrastar la hipótesis nula de ausencia de asociación entre ambas variables. Además, las conclusiones que pudiesen extraerse de los resultados que se obtuvieran luego de aplicar métodos exactos serían extremadamente inestables desde un punto de vista inferencial por el pequeño tamaño de los estratos, y no aumentarían la validez externa del estudio en modo alguno.

TABLA 28. Distribución porcentual, con los correspondientes intervalos de confianza del 95% (IC95%), de los tipos de estudio presentados en los 196 artículos evaluados según la revista y su estado de publicación.

Tipo de estudio	Revista							
	Atención Primaria			Medicina Clínica				
	Publicado		Rechazado	Publicado		Rechazado		
%	IC 95%	%	IC 95%	%	IC 95%			
Descriptivo	2,1	0 - 4,8	14,4	7,4 - 21,4	4,0	0,1 - 7,9	5,0	0,7 - 9,3
Comparativo	11,3	5,5 - 12,6	17,5	9,9 - 25,1	26,2	17,5 - 34,9	24,2	15,8 - 32,6
Prevalencia	29,9	20,8 - 36,9	8,2	2,7 - 13,7	12,1	5,7 - 18,5	11,1	4,9 - 17,3
Casos y controles	0		1,0	0 - 3,0	0		0	
Cohortes	0		3,0	0,05 - 7,3	0		2,0	0 - 4,8
Ensayos clínicos	0		0		0		0	
Otros estudios experimentales	0		0		2,0	0 - 4,8	2,0	0 - 4,8
Cuasi-experimentales	6,1	1,3 - 10,9	3,1	0,05 - 7,0	1,0	0 - 3,1	2,0	0 - 4,8
Otros	2,0	0 - 4,8	1,0	0 - 3,0	5,0	0,7 - 9,3	3,0	0 - 6,4
Total	51,5	41,6 - 61,4	48,4	38,5 - 58,3	50,5	40,7 - 60,3	49,4	39,6 - 59,2

4. 1. 6. 1 Tipo de estudio y calidad

En la tabla 29 se presentan los distintos tipos de estudios de todos los artículos revisados según su calidad. Los datos sugieren que en la muestra de artículos evaluados existe una asociación entre el tipo y la calidad del estudio ($\chi^2(7) = 22,373; P = 0,002$).

TABLA 29. Relación entre el tipo de estudio de los 196 artículos evaluados y su calidad.

Tipo de estudio	Calidad		Total
	Baja	Intermedia	
Descriptivo	16	9	25
Comparativo	41	37	78
Prevalencia	16	44	60
Casos y controles	1	0	1
Cohortes	5	0	5
Ensayos clínicos	0	0	0
Otros experimentales	2	2	4
Cuasi-experimentales	5	7	12
Otros	3	8	11
Total	89	107	196

$$\chi^2(7) = 22,373; P = 0,002.$$

Si se revisa detenidamente la tabla se aprecia que, globalmente, la cifra de artículos de calidad intermedia es más alta que la de los de calidad baja para casi todos los tipos de estudios, lo cual tiende a convertirse en tendencia a medida que aumenta la complejidad del tipo de estudio. Excepciones a esta tendencia general fueron los estudios de casos y controles (sólo hubo uno y fue de calidad baja) y los de cohortes (los cinco de la muestra fueron de calidad baja). Lo contrario se observa con los estudios menos complejos (descriptivos y comparativos): en ellos, el número de estos estudios que fueron de baja calidad fue mayor que el de los de calidad intermedia. Además, el mismo número de otros estudios experimentales fueron de calidad baja y alta.

A la luz de estos resultados y con el propósito de profundizar en el análisis de esta asociación, la variable tipo de estudio se transformó en una variable dicotómica, combinando los tipos menos complejos --descriptivos y comparativos-- en una categoría y los restantes en otra. La tabla 30 muestra la asociación detectada entre las variables tipo de estudio recategorizada y calidad de los artículos de las dos revistas analizadas

conjuntamente. Como puede observarse, la posibilidad de que un estudio descriptivo y comparativo sea de calidad baja es, con una confianza del 95%, 2,36 veces más alta que la de los seis tipos restantes (prevalencia, casos y controles, cohortes, cuasi-experimentales, otros estudios experimentales y <<otros>>). Dicha asociación se explica por el hecho de que más estudios descriptivos y comparativos fueron de calidad baja que intermedia y porque de los tipos de estudios restantes los de calidad intermedia fueron el doble de los de calidad baja.

TABLA 30. Relación entre el tipo de estudio (reategorizado en dos grupos) de los 196 artículos evaluados y su calidad.

Calidad	Tipo de estudio		Total
	0	1	
Baja	57	32	89
Intermedia	46	61	107
Total	103	93	196

$RP = 2,36$; (IC95%: 1,27 - 4,39).

0 comprende los estudios descriptivos y comparativos y 1, los restantes

Al estratificar esta asociación por revista se obtuvieron resultados coherentes con los que arrojó el análisis combinado. Los resultados obtenidos tanto en *Atención Primaria* como en *Medicina Clínica* son compatibles, con una confianza del 95%, con la existencia de una asociación débil entre el tipo de estudio y la calidad (tabla 31). Claramente, los posibles valores del parámetro RP son más elevados en *Atención Primaria* que en *Medicina Clínica*. En *Atención Primaria*, la asociación podría ser incluso moderada a nivel poblacional. Los valores altos estimados de la fuerza de la asociación (todos ellos mayores de 1,0) son atribuibles a que hubo más tipos de estudio con diseños menos complejos de calidad baja y más estudios con diseño más complejo cuya calidad fue intermedia. Por el contrario, en *Medicina Clínica* algunos valores estimados del la RP fueron menores de 1,0, lo cual indica que los datos obtenidos son también compatibles con que en la población de la cual proceden los artículos de esta submuestra la posibilidad de que los artículos con diseños menos complejos sean de calidad más alta es menor que la de los artículos con diseño más complejo. Los

resultados que atañen a esta revista se explican porque, a diferencia de *Atención Primaria*, hubo más artículos de calidad intermedia en las dos categorías de la variable tipo de estudio.

TABLA 31. Relación entre el tipo de estudio (recategorizado en dos grupos) de los 196 artículos evaluados y su calidad, según la revista.

Atención Primaria

Calidad	Tipo de estudio		Total
	0	1	
Baja	31	19	50
Intermedia	13	34	47
Total	44	53	97

$RP = 4,27$; (IC95%: 1,67 - 11,07).

Medicina Clínica

Calidad	Tipo de estudio		Total
	0	1	
Baja	26	13	39
Intermedia	33	27	60
Total	59	40	99

$RP = 1,64$; (IC95%: 0,65 - 4,13).

0 comprende los estudios descriptivos y comparativos y 1, los restantes

Dado que la diferencia de calidad más notable se observó en los estudios de prevalencia (tabla 29), la variable tipo de estudio se recategorizó de nuevo, traspasando los estudios de prevalencia a la categoría estudios descriptivos y comparativos. Al analizar los datos obtenidos se constata (tabla 32) que la asociación entre tipo de estudio prácticamente desaparece (la RP está muy próxima a 1,0) y los valores estimados de la RP son congruentes tanto con la presencia en ambos sentidos como con la ausencia de dicha asociación.

TABLA 32. Relación entre el tipo de estudio (recategorizado en dos grupos) de los 196 artículos evaluados y su calidad.

Calidad	Tipo de estudio		Total
	0	1	
Baja	82	17	99
Intermedia	81	16	97
Total	163	33	196

$RP = 0,95$; (IC95%: 0,42 - 2,14).

0 comprende los estudios descriptivos, comparativos y de prevalencia y 1, los restantes

4.1.7 Complejidad del análisis

De los 196 artículos revisados, 191 presentaban análisis estadísticos de algún tipo. En 128 de ellos (67,0%; IC95%: 60% - 74%) la complejidad fue intermedia, es decir, los análisis consistieron en pruebas de hipótesis (la mayor parte) y de estimación a nivel bivariante. Les siguieron en orden de frecuencia los artículos con análisis de complejidad baja (40 o 20,9%; IC95%: 14,9% - 26,9%), es decir, los que contienen solamente análisis estadísticos univariantes. Por último, en 23 artículos (12,9%; IC95%: 7,9% - 17,9%) se llevaron a cabo análisis de complejidad alta, o sea, aquellos en los que se presentó algún modelo estadístico multivariante. Por consiguiente, al considerar en conjunto ambas revistas, se constata que buena parte de los artículos recibidos presentan análisis de complejidad intermedia.

A nivel inferencial, los resultados permiten afirmar, con una confianza del 95%, que el porcentaje de estudios en los que se han realizado análisis de complejidad baja no difiere del de aquellos que presentan análisis de complejidad intermedia, pues sus intervalos de confianza se solapan.

Al comparar la complejidad de los análisis de los artículos publicados y rechazados por ambas revistas conjuntamente (tabla 33) se observa que se rechazaron más artículos con análisis de complejidad baja y alta de los que se publicaron, y que se publicaron más de complejidad intermedia de los que se rechazaron. La única asociación detectada entre la complejidad de los análisis y el estado de publicación corresponde a los estudios con análisis de complejidad baja e intermedia: la posibilidad

de que se publique un artículo de complejidad intermedia es 1,6 veces más alta que la que se publique uno de complejidad baja.

TABLA 33. Relación entre la complejidad de los análisis estadísticos presentados en los artículos evaluados y su estado de publicación.

Complejidad	Estado de publicación		Total
	Publicado	Rechazado	
Baja	13	27	40
Intermedia	73	55	128
Alta	10	13	23
Total	96	95	191

RP (baja frente a intermedia) = 0,36 (IC95%: 0,16 - 0,81).

RP (intermedia frente a alta) = 1,73 (IC95%: 0,65 - 4,63).

RP (baja frente a alta) = 0,63 (IC95%: 0,19 - 2,05).

No obstante, al estratificar este análisis según la revista, no se detectó asociación alguna entre ambas variables (tabla 34).

TABLA 34. Relación según la revista entre la complejidad de los análisis estadísticos presentados en los artículos evaluados y su estado de publicación.

Atención Primaria

Complejidad	Estado de publicación		Total
	Publicado	Rechazado	
Baja	12	20	32
Intermedia	36	25	61
Alta	2	2	4
Total	50	47	97

RP (baja frente a intermedia) = 0,42 (IC 95%: 0,16 - 1,09).

RP (baja frente a alta) = 0,60 (IC 95% (límites exactos): 0,04 - 9,43).

RP (intermedia frente a alta) = 1,44 (IC 95% (límites exactos): 0,10 - 1,44).

Medicina Clínica

Complejidad	Estado de publicación		Total
	Publicado	Rechazado	
Baja	1	7	8
Intermedia	37	30	67
Alta	8	11	19
Total	46	48	94

RP (baja frente a intermedia) = 0,12 (IC 95%(límites exactos): 0,00 - 1,01).

RP (baja frente a alta) = 0,20 (IC 95% (límites exactos): 0,00 - 2,18).

RP (intermedia frente a alta) = 1,70 (IC 95%: 0,54 - 5,37).

4. 1. 7 Complejidad y calidad

Al analizar la relación entre estas dos variables, fue preciso combinar las categorías intermedia y alta de la complejidad de los análisis, dado que algunos efectivos observados fueron iguales a cero. Tanto en el análisis combinado (tabla 35) como en el estratificado por revista (tabla 36), los datos son congruentes con la ausencia de asociación entre complejidad y calidad de los artículos. en las dos revistas

TABLA 35. Relación entre la calidad de los artículos evaluados y la complejidad de los análisis estadísticos presentados en ellos, luego de combinar las categorías intermedia y alta de la variable complejidad.

Calidad	Complejidad		Total
	Baja	Intermedia y alta	
Baja	22	66	88
Intermedia	18	85	103
Total	40	151	191

$RP = 1,57$ (IC 95%: 0,74 - 3,36).

TABLA 36. Relación según la revista entre la calidad de los artículos evaluados y la complejidad de los análisis estadísticos presentados en ellos, luego de combinar las categorías intermedia y alta de la variable complejidad.

Atención Primaria

Calidad	Complejidad		Total
	Baja	Intermedia y alta	
Baja	17	33	50
Intermedia	15	32	47
Total	32	65	97

$RP = 1,10$ (IC 95%: 0,43 - 2,80).

Medicina Clínica

Calidad	Complejidad		Total
	Baja	Intermedia y alta	
Baja	5	33	38
Intermedia	3	53	56
Total	8	86	94

$RP = 2,68$ (IC 95% (límites exactos): 0,48 - 18,18).

4. 1. 8 Significación estadística

De los 196 artículos revisados, en 146 (62 de *Atención Primaria* y 84 de *Medicina Clínica*) se aplicaron pruebas de hipótesis. En los 50 restantes (35 de *Atención Primaria* y 15 de *Medicina Clínica*) no se aplicaron pruebas de hipótesis ni de inferencia estadística. En estos últimos sólo se realizaron análisis descriptivos.

De los 146 artículos en los que realizaron pruebas de hipótesis o de significación estadística, 143 (97,9%) fueron positivos y 3 (2,1%), negativos. Por consiguiente, en la

práctica totalidad de los artículos que se enviaron a publicación a ambas revistas y en los que se habían hecho pruebas estadísticas, éstas eran de significación, y sus resultados fueron estadísticamente significativos. En otras palabras, la inmensa mayoría de los estudios en los que se habían realizado pruebas de hipótesis y se enviaron a publicación fueron estudios positivos.

Al analizar estos resultados según el estado de publicación, se observa que de los 146 artículos con pruebas de hipótesis, 79 se publicaron y 67 se rechazaron (tabla 37). De los 3 estudios negativos, 2 fueron publicados por *Medicina Clínica* y 1, rechazado por *Atención Primaria* (tabla 38) Por tanto, obviamente se constata que en la muestra estudiada no existe asociación entre la significación estadística de un estudio y su publicación o rechazo.

TABLA 37. Relación entre el nivel de significación estadística notificado en los artículos evaluados y su estado de publicación.

Significación estadística	Estado de publicación		Total
	Publicado	Rechazado	
Significativo	77	66	143
No significativo	2	1	3
Total	79	67	146

Si estos últimos resultados se estratifican por revista, se obtienen los que figuran en la tabla 38.

TABLA 38. Relación entre el nivel de significación estadística notificado en los artículos evaluados y su estado de publicación según la revista.

Atención Primaria

Significación estadística	Estado de publicación		Total
	Publicado	Rechazado	
Significativo	34	27	61
No significativo	0	1	1
Total	35	28	62

Medicina Clínica

Significación estadística	Estado de publicación		Total
	Publicados	Rechazados	
Significativo	43	39	82
No significativo	2	0	2
Total	44	40	84

Como se puede apreciar, ambas tablas 2×2 contienen valores observados menores de 5. Ello, aunado a las características que confieren a las tablas los restantes datos, excluye la necesidad de realizar pruebas estadísticas de asociación para poder afirmar, sin lugar a dudas, que en las muestras estudiadas de *Atención Primaria* y *Medicina Clínica* no existe asociación entre el estado de publicación y la presencia o ausencia de significación estadística de un estudio.

Lógicamente, carece de sentido analizar si existe asociación entre la presencia de significación estadística y la calidad de los artículos, dado que sólo en 3 de ellos los resultados no fueron estadísticamente significativos.

4.12 Análisis multivariantes

Una vez realizado el análisis bivalente entre las distintas variables consideradas en este estudio, se procedió a analizar la existencia de posibles asociaciones multivariantes entre el estado de publicación --tratada como variable dependiente de los modelos-- y las restantes variables, introducidas en los modelos como independientes.

Con este propósito, para cada revista se construyó un modelo saturado con todas las variables independientes (procedencia del estudio, tamaño muestral, objeto del estudio, tipo de diseño, complejidad del análisis estadístico y calidad) y los respectivos términos de interacción. Luego de realizar una regresión logística de la covarianza (*logistic analysis of covariance*) paso a paso (en la cual se eliminan del modelo todas aquellas variables independientes que no se asocian con la dependiente, utilizando como criterio para ello la ausencia de resultados estadísticamente significativos en la prueba de significación de Wald de los coeficientes de regresión, que sigue una distribución de ji cuadrado), se obtuvieron los modelos finales, que se describen a continuación.

En la tabla 39 se presenta el modelo multivariante final para *Atención Primaria*. Como puede observarse, a nivel multivariante, las dos únicas variables independientes asociadas con el estado de publicación fueron el tipo de estudio y su calidad. La fuerza de la asociación entre el tipo de estudio y el estado de publicación fue $RP = 4,30$ (IC95%: 1,45 - 12,50), y entre la calidad y el estado de publicación, $RP = 10,42$ (IC95%: 3,45 - 31,25).

TABLA 39. Resultados del modelo de regresión logística final de la asociación entre el estado de publicación y las variables consideradas en este estudio correspondiente a *Atención Primaria*.

	β	EE (β)	RP	IC 95%
Constante	1,6511	0,4524		
Tipo de estudio	-1,4596	0,5605	4,30	1,45 - 12,50
Calidad	-2,3437	0,5671	10,42	3,45 - 31,23

Bondad del ajuste: $\chi^2_{(2)} = 38,218$; $P = 0,0001$.

Estos datos significan, primero, que en *Atención Primaria* el tipo de estudio influye en la publicación o el rechazo de un artículo. Este resultado del análisis multivariante corresponde al resultado obtenido al analizar la asociación entre tipo de estudio y estado de publicación en el análisis bivariante: *Atención Primaria* publica más estudios de prevalencia de los que rechaza y rechaza más estudios descriptivos de los que publica. Segundo, la posibilidad de que la revista publique un estudio de calidad

alta es entre 3,45 y 31,25 veces mayor que la posibilidad de que publique un estudio de calidad baja. Además, los datos indican que la asociación de la calidad con el estado de publicación es más fuerte que la de ésta última variable y el tipo de estudio.

Al interpretar estos resultados es preciso tener en cuenta que estas *RP* son razones ajustadas según la otra variable incluida en el modelo. En otras palabras, la *RP* estimada de la asociación entre el tipo de estudio y el estado de publicación es 4,30 cuando la calidad de los artículos se mantiene constante. Por lo tanto, al margen de que la calidad varíe o no según el tipo de estudio, en *Atención Primaria* es más probable que se publiquen o rechacen artículos en los que se describen los diseños mencionados.

La interpretación de la *RP* correspondiente a la asociación entre estado de publicación y calidad de los estudios sigue el mismo razonamiento: en *Atención Primaria* es más probable que se publique un estudio de calidad intermedia que baja, y ello es independiente de la asociación entre el estado de publicación y el tipo de estudio, pues la *RP* de esta última asociación se estimó manteniendo constante el tipo de estudio o diseño.

En lo que hace referencia a la interpretación de estos modelos de regresión logística cabe decir por último que el modelo tiene una validez aceptable; es decir, el modelo se adapta, se ajusta o explica razonablemente bien los datos observados, por cuanto la prueba de bondad del ajuste utilizada ($\chi^2 = 38,218$; $P = 0,0001$) es estadísticamente muy significativa (tabla 39).

El modelo final de regresión logística correspondiente a *Medicina Clínica* se resume en la tabla 40.

TABLA 40. Resultados del modelo de regresión logística final de la asociación entre el estado de publicación y las variables consideradas en este estudio correspondiente a *Medicina Clínica*.

	B	EE(B)	RP	IC 95%
Constante	2,0369	0,6138		
Calidad	-3,3950	0,7187	30,3	7,14 - 125,0

Bondad del ajuste: $\chi^2_{(1)} = 33,345$; $P = 0,0001$.

En esta revista sólo la calidad se asoció a nivel multivariante con el estado de publicación. La fuerza de esta asociación fue $RP = 30,3$ ($IC95\%: 7,14 - 125,0$). Es decir, en *Medicina Clínica* la posibilidad de que un artículo de calidad alta se publique es de 7,14 a 125 veces más elevada que la que se publique un estudio de calidad baja. La validez de este modelo también es aceptable, ya que la prueba de bondad de ajuste también fue altamente significativa ($\chi^2 = 33,345$; $P = 0,0001$). Parafraseando, el modelo se ajusta o explica los datos observados en esta revista de modo razonable.

Finalmente, la ausencia en ambos modelos de términos de interacción estadísticamente significativos entre variables confirma el cumplimiento del supuesto de paralelismo.

Por consiguiente, a juzgar por los resultados del análisis multivariante, se puede concluir que tanto en *Atención Primaria* como en *Medicina Clínica* la calidad es la variable asociada por antonomasia con el estado de publicación, o sea, que en ambas revistas la calidad parece regir la selección de artículos para publicación. Además, en *Atención Primaria*, el tipo de estudio también se asocia con la publicación o el rechazo de los artículos, al margen de la calidad de éstos, y que esta asociación es tributaria de los estudios de prevalencia y descriptivos.

CAPÍTULO 5. DISCUSIÓN

5.1 Comentarios generales

El principal valor de esta memoria de investigación estriba en tres hechos íntimamente relacionados entre sí y estrechamente vinculados con sus objetivos. Primero, ofrece un modelo que permite analizar y monitorizar factores que rigen la selección de artículos científicos originales para publicación. Segundo, dicho modelo se basa en un cuestionario y una escala originales y confiables, que fueron construidos para medir la calidad de los artículos recibidos por las revistas para publicación. Y, tercero, como resultante de los anteriores, que dicho modelo permite estimar la existencia de sesgos de publicación en el proceso de selección de manuscritos.

De estas tres razones, la estimación de la calidad de los artículos constituye la piedra angular del estudio. La principal limitación de las investigaciones realizadas hasta la fecha sobre este tema es que en ellas no se han analizado simultáneamente factores importantes que intervienen en la revisión y la selección de artículos para publicación. Unos estudios han sido más exhaustivos que otros en considerar las variables que vertebran la revisión y selección de manuscritos, pero su revisión minuciosa revela la omisión sistemática de la calidad como variable interviniente esencial en la toma de decisiones editoriales a este respecto.

Como se mencionó en las secciones que anteceden, la estimación de la *calidad* de un artículo entraña grandes dificultades. No obstante, en esta memoria se ha propuesto una definición operacional de este concepto y un instrumento para medirlo cuya validez, a juzgar por los resultados obtenidos, es aceptable.

En estos hechos radica la innovación que aporta el estudio: la posibilidad de medir con un grado de rigor aceptable la calidad de un artículo científico original y, con ello, analizar la validez de la revisión y selección editoriales de manuscritos. El escrutinio de la

validez de estos dos procesos, para ser completo, debe incluir obligatoriamente el estudio de los sesgos de publicación.

Para cumplir con los objetivos marcados fue preciso ampliar el marco de utilización del diseño de casos y controles (Breslow NE, Day NE, 1980; Schlesselman JJ, 1982). Por añadidura, a diferencia de la mayor parte de estudios precedentes, en esta memoria de investigación se introdujo de modo novedoso la evaluación de artículos rechazados por las revistas estudiadas. Sin esta evaluación no es posible obtener una imagen válida, y sobre todo completa, del entramado de factores que conforman la publicación de artículos originales. Sólo es posible analizar con rigor la validez del proceso de publicación cuando dicho análisis incorpora artículos rechazados. La omisión de éstos torna el análisis incompleto.

El estudio se llevó a cabo con 196 de los 200 artículos que en la etapa de diseño de esta memoria de investigación se decidió evaluar. Por tanto, la pérdida de observaciones fue exigua. Los artículos excluidos del estudio fueron incluidos en la muestra por los responsables de las dos revistas de forma desapercibida, es decir, sin reparar en que no cumplían el criterio de inclusión relativo al tipo de estudio. Dado que la exclusión de estos cuatro artículos sólo responde a la razón mencionada, a este nivel del muestreo no se introdujo sesgo alguno de selección en la muestra.

En la práctica, puede considerarse que también se alcanzó el tamaño muestral preestablecido para el diseño de casos y controles empleado: la muestra final estuvo constituida por 100 artículos publicados (casos) –50 por cada revista– y por 96 rechazados (controles) de los 100 rechazados (controles) que dicta el diseño de la muestra.

5.2 Estado de publicación

Los 4 artículos que se excluyeron de este estudio eran controles, es decir, artículos rechazados. Esta cifra es exigua en relación con el tamaño de la muestra final. Por ello, aunque la razón final de casos y controles (1:0,96) no coincide con la fijada en el diseño (1:1), la diferencia es nimia y no introduce desequilibrio alguno en el diseño ni sesgos en esta etapa de selección.

5.3 Calidad de los artículos

El primer resultado que cabe destacar de los artículos analizados en relación con su calidad es que la media de esta variable —medida en una escala de 1 a 5— se encuentra en el rango de calidad de 1,5 a 2,9. La precisión de las estimaciones de este parámetro puede considerarse bastante aceptable. Si bien la media no es una medida robusta de la tendencia central de una variable (por cuanto está muy influida por los valores extremos de ésta), en la serie de artículos evaluados la variabilidad de las medias estimadas fue baja, como lo demuestra la baja magnitud de los errores estándar de las medias y, por ende, la baja amplitud de los intervalos de confianza calculados.

En promedio, por consiguiente, la calidad de los artículos evaluados osciló entre intermedia y baja. Este dato ha de interpretarse con suma cautela. De antemano, es preciso adoptar un referente de calidad, como, por ejemplo, el que proveen las revistas de medicina más prestigiosas cuya cobertura temática es similar a la de *Atención Primaria y Medicina Clínica*.

Las diferencias de calidad entre ambos grupos de revistas pueden atribuirse a un entremado complejo de factores. Si los recursos de investigación disponibles (formación y experiencia investigadoras, medios materiales y financieros y asesoramiento para la investigación) de los autores de artículos enviados a *Atención Primaria y Medicina Clínica*

son más precarios que los de los autores que publican en las revistas de renombre, cabe esperar que la calidad de los estudios que se remitan a publicación y se publiquen en estas dos revistas sea menor. Por añadidura, cierto número desconocido de artículos de alta calidad escritos por investigadores españoles se <<fugan>> a revistas de mayor prestigio y factor de impacto por motivos cuyo análisis trasciende los objetivos de esta memoria de investigación. Por otra parte, es poco factible que la capacidad y experiencia de revisores y editores de las dos revistas evaluadas para seleccionar los mejores artículos sea menor que la que exhiben los de revistas de mucho mayor prestigio, pues, como se comentará más adelante, los resultados obtenidos demuestran su marcada capacidad de distinguir y seleccionar para publicación los artículos de calidad más elevada. A todo ello cabe añadir que, en ocasiones, la baja demanda de publicación pueda forzar a los editores a seleccionar artículos de menor calidad de la deseada (diversas comunicaciones personales, 1998-1996, cuya procedencia debe mantenerse en el anonimato por razones éticas).

Por otro lado, la calidad de los artículos publicados fue sistemáticamente mayor que la de los rechazados y la diferencia observada no es atribuible al azar. Este resultado es extremadamente importante, como se detalla a continuación. En primer lugar, revela que el sistema de revisión y selección de artículos de las dos revistas (y, por ende, de editores y revisores) cumple con su cometido primigenio: aceptar los artículos de mayor calidad que reciben y rechazar los de calidad menor. Por lo tanto, si bien la calidad de los estudios que intentan publicarse en *Medicina Clínica y Atención Primaria* (lo que podríamos denominar bajo el epígrafe <<la cantera de estudios>>) fue intermedia y baja, los revisores y editores de estas revistas seleccionan entre ellos para publicación los de mayor calidad. Segundo, pone de manifiesto que, al margen de que otros factores puedan influir e incluso sesgar la selección de estudios para publicación, la calidad, como cabía esperar, rige claramente su selección en las dos revistas, lo cual confirma una de las hipótesis de esta memoria.

5.3.1 Calidad de las secciones

La diferencia observada entre las medias de la calidad global de los artículos publicados y rechazados por ambas revistas se convierte en un hallazgo congruente al constatar que la calidad de cada una de las seis secciones de los artículos publicados fue mayor que la de los rechazados. Por consiguiente, la diferencia relativa a la calidad global es congruente, no resulta de la influencia en la media global de valores extremos de calidad elevados de alguna sección y, con una confianza del 95%, no es atribuible al azar o a la variabilidad del muestreo.

La Introducción fue la sección de calidad más alta de los artículos publicados y rechazados por ambas revistas. Es legítimo pensar que la notificación del marco general del estudio, sus antecedentes y la formulación de hipótesis e objetivos entrañen menor complejidad que la descripción de los métodos empleados, la presentación acendrada de los resultados y su correcta interpretación. Acaso ello explique que, si bien la calidad global de las introducciones no es alta, exceda sistemáticamente la de los restantes apartados del artículo científico original.

Asimismo, el lugar que ocupan en la jerarquía de calidad los restantes apartados fue muy pareja en todos los artículos, excepto los publicados por *Medicina Clínica*. Los Resultados ocuparon el segundo, seguidos de Materiales y Métodos y de la Discusión, lo cual puede interpretarse siguiendo el mismo razonamiento que en la Introducción. No obstante, en los artículos publicados por *Medicina Clínica* la calidad de los Resultados fue menor que la de la Discusión, la sección supuestamente más intrincada del artículo.

En cualquier caso, la interpretación más detallada de estos hallazgos exige analizar pormenorizadamente la calidad de los subapartados que integran estas cuatro secciones.

5.3.2 Calidad de los subapartados

Al pormenorizar el análisis de la calidad de la Introducción se aprecia que la de los cinco subapartados de esta sección fue mayor en los artículos aprobados que en los rechazados de las dos revistas. Este hallazgo pone de relieve nuevamente que sus sistemas de revisión editorial discernen los artículos de alta y baja calidad en relación con los propósitos de la Introducción: descripción del marco en que se realizaron los estudios, formulación del problema y pregunta de investigación y su respaldo conceptual y bibliográfico, indicación de la utilidad y aplicabilidad del estudio, y definición de sus objetivos e hipótesis. Sin embargo, la calidad de estos subapartados sólo fue mayor de 3,0 en los artículos publicados por *Medicina Clínica*, lo cual puede responder a los mismos motivos que explican el gradiente de calidad (de intermedia a baja) de los artículos revisados en esta memoria de investigación.

Es preciso destacar que la menor calidad de esta sección correspondió al subapartado relativo a la formulación de objetivos e hipótesis. Ello es debido sobre todo a la imprecisión y vaguedad con que se definen, un hallazgo encontrado reiteradamente en la literatura nacional e internacional, no sólo de artículos originales, sino también en protocolos de investigaciones, evaluaciones de programas y servicios de salud, y en análisis y síntesis de políticas de salud (Majchrzak A, 1984; Dever GEA, 1984; Ginzberg E, 1991; Pathmanatan I, Nik-Safiah NI, 1991; Starfield B, 1985; US Preventive Services Task Force, 1989; Ganiats TG, Wong AF, 1991; Jefferson T, Demicheli V, 1995).

La calidad de los seis subapartados de la sección de Materiales y Métodos de los artículos publicados por las dos revistas también fue mayor que la de los rechazados, lo cual se atribuye a las mismas razones esgrimidas en los párrafos que anteceden.

En esta sección sobresale la deficiente descripción de los procedimientos de recolección, control de calidad y análisis de los datos, así como del diseño utilizado, de los

métodos de análisis y del muestreo. Los artículos incumplen el requisito de incluir suficiente información para que un investigador avezado pueda repetir el estudio. La insuficiente descripción de la selección de la muestra limita la capacidad del lector de juzgar su grado de representatividad y la posibilidad de que se hayan introducido distintos tipos de sesgos (como el de selección). En muchos casos no se describe el tipo de muestreo ni los criterios de inclusión y exclusión de los participantes en el estudio. También es frecuente la ausencia de los supuestos aceptados para calcular el tamaño muestral y establecer la potencia estadística del estudio. Por añadidura, a menudo no se presentan con suficiente nivel de detalle la unidad o las unidades de análisis ni los métodos, modelos y pruebas estadísticas aplicadas. Ello impide valorar la idoneidad de su utilización y el rigor de los resultados presentados, interpretarlos adecuadamente y apreciar algunas de las posibles limitaciones del estudio. Todas las deficiencias señaladas limitan a la postre la posibilidad de que el lector evalúe la parcela de la validez interna y externa del estudio dependiente de estos datos.

Por otro lado, la idoneidad del diseño utilizado para los objetivos e hipótesis del estudio fue el subapartado de mayor calidad. En algunos estudios revisados se emplean diseños que permiten responder, en principio, a los objetivos del estudio y contrastar sus hipótesis. Empero, la deficiente descripción de la información que debe incluirse en Materiales y Métodos restringe la capacidad de valorar con rigor la idoneidad real del diseño y, por ende, sus limitaciones. En otros estudios, por el contrario, no se utilizan diseños adecuados para alcanzar los objetivos fijados, fundamentalmente en aquellos en los que se deben realizar extrapolaciones, evaluar la efectividad de programas y servicios de salud o sugerir relaciones de causalidad.

El patrón de la calidad de los subapartados de la sección de Resultados fue idéntico al de las dos secciones precedentes. La calidad de los tres subapartados en los artículos

publicados por *Medicina Clínica y Atención Primaria* fue mayor que la de los rechazados. No obstante los valores bajos de calidad estimados (ningún límite superior de los intervalos de confianza de 95% fue mayor de 3,0), la construcción y corrección de los modelos de análisis fue el subapartado de calidad más alta, seguido por la del muestreo y, finalmente, por la adecuación de la presentación general y específica de los resultados.

La baja calidad estimada para estos tres subapartados es atribuible a diversas causas. Primero, en algunos estudios figuran los resultados de análisis no descritos en Materiales y Métodos. En otros, los resultados presentados corresponden a análisis que no permiten responder a los interrogantes de la investigación. Segundo, se observa una tendencia muy marcada y generalizada a limitar la presentación de resultados a un conjunto de valores P . Ello trasluce la obstinación inopinada y vana de muchos investigadores en obtener valores P significativos sin tener en cuenta las limitaciones y peligros que encierra el empleo exclusivo de pruebas de hipótesis (Gardner MJ, Altman DG, 1986; Walter SD, 1995; Campillo C 1996), así como su reticencia --contraviniendo las recomendaciones de muchas revistas y comités internacionales de revistas médicas-- a realizar, en su lugar, pruebas de estimación (Altman DG, Gore SM, Gardner MJ, Pokock SJ, 1983; Gardner MJ, Altman DG, 1986; Home PD, 1988; Bailar III JC, Mosteller F, 1988; Campillo C, 1996). Tercero, pocos artículos incluyen medidas de la variabilidad e incertidumbre de las estimaciones presentadas y de los errores de las mediciones efectuadas. Cuarto, cuando se utilizan diseños y se construyen modelos de complejidad intermedia o alta (grados 2 y 3 de complejidad en el cuestionario) en muy pocos casos se incluyen los resultados del análisis de la validez de los modelos, lo cual impide conocer en qué medida los modelos explican o se ajustan a los datos observados. Este problema se ha puesto de relieve frecuentemente en la literatura, al igual que la necesidad de formular recomendaciones específicas para la presentación de modelos lineales generalizados (Hosmer DW, Lomeshow S, 1991; Concato

J, Feinstein AR, Holford TR, 1993; Campillo C, 1993). Por último, es preciso resaltar que en algunos artículos se presentan resultados de análisis complejos (por ejemplo, modelos lineales generalizados y otros análisis multivariantes) omitiendo los resultados de análisis más sencillos —acaso porque no se realizaron— que deben preceder la construcción de modelos de esa complejidad (vgr., asociaciones bivariantes entre la variable dependiente y las independientes, estudio de la multicolinealidad, análisis de interacciones y de factores de confusión putativos) o sin comprobar que se cumplen los supuestos, a veces intrincados, de estos modelos. Si tales requisitos no se cumplen, los resultados carecen de sentido y el estudio entorpece el avance del conocimiento por limitado que sea su alcance. El grado de complejidad de un modelo o de un diseño *per se* no confiere mayor importancia a los resultados ni aumenta el rigor de un análisis, sobre todo cuando éste no responde fehacientemente a los objetivos e hipótesis del estudio o cuando carece de fundamento formal.

La calidad de los nueve subapartados que conforman la Discusión fue entre intermedia y baja en las dos revistas, y la de los artículos publicados fue mayor que la de los rechazados. Las medias de calidad de los publicados fueron estadísticamente distintas de las de los rechazados. Por consiguiente, los sistemas de revisión de *Medicina Clínica* y *Atención Primaria* también discernen la calidad de los artículos en relación con este subapartado. Sólo cinco límites superiores de los intervalos de confianza de las 36 medias estimadas fueron mayores de 3,0. Dos corresponden a las medias de subapartados de artículos publicados por *Atención Primaria* y tres, a publicados por *Medicina Clínica* (dos de esos subapartados coinciden en ambas revistas).

A la luz de estos hallazgos puede afirmarse, en primer lugar, que, en promedio y en ambas revistas, el nivel de congruencia de la interpretación de los resultados respecto a los objetivos e hipótesis de los estudios varía entre bajo e intermedio y que los estudios, a pesar

de sus notables limitaciones metodológicas y formales, aportan datos novedosos o necesarios. Por tanto, las conclusiones extraídas en los artículos publicados por las dos revistas se ajustan en cierta medida a los objetivos. Debe hacerse hincapié en que, sobre todo en algunos artículos rechazados, dicha congruencia es falaz, pues la interpretación de los resultados, aunque es acorde con los objetivos, resulta metodológicamente incorrecta. A modo de ejemplo valga citar que las inferencias de algunos estudios responden a sus objetivos e hipótesis, pero soslayan los requisitos que dichas inferencias han de cumplir, establecen relaciones de causalidad sin cumplir con los criterios de causalidad aceptados (Hill AB, 1965) o atribuyen resultados a la efectividad del programa o servicio de salud evaluado sin utilizar un diseño que permita hacer dichas atribuciones.

En el subapartado referente a la capacidad de extraer las conclusiones presentadas a partir de los métodos y resultados de los estudios, las medias de calidad estimadas oscilaron entre 2,53 y 3,19 en los artículos publicados en ambas revistas y entre 1,69 y 2,24 en los rechazados. Por consiguiente, fueron estadísticamente distintas. En general, las conclusiones de los artículos trasgreden los límites de extrapolación que imponen el diseño, los análisis y los resultados. En otras palabras, el alcance de las conclusiones suele exceder los límites a que deberían circunscribirse si, en rigor, se ajustaran a la capacidad de extraerlas que acotan el diseño, la muestra, el análisis y los resultados. Dos de los más claros exponentes de ello los constituyen la mencionada atribución de efectividad en evaluaciones de programas y servicios y el establecimiento de relaciones de causalidad cuando el diseño o el estudio en conjunto no lo permiten desde un punto de vista metodológico. En síntesis, y desde otra perspectiva, puede afirmarse que en ciertos artículos se sobreestima tanto la validez interna como la externa del estudio que notifican.

La media de la calidad de los restantes subapartados de la Discusión fue baja o muy baja. Así, raras veces se comparan e interpretan los resultados teniendo en cuenta

realmente los de otros estudios pertinentes. En buena parte de los artículos sólo se sacan a colación formalmente resultados de otros estudios, pero no se interpretan críticamente poniendo de relieve sus limitaciones y circunscribiendo sus resultados a las peculiaridades del estudio realizado. Tampoco se señalan faltas de asociación ni discrepancias con otros estudios, ni se intenta buscar los motivos que puedan explicarlas. En muy pocos artículos se formulan preguntas genuinas para futuras investigaciones, ni se acota con precisión el marco de futuras investigaciones encaminadas a ahondar en el tema o a superar los obstáculos que han impedido contrastar las hipótesis o alcanzar los objetivos completamente, ni se exponen de forma diáfana las consecuencias teóricas y aplicaciones prácticas de los estudios notificados.

Al margen de la importancia de estas deficiencias, su gravedad estriba en que se aprecian con bastante asiduidad en artículos publicados. Estos hallazgos coinciden con los de diversas evaluaciones de artículos científicos realizadas en el seno de revistas nacionales e internacionales al tiempo que confirman las limitaciones y deficiencias también señaladas en evaluaciones publicadas de artículos de revisión de la literatura (Feinstein AR, 1974; Emerson JD, Colditz GA, 1983; Nolasco A, Gascón E, Mur P, Ferrándiz E, Álvarez-Dardet C, 1986; Álvarez-Dardet C, Mur P, Gascón E, Nolasco , Bolumar F, 1987; Bailar III JC, Mosteller F, 1988; Bailar III JC, Mosteller F, 1992; 1992; Silva LC, Pérez-Nieves, Cuellar L, 1995; González JC, Pulido M, Sanz F, 1995; Gómez C, et al., 1996). Al cabo, las revisiones de la literatura no deberían distinguirse de los artículos científicos originales en los métodos científicos empleados, sino solamente en sus cometidos y en la unidad de análisis (Mulrow CD, 1987; Oxman AD, Guyatt GH, 1988).

Por último, la notificación explícita y clara de las limitaciones del estudio fue el subapartado de menor calidad de la Discusión. En este caso, las medias de calidad de los artículos publicados también fueron distintas de las de los rechazados y, tanto en unos como

en otros, bajas o muy bajas. Este hecho pone de relieve dos fenómenos palmarios. Primero, la reticencia a reconocer los errores en que se ha incurrido en una investigación y las limitaciones inherentes al estudio, sobre todo las de índole metodológica. Obviamente, muchos autores parecen escudarse en la falacia de que la notificación de dichas limitaciones resta importancia o valor al estudio o que de algún modo lo desmerece (Szklo M, comunicación personal, 1992; Day R, 1995). Nada más alejado de la realidad: ninguna investigación está exenta de defectos y limitaciones metodológicas (tales como sesgos, factores de confusión, bajo tamaño muestral y baja potencia estadística, poca representatividad muestral o baja validez externa), y es su reconocimiento explícito --no su omisión-- lo que aumenta la calidad de la notificación del estudio y deja traslucir el rigor y la capacidad analítica de los investigadores. Segundo, en muchos casos es posible concluir sin rebujos que este problema es fruto de la baja formación en investigación de los autores. Este desconocimiento les impide o limita su capacidad para detectar y escrutar las limitaciones del estudio y, por ende, notificarlas debidamente.

Todo ello es consecuencia de que en la Discusión, por un lado, se concentran y ponen de manifiesto las consecuencias de las limitaciones señaladas en las secciones precedentes. Y, por otro, de que es en ella donde el autor debe mostrar su capacidad crítica, de interpretación, de análisis y de síntesis (Kepple SR, 1991; International Committee of Medical Journal Editors, 1993; Day R, 1995). La deficiente formación en investigación mentada en el párrafo que antecede también da cuenta del nivel de calidad estimado para cada uno de los subapartados de la Discusión.

La estimación de la calidad de las referencias bibliográficas de los artículos se fundamentó en su pertinencia, actualidad e integridad (International Committee of Medical Journal Editors, 1993; Day R, 1995). En ambas revistas las medias de la calidad de las referencias de los artículos publicados fueron más altas que la de los rechazados, si bien

sólo fueron distintas estadísticamente las de *Medicina Clínica*. Por otra parte, la calidad varió entre baja (sobre todo en los artículos rechazados) y baja e intermedia en los publicados, lo cual se atribuye a los siguientes motivos. En algunos artículos no se respaldan con referencias numerosas afirmaciones, lo cual contraviene las normas internacionales de edición (International Committee of Medical Journal Editors, 1993). Asimismo, muchos artículos contienen mayoritariamente referencias en castellano, lo cual traduce a buen seguro el desconocimiento que tienen los autores de otros idiomas. Ello reduce sobremanera el volumen y la calidad de la información de que se parte para diseñar y acometer una investigación, así como el rigor de las interpretaciones de los resultados que se obtienen y de la Discusión en general, pues reduce la disponibilidad de datos necesarios para cumplimentar adecuadamente esta sección.

Llama también la atención el escaso número de referencias que muchos artículos incluyen respecto al alcance de sus afirmaciones. Ello puede deberse básicamente a dos razones: falta de acceso a fuentes bibliográficas --computadorizadas o manuales-- y a ineficiencia de las búsquedas en los casos en que se tenga acceso a ellas. Esta deficiencia es un problema universal que se ha estudiado ampliamente (Poynard T, Conn HO, 1985; Gotzsche PC, Lange B, 1991; Jadad AR, McQuay HJ, 1993; Dickersin K, Scherer R, Lefebvre C, 1994). En este sentido, ha de tenerse en cuenta el lugar geográfico en que se realizó el estudio y los escollos que ello crea para tener acceso a la información. Es muy probable que la poca actualización de las referencias detectada en algunos artículos responda a los mismos motivos.

Por último, la calidad de la presentación de la información se estimó utilizando como referentes y criterios las normas de presentación recomendadas por comités internacionales de editores de revistas médicas y por instituciones integradas por editores y expertos en metodología de la investigación (International Committee of Medical Journal

Editors, 1993; Altman DG, Gore SM, Gardner MJ, Pocock SJ, 1983; Gardner MJ, Altman DG, 1986; Home PD, 1988; Bailar III JC, Mosteller F, 1988). No se tuvo en cuenta, por el contrario, las normas de presentación y las recomendaciones propias de *Atención Primaria* ni de *Medicina Clínica*, ya que no se formuló como objetivo de esta memoria de investigación evaluar el grado de adherencia de estas revistas a sus propias normas ni a las internacionales. Valga añadir que en esta memoria las normas de los comités internacionales, y no las de alguna revista concreta, se aceptaron como el patrón normativo de referencia con fines de evaluación, pues las primeras surgieron de un consenso más amplio y heterogéneo en cuanto a la formación, especialidad y opiniones de sus integrantes, que las del consejo editorial de alguna revista particular. Por añadidura, las recomendaciones de estos comités son aceptadas cada vez por más revistas como normas de presentación preceptivas. De hecho, constituyen en la actualidad un requisito de indización y aceptación en muchas bases de datos y registros de estudios y su formulación precede en buena medida a las de las revistas individuales (Easterbrook PJ, 1992). No obstante, y a causa de estas suposiciones y de los criterios adoptados, los resultados obtenidos pueden servir de estimación indirecta del grado de adherencia de las dos revistas a esas recomendaciones y sugerir indirectamente medidas que pueden adoptarse para aumentar la calidad de la presentación y notificación de resultados.

La calidad de los dos subapartados de la Presentación fue intermedia. En el caso de los artículos publicados, este resultado puede interpretarse como medida aproximada del grado de adherencia de las dos revistas a los requisitos internacionales, ya que el cuestionario empleado para evaluar los artículos se basa en dichos requisitos. La calidad estimada del primer subapartado expresa la facilidad con que se leen, la calidad de la sintaxis y de la lexis, así como la adecuación de la información presentada a la estructura del artículo (Day R, 1995). En general, la información de los artículos revisados está bien

estructurada, en el sentido de que las secciones de los artículos (IMRYD) no contienen datos que pertenezcan a otras secciones. Sin embargo, se observan numerosos errores de sintaxis y anglicismos (más abundantes en los artículos rechazados), que podrían subsanarse si en las revistas se revisara el estilo de un modo acendrado. Es deseable mejorar, además, la presentación de datos en tablas y figuras, dado que en algunos casos se presenta la misma información en ellos y en el texto, en otros no está justificado emplearlos --pues la información podría resumirse en frases breves, como lo ejemplifican las meras enumeraciones-- y en otros --sobre todo las presentaciones de modelos estadísticos en tablas-- carecen de datos esenciales para que el lector pueda interpretarlos y valorar su bondad de ajuste, una deficiencia encontrada con frecuencia en revistas médicas (Hosmer DW, Lomeshow S, 1991; Campillo C, 1993).

La calidad del último subapartado, el referente al resumen, osciló entre baja e intermedia, lo cual entraña una limitación importante de la calidad global de los artículos por cuanto el resumen es lo primero que el lector consulta a fin de decidir si debe leer el artículo o no. Un agravante es que el resumen suele indizarse y aparecer en muchas bases de datos y revistas, es decir, es la parte del artículo a la que los lectores suelen tener acceso en primera instancia. Muchos resúmenes, aunque son informativos (en contraposición a los descriptivos) no reúnen toda la información básica que estipulan las normas internacionales de elaboración de resúmenes (Mulrow CD, Stepher B, Thacker SB, Ough JA, 1988; Haynes RB, Mulrow CD, Huth J, Altman DG, Gardner MJ, 1990). Debe hacerse hincapié, no obstante, en que la reciente adopción por ambas revistas de resúmenes estructurados ha supuesto un mejoramiento palmario de la capacidad informativa de los resúmenes que publican.

Empero, algunos de los resúmenes revisados pueden catalogarse como híbridos, es decir, todavía incluyen información y comparten atributos (fundamentalmente en los

apartados de antecedentes y conclusiones) que son más propios de los resúmenes descriptivos que de los informativos. La calidad global estimada de este subapartado destaca la magnitud de los cambios que les resta introducir en ellos, y que se reducen en definitiva a la selección de las parcelas de información y datos específicos que todavía han de añadirse a los resúmenes. Estos comentarios son congruentes con el hecho de que la calidad de los resúmenes de los artículos publicados y rechazados de *Atención Primaria* y de *Medicina Clínica* no sea demasiado dispareja. Ambas revistas han concluido la primera etapa: estructurar sus resúmenes. El paso siguiente consiste en seleccionar con criterios muy precisos la información que se incluya en cada sección de dichos resúmenes.

5.4 Procedencia de los artículos

En varias comunicaciones personales (que deben mantenerse en el anonimato por razones éticas) se señaló que las dos revistas aceptaban sistemáticamente más artículos procedentes de Cataluña que de otras regiones del territorio español y que la calidad de los artículos no intervenía en esta selección sesgada. Este elemento introducido sistemáticamente en la aceptación editorial se acuñó para los propósitos de este estudio bajo el epígrafe de *sesgo de procedencia*. Asimismo, y habida cuenta de las similitudes que guarda este sesgo con el de publicación primigenio (Smith ML, 1980), se decidió incluirlo en el acápite genérico de sesgos de publicación. Por estos motivos, se investigó si este sesgo desempeñó algún papel en los procesos de selección de artículos en *Atención Primaria* y *Medicina Clínica*. Como ponen de manifiesto los resultados, *Atención Primaria* recibió más artículos de Cataluña que de otras regiones que *Medicina Clínica*, pero la procedencia no se asoció con su publicación o rechazo en ninguna de las dos revistas consideradas en conjunto ni por separado.

Por otra parte, el análisis combinado de ambas revistas mostró que la calidad se asoció con la procedencia de los artículos. Es más probable que la calidad de un artículo procedente de Cataluña sea más baja que la de uno procedente de otra región. No obstante, cuando se estratificaron los resultados, dicha asociación no se encontró en *Atención Primaria*. Por el contrario, en *Medicina Clínica* la asociación fue palmaria: fue más probable que la calidad de un artículo proveniente de Cataluña fuese más baja que la de un artículo de otra región.

No se detectó asociación entre la procedencia y el estado de publicación de un artículo al analizar ambas revistas en conjunto ni al estratificar dicha asociación por revista. Si este hallazgo se interpreta teniendo en cuenta las asociaciones encontradas entre procedencia y calidad y entre estado de publicación y calidad, se plantea, al menos en apariencia, una contradicción: por un lado, aunque la aceptación y el rechazo de artículos dependen de su calidad y son independientes de su procedencia, la calidad se asoció con la procedencia en el análisis combinado y en *Medicina Clínica*. Por tanto, cabe preguntarse: ¿por qué el estado de publicación no se asocia con la procedencia de los artículos cuando la calidad de los que proceden de otra región es más alta que la de los Cataluña en el análisis combinado y en *Medicina Clínica*?

La respuesta a este interrogante la proporciona el análisis estratificado. Primero, la asociación entre procedencia y calidad del análisis combinado fue tributaria de la asociación encontrada entre estas dos variables en *Medicina Clínica*. Segundo, en esta revista, aunque la publicación o el rechazo de un artículo no dependen de su procedencia, la calidad de los artículos de otra región fue más alta que la de los de Cataluña. Por consiguiente, es lícito esperar que los artículos de Cataluña se rechazasen más que los de otra región. Si se analiza en detalle el cuadro 14 correspondiente a *Medicina Clínica*, se aprecia que, aunque el intervalo de confianza del 95% de la *RP* de la asociación entre

procedencia y estado de publicación contiene del valor 1,0 de ausencia de asociación, la mayor parte de sus valores son menores de 1,0. Ello indica que es posible que el verdadero valor del parámetro sea menor de 1,0. Es legítimo pensar entonces que, si el tamaño de esta submuestra fuese mayor, la precisión de la estimación de la *RP* aumentaría y, por tanto, la amplitud del intervalo de confianza se reduciría para comprender solamente valores menores de 1,0. Ello significaría que la posibilidad de que se publique un artículo procedente de otra región sería más alta que la de un artículo de Cataluña. El que los resultados sean coherentes con la ausencia de asociación se debe a que el intervalo de confianza contiene el valor 1,0, lo cual acaso sea atribuible al pequeño tamaño de la submuestra.

De este modo, es posible que *Medicina Clínica* rechace más artículos de Cataluña, no porque introduzca un sesgo de procedencia en la selección de los artículos para publicación, sino porque la calidad de los artículos de Cataluña es en conjunto más baja. Así, se concluye que ni en el proceso de selección de artículos de *Atención Primaria* ni en el de *Medicina Clínica* se introduce sesgo de procedencia, lo cual desmiente el contenido de las comunicaciones personales mentadas.

5.5 Tamaño de la muestra

En 92,86% (IC95%: 89,26% - 96,46%) de los artículos revisados se estudiaron muestras. De ahí se colige que sólo en 7,14% de los artículos que reciben ambas revistas se estudiaron poblaciones. En estos artículos se analizaban bases de datos como registros o estadísticas poblacionales.

Por otro lado, ni en el análisis conjunto ni por separado de las dos revistas el tamaño muestral se asoció con el estado de publicación. Sobre este particular cabe señalar que en algunos estudios se ha indicado que el pequeño tamaño de las muestras a veces se interpreta

sistemática e injustificadamente como una limitación universal de las investigaciones (Meinert CL, 1986). Si bien ello puede ser cierto cuando las muestras son extremadamente pequeñas y, sobre todo, cuando, además de su pequeño tamaño, no son representativas, en general cabe esperar que el tamaño muestral *per se* no determine ni menoscabe el rigor de una investigación. Si el estudio se ha llevado a cabo rigurosamente desde un punto de vista metodológico, el pequeño tamaño muestral por sí solo reduce la precisión de las estimaciones y afecta su validez externa en la medida en que la muestra sea más o menos representativa de la población de la cual procede. Lamentablemente, la limitación de recursos –entre los que se incluye a menudo la falta de datos– obliga a los investigadores a trabajar con muestras de tamaño menor del deseable. Ello explica parte de la difusión del metanálisis, de las revisiones sistemáticas y, más recientemente, de la Colaboración Cochrane (Chalmers I, 1992) como herramientas destinadas a minimizar los efectos, entre otras muchas cosas, del reducido tamaño muestral de muchos estudios. Estas afirmaciones se ven respaldadas por el resultado que se interpreta seguidamente.

En esta memoria de investigación, el tamaño muestral tampoco se asoció con la calidad al analizar las dos revistas en conjunto ni al estratificar la asociación entre calidad y estado de publicación según la revista. Por tanto, los resultados del análisis bivariante permiten afirmar, con una confianza del 95%, que el tamaño muestral no influye en la publicación o el rechazo de artículos en ninguna de las dos revistas y que dicha ausencia de asociación no responde al efecto de la calidad como posible factor de confusión.

5.6 Objeto del estudio

El primer dato que destaca del análisis de esta variable es la ausencia de artículos sobre análisis o síntesis de políticas de salud en los cuales se hayan utilizado alguno de los diseños contemplados en los criterios de inclusión de artículos. Ello puede ser el reflejo del

campo temático de las revistas. Es probable que en el transcurso de su trayectoria ninguna de ellas haya publicado este tipo de estudios o solamente un número muy bajo de ellos. De ser así, los estudios de esta índole realizados en España deben de publicarse en otras revistas, tanto nacionales como extranjeras.

El resultado más sobresaliente del análisis estratificado de esta variable es la distinta distribución de frecuencias de los artículos según el objeto de estudio de cada revista. En *Atención Primaria*, los objetos de estudios más frecuentes fueron, por orden decreciente de frecuencia, servicios, programas o proyectos y poblaciones, comunidades o personas. Ambos contabilizaron el 88,7% de todos los artículos de la muestra de estudiada de esta revista. Por su parte, en *Medicina Clínica*, los dos objetos de estudio más frecuentes fueron poblaciones, comunidades o personas, y desarrollo, utilización y evaluación de tecnología médica, que en total contabilizaron 95,0% de los artículos revisados de esta revista. Sólo en 3% de los artículos el objeto de estudio fue servicios, programas y proyectos.

Estos resultados traslucen las diferencias referentes al perfil temático de ambas revistas. El de *Atención Primaria* da mucha mayor cabida a investigaciones de servicios sanitarios y a estudios de alcance poblacional y comunitario que *Medicina Clínica*. En la mayor parte de los estudios de *Atención Primaria* sobre poblaciones, comunidades o personas revisados el objeto real fue poblaciones y comunidades, y el enfoque, eminentemente comunitario. Por el contrario, *Medicina Clínica* publica y recibe más artículos de naturaleza clínica. De ahí que en ella la mayor parte de los artículos en los que se estudiaron poblaciones, comunidades y personas el enfoque fuese primordialmente clínico y que esta categoría se nutra sobre todo de estudios de grupos de pacientes o personas más que de comunidades y poblaciones. El enfoque clínico prominente de *Medicina Clínica* también explica que en ella se reciban y publiquen más artículos sobre

utilización y evaluación de tecnología médica que en *Atención Primaria*. Dado el auge reciente a nivel internacional —que también comienza a contemplarse en España, aunque un tanto a la zaga de los países anglosajones— de los estudios de evaluación de tecnología médica (Ganiats TG, Wong AF, 1991; Banta HD, Luce BR, 1993), cabe esperar que la cifra absoluta de los estudios de este tipo que reciban ambas revistas aumente, aunque todo parece indicar que el espacio editorial que *Medicina Clínica* les ha dedicado hasta la fecha es mucho mayor que el de *Atención Primaria*.

No se encontró asociación entre el estado de publicación y el objeto de estudio. Es decir, los artículos publicados y rechazados por las dos revistas no se diferencian en cuanto al objeto de estudio. La única excepción a este respecto apareció en *Medicina Clínica*: esta revista publicó el doble de artículos sobre servicios, programas y proyectos de los que rechazó. Por este motivo, los intervalos de confianza de las proporciones no se solapan. Sin embargo, este resultado ha de interpretarse con precaución, especialmente al extraer conclusiones y hacer extrapolaciones, ya que se cifra en una diferencia de 2 artículos publicados frente a 1 rechazado. Es decir, el tamaño de esta submuestra es tan pequeño que no permite, en rigor, extraer conclusión alguna. Por consiguiente, a efectos prácticos puede concluirse que, en las muestras estudiadas, el objeto de estudio no se asoció con el estado de publicación.

Tampoco se encontró asociación alguna en el análisis combinado ni en el estratificado según la revista entre el objeto de estudio y su calidad. Esto es congruente con la ausencia de asociación entre el estado de publicación y el objeto de estudio, pues sería de esperar que, si la calidad de los artículos con un objeto de estudio determinado fuese mayor, también se publicarían más, ya que la calidad se asoció con el estado de publicación.

Cabe destacar que la variable objeto de estudio no se ha definido ni considerado en estudios similares publicados en revistas nacionales ni internacionales. La ventaja comparativa que aporta su inclusión en esta memoria de investigación es triple. Por un lado, aumenta el contenido informativo del estudio y permite ahondar en el análisis de los perfiles temáticos de las revistas estudiadas. Por otro, si bien se trata de una variable distinta del tipo o diseño del estudio, ayuda a garantizar que esta diferenciación se preserva en la evaluación de los artículos y aumenta la especificidad de la clasificación del diseño de los estudios notificados. Y, por último y como consecuencia de las dos ventajas anteriores, aumenta la reproducibilidad y, por consiguiente, la validez del cuestionario de revisión utilizado en esta memoria de investigación, así como la de las estimaciones de la calidad obtenidas.

Para ilustrar y respaldar estas aseveraciones pueden citarse dos ejemplos. El primero consiste en la evaluación de un programa de prevención en la comunidad mediante un diseño cuasiexperimental pre-post con grupo control. En este caso, el objeto del estudio es la comunidad y el diseño corresponde a una variante de los estudios cuasiexperimentales. En algunas clasificaciones dicho estudio se registraría exclusivamente como evaluación de programa o como estudio evaluativo –de los cuales hay muchas variantes–, a expensas de perder el componente comunitario y la especificidad metodológica que encierra un diseño de este tipo. Otro ejemplo lo aporta la evaluación de una tecnología médica en la cual se realiza un análisis de efectividad en función del coste y un estudio de casos y controles. Este estudio podría clasificarse asimismo como evaluación, como estudio de casos y controles o, en aquellas clasificaciones que incluyen los diseños de evaluación socioeconómicas de programas, como análisis de efectividad en función del coste. Obviamente, todas estas categorizaciones son inespecíficas e incompletas y reducen presumiblemente la reproducibilidad del cuestionario utilizado. Asimismo, no debe

confundirse la complejidad del diseño con la de los análisis estadísticos, que se abordará más adelante.

5.7.1 Tipo de estudio

Tal como se indicó en la sección de Materiales y Métodos de esta memoria de investigación, los resultados relativos a los tipos de estudios incluidos en el presente estudio deben analizarse a tenor de la complejidad de su diseño, el nivel de control de las variables que permite cada uno y su capacidad para sugerir hipótesis y relaciones causales. Así, los estudios descriptivos se sitúan en el extremo inferior de la escala de estos atributos y los ensayos clínicos y otros estudios experimentales, en el extremo superior.

La clasificación de los tipos o diseños de estudios es singular, pues no coincide con las utilizadas en otros estudios (Feinstein AR, 1974; Emerson JD, Colditz GA, 1983; Nolasco A, Gascón E, Mur P, Ferrándiz E, Álvarez-Dardet C, 1986; Álvarez-Dardet C, Mur P, Gascón E, Nolasco, Bolumar F, 1987; Bailar III JC, Mosteller F, 1988; Bailar III JC, Mosteller F, 1992; 1992\$; Silva LC, Pérez-Nieves, Cuellar L, 1995; González JC, Pulido M, Sanz F, 1995; Gómez C, et al., 1996). Entre estas últimas sobresale la de Fletcher (Fletcher RH, Fletcher SW, 1979) por la frecuencia con que se han utilizado versiones en ella, incluso en estudios publicados en revistas médicas españolas (Álvarez-Dardet C, Mur P, Gascón E, Nolasco, Bolumar F, 1987). Por estas razones, la clasificación empleada en esta memoria de investigación presenta tanto ventajas como desventajas. La principal desventaja es que impide realizar comparaciones con los resultados de otros estudios. No obstante, en ninguno de los objetivos de esta memoria se contempló llevar a cabo comparaciones con otros estudios. Por el contrario, el propósito más general consistió en configurar un modelo de análisis de los factores que permitan explicar la publicación y el rechazo de artículos. Las ventajas que ofrece esta clasificación son de diversa índole. Por un

lado, sólo se dirige a clasificar estudios observacionales, cuasiexperimentales y experimentales, lo cual aumenta su especificidad, responde a los objetivos generales y específicos de esta memoria de investigación y la convierte en una clasificación más exhaustiva que las restantes para los tipos generales de estudios consignados. En este sentido supera a las restantes clasificaciones, que no incluyen varios diseños empleados, sobre todo, en evaluación de programas y servicios y en los cada vez más frecuentes análisis socioeconómicos de tecnologías médicas (Luce BR, Elixhauser A, 1990; Ganiats TG, Wong AF, 1991; Drummond M, Brandt A, Luce B, Rovira J, 1993). Por otro lado, es una clasificación sencilla y altamente fiable, ya que la concordancia obtenida con ella en los estudios pilotos sobrepasó 85% (Campillo C, de los Ríos R, Flores R, 1996).

Como cabe esperar, en la muestra estudiada global el número de estudios con diseños menos complejos (descriptivos, comparativos y de prevalencia) superó con mucho al de estudios con diseños de mayor complejidad (163 frente a 33, respectivamente). Este resultado también se constató al estratificar el análisis por revista. En *Atención Primaria*, la relación fue de 81 a 16 y en *Medicina Clínica*, de 82 a 17, ambas prácticamente idénticas. Esto remeda muy de cerca y es muy congruente con las distribuciones de frecuencias de los diseños de estudios notificadas en revistas nacionales --incluida *Medicina Clínica*-- (Álvarez-Dardet C, Mur P, Gascón E, Nolasco , Bolumar F, 1987) e internacionales (Feinstein AR, 1974; Emerson JD, Colditz GA, 1983; ; Bailar III JC, Mosteller F, 1988; Bailar III JC, Mosteller F, 1992; 1992\$; Silva LC, Pérez-Nieves, Cuellar L, 1995).

Por añadidura, el escrutinio más pormenorizado de estos resultados también ha de tomar en consideración el tamaño muestral y el grado de representatividad de la muestra de este estudio. Si la muestra global y la submuestra de cada revista son representativas de la población de artículos de ambas revistas, es preciso preguntarse por qué en las muestras estudiadas aparecen algunos de los estudios con diseños más complejos --como los cuasi-

experimentales— con mayor frecuencia que estudios menos complejos. El pequeño tamaño de varias subcategorías del tipo de estudios respalda la idea de que la variabilidad muestral subsecuente explique dicha distribución muestral. Las ligeras diferencias entre la distribución de frecuencias de los diseños encontradas en este estudio y las observadas en *Medicina Clínica* en estudios efectuados con anterioridad también pueden responder a este motivo (Álvarez-Dardet C, Mur P, Gascón E, Nolasco , Bolumar F, 1987).

Llama la atención que en la muestra no se encontró ningún ensayo clínico. Ello puede deberse a dos factores. Primero, a la variabilidad del muestreo. Segundo, a que, por las características de las normas que rigen en España la realización de estos estudios, la mayor parte de ellos se registren en el *Registro Español de Ensayos Clínicos* del Ministerio de Sanidad y Consumo (Easterbrook PJ, 1992) y se publiquen en un directorio específico de esta institución y no en revistas médicas (Easterbrook PJ, 1992). Asimismo, otras revistas nacionales y extranjeras podrían ser el destino de estos estudios.

Sobresale también el bajo número de estudios de casos y controles y de cohortes recibidos por ambas revistas en el período que abarca el estudio. Ello puede deberse, por un lado, a que se realicen pocos estudios de este tipo o a que los que se lleven a cabo se publiquen en otras revistas. Por otro, también ha de tenerse en cuenta la variabilidad del muestreo, sobre todo a la luz de los resultados de otros estudios realizados con *Medicina Clínica* y otras revistas nacionales en los cuales la frecuencia relativa de este diseño fue más alta, si bien en esos estudios sólo se incluyeron artículos publicados. Este hecho respaldaría en todo caso los resultados obtenidos en la presente memoria.

El tipo de estudio no se asoció con el estado de publicación, ni al analizar conjuntamente los 196 artículos ni al estratificar ambas variables por revista. La única salvedad a este hallazgo se observó en *Atención Primaria*, en la cual se rechazaron más estudios descriptivos de los que se publicaron y se publicaron más de prevalencia de los que

se rechazaron. Esta excepción se explica al constatar que la media de la calidad de los estudios de prevalencia publicados fue más alta que la de los rechazados, al tiempo que la de los estudios descriptivos publicados también fue más elevada que la de los rechazados. Por consiguiente, las diferencias observadas entre esos dos tipos de estudios no son atribuibles a la publicación sesgada de artículos en función atribuible al diseño, sino a que, presumiblemente, la calidad fue el criterio que gobernó la selección de artículos para publicación.

Al considerar los resultados de este apartado junto con los del objeto de los estudios se aprecia que los perfiles de ambas revistas no difieren sustancialmente en cuanto al tipo de estudio sino en relación con su objeto. Estas diferencias contribuyen a explicar los distintos perfiles temáticos de las dos revistas.

~~5.7.2 Tipo de estudio y calidad~~

El análisis bivariante inicial, realizado con la variable tipo de estudio con nueve categorías, reveló la presencia de una asociación débil entre el tipo o diseño del estudio y la calidad. Como ilustran los resultados, esta asociación indica que, a medida que aumenta la complejidad del diseño de un estudio también lo hace su calidad en términos generales. Es decir, es más probable que en ambas revistas la calidad de un estudio descriptivo o comparativo sea más baja que la de un estudio con diseño más intrincado, como, por ejemplo, los diseños experimentales o los cuasiexperimentales.

Como se comentará en el apartado sobre la complejidad de los análisis estadísticos, este resultado es muy congruente con la asociación encontrada entre dicha complejidad y la calidad del estudio. Cabe suponer que la complejidad de los análisis estadísticos corre pareja a la del diseño (Fleiss JL, 1986; Armitage P, Berry G, 1987). Por ello, al igual que se señala al interpretar los resultados de la variable complejidad, es muy verosímil que los

investigadores que llevan a cabo estudios con diseños complejos dispongan de más recursos y sean más avezados en la utilización de métodos de investigación. De ahí se desprende como corolario que la distribución de la calidad sea paralela a la de la complejidad tanto del diseño como del tratamiento estadístico de los datos de un estudio. Cuanto menores son los conocimientos, la experiencia y los recursos para investigación de un autor, menor calidad cabrá esperar de los estudios y actividades de investigación que acometa. Los resultados obtenidos son coherentes con estos principios, tanto los que brindó el análisis conjunto como los obtenidos al estratificar dicha asociación por revista.

La recategorización de la variable tipo de estudio es congruente con la tendencia general que reflejan los resultados: el número de los dos tipos de estudios menos complejos de baja calidad fue más elevado que el de los de calidad intermedia. Por este motivo se combinaron en una sola categoría. El análisis bivariante de estas variables mostró una débil asociación entre ellas. Teniendo en cuenta esta categorización de la variable tipo de estudio, la asociación encontrada puede interpretarse en los siguientes términos: en el análisis combinado, la calidad de los artículos revisados tiende a aumentar con el diseño, el nivel de control de las variables y la capacidad de los estudios para sugerir relaciones causales.

Por otro lado, al redefinir nuevamente las categorías de la variable tipo de estudio, traspasando los estudios de prevalencia al estrato integrado por los descriptivos y los compartivos, la asociación entre el tipo y la calidad desaparece a juzgar los resultados por la estimación puntual de la *RP*. Los valores de su intervalo de confianza del 95% no permiten extraer conclusión definitiva alguna en relación con la asociación que nos ocupa, porque los valores del parámetro estimados son congruentes, con una confianza del 95%, tanto con la inexistencia en la población de dicha asociación como con que ésta se dirija en uno u otro sentido.

Al interpretar este último resultado es menester tener cierta cautela, ya que los estudios de prevalencia por sí solos contabilizaron el 30,61% ($n = 60$) de los 196 artículos evaluados. De ellos, el 73,33% fueron de calidad intermedia y el 26,67% restante, de calidad baja. Por esta razón, su inclusión en la categoría que comprende los descriptivos y los comparativos decanta la *RP* hacia 1,0, es decir, en el sentido de la ausencia de asociación. Por el contrario, al integrarlos en la categoría de los estudios con diseños más complejos, se obtiene el efecto contrario: la *RP* tiende a aumentar por encima de 1,0. Se trata de un efecto similar al de la influencia de los sesgos de clasificación errónea no diferencial (*non differential misclassification bias*, en inglés) en las *RP* (Schlesselman JJ, 1982).

5.8 Complejidad

La clasificación de la complejidad de los análisis estadísticos utilizada en esta memoria de investigación responde a criterios académicos y de aplicabilidad práctica. El criterio académico arraiga en que la mayor parte de los cursos de estadística se estructuran en tres niveles de complejidad creciente: análisis univariantes o de complejidad baja, bivariantes o intermedia, y multivariante o alta. Al mismo tiempo, esta clasificación provee una escala de complejidad fácilmente aplicable en la práctica y exenta de equívocos. El grado de complejidad de un estudio equivale, según esta clasificación, al del análisis o prueba estadística aplicada de mayor complejidad.

Casi la totalidad de los artículos revisados presentaron análisis estadísticos de algún tipo. Ello se debe a los diseños de estudios que se incluyeron en esta memoria de investigación: en los estudios observacionales, experimentales y cuasiexperimentales siempre se llevan a cabo análisis estadísticos (Campbell DT, Stanley JC, 1963; Fleiss JL, 1986; Armitage P, Berry G, 1986). Los artículos que no presentaron alguno de esos diseños

fueron excluidos de la revisión. En estos artículos tampoco se realizaron análisis estadísticos.

Al considerar en conjunto ambas revistas se constató que buena parte de los artículos que éstas reciben presentan análisis de complejidad intermedia, es decir, bivariantes. Este resultado era esperable. En muchos de los estudios revisados se formulaban objetivos y se pretendían contrastar hipótesis por medio de diseños y análisis estadísticos no extremadamente complejos. A tenor de los hallazgos de diversos estudios, todo parece indicar que los análisis realizados con mayor frecuencia en los artículos publicados en revistas de ciencias de la salud son bivariantes, es decir, de complejidad intermedia, y que, si bien la aplicación de métodos estadísticos complejos (o multivariantes) ha aumentado notablemente en la últimas dos décadas (Hosmer DW, Lomeshow S, 1991; Bailar JC III, Mosteler F, 1992; Concato J, Feinstein AR, Holford TR, 1993), su frecuencia no ha superado la de los análisis bivariantes. Estas conclusiones también se desprenden de los resultados obtenidos en estudios de diversas revistas médicas españolas y coinciden con los análisis efectuados en el pasado con *Medicina Clínica* (González JC, Pulido M, Sanz F, 1995).

Desde la perspectiva de la inferencia estadística, los resultados indican que los porcentajes de análisis de complejidad baja exclusivamente y alta no difieren en la población de la cual se extrajeron las muestras. Como se ha mencionado, la frecuencia de artículos en los que se notifican resultados de análisis multivariantes ha aumentado y, probablemente, este también haya sido el caso en las dos revistas estudiadas. Llama la atención, no obstante, que a nivel poblacional dicha frecuencia pueda ser similar a la de los análisis univariantes. De hecho, en total se contabilizaron 40 (20,9%; IC95%: 14,9% - 26,9%) artículos de complejidad baja y 23 (12,9%; IC95%: 7,9% - 17,9%) de complejidad alta. Es decir, en cifras absolutas, los análisis univariantes fueron más frecuentes que los

multivariantes. Empero, los únicos valores del parámetro estimado que se solapan en ambos intervalos de confianza oscilan de 14,9 a 17,9. Por tanto, es probable que en la población de artículos de la que procede la muestra estudiada los análisis de complejidad baja exclusivamente sean más frecuentes que los de complejidad alta, dado que gran parte de los valores del intervalo de confianza de la proporción de los primeros son más altos y no se solapan con los del intervalo de confianza de la proporción de los segundos.

De esos 40 artículos, 32 fueron recibidos en *Atención Primaria* y 8, en *Medicina Clínica*. Este hallazgo puede ser tributario de dos factores. Primero, las diferencias relativas a la cobertura temática de las dos revistas. A juzgar por las muestras estudiadas, *Atención Primaria* recibe más estudios sobre medicina comunitaria y atención primaria que *Medicina Clínica* cuyos objetivos no exigen realizar inferencias. Segundo, las diferencias en el tipo de estudios que reciben las dos revistas: *Atención Primaria* recibe más estudios meramente descriptivos que *Medicina Clínica*. Consiguientemente, era de esperar que en *Atención Primaria* la frecuencia de análisis de complejidad baja fuese más elevada que la de *Medicina Clínica*.

Al investigar la relación entre estado de publicación y complejidad en las dos revistas combinadas se observó que ambas rechazaron más artículos con análisis de complejidad baja y alta y publicaron más con análisis de complejidad intermedia. A primera vista, este hallazgo podría deberse, primero, a que la capacidad investigadora de los autores que envían a publicación artículos de complejidad baja sea menor y, por ende, la calidad final del estudio y de la presentación de los datos en el artículo. Por ello se rechazaron más. Y, segundo, a que aquellos que realizan análisis más complejos no alcancen el nivel de formación en métodos de investigación que exige el manejo de estos análisis, lo cual explicaría su menor calidad y, por tanto, su mayor rechazo. A esto último habría que añadir el hecho de que la extraordinaria difusión de paquetes de análisis

estadísticos no ha ido a la par del aumento paralelo necesario de la formación de sus usuarios en métodos de investigación (Concato J, Feinstein AR, Holdford TR, 1993; González JC, Pulido M, Sanz F, 1995; Campillo C, 1996). Por tanto, aunque el número de artículos que presentan resultados de modelos multivariantes haya aumentado, es posible que el rigor metodológico de los estudios que los emplean no alcance todavía las cotas de calidad de los artículos que utilizan análisis de complejidad intermedia. Estas explicaciones suponen, implícitamente, que los revisores de ambas revistas evalúan correctamente los modelos multivariantes, aunque su presentación podría mejorar sustancialmente como se ha mencionado en otros estudios realizados con revistas españolas (en algunos de los cuales se incluyó *Medicina Clínica*) y extranjeras (Hosmer DW, Lomeshow S, 1991; Bailar JC III, Mosteler F, 1992; Concato J, Feinstein AR, Holford TR, 1993; Palmas W, Denton TA, Diamond GA, 1993; Campillo C, 1993; Goodman SN, 1993; González JC, Pulido M, Sanz F, 1995).

Sin embargo, en el análisis estratificado por revista no se detectó asociación alguna entre el estado de publicación y la complejidad de los análisis. Tampoco se detectó ninguna asociación entre la calidad y la complejidad, ni en el análisis combinado ni en el estratificado. Por consiguiente, los datos obtenidos son compatibles con la ausencia de asociación entre complejidad y estado de publicación de los artículos evaluados en ambas revistas y descartan el posible efecto confusor de la calidad en dicha asociación.

5.9 Significación estadística

El primer resultado destacable que brindó el análisis de esta variable es que, de 196 artículos revisados, en 146 se realizaron pruebas estadísticas y que todas ellas fueron pruebas de hipótesis. Parafraseando, en ningún artículo en el que se presentaron resultados

de análisis estadísticos se efectuaron mayoritariamente pruebas de inferencia estadística. Éstas fueron pruebas puntuales.

La importancia de este hallazgo estriba en lo siguiente. A pesar de las discrepancias y debates que desde hace tiempo se mantienen en torno a las ventajas y desventajas de las pruebas de inferencia estadística frente a las de hipótesis, en ningún artículo revisado --sin excepción alguna-- se llevaron a cabo exclusivamente inferencias.

En esta discusión no se profundizará en ese debate, pues ello excede con creces los objetivos de esta memoria de investigación. Valga señalar, sin embargo, que la mayor parte de las revistas indizadas en bases bibliográficas de prestigio internacional (como *Index Medicus*, *Current Contents* o *Excerpta Medica*, por citar algunos ejemplos) siguen, con mayor o menor adhesión y rigor, buena parte de las normas y recomendaciones dictadas por los comités internacionales de editores de revistas de ciencias de la salud (Altman DG, Gore SM, Gardner MJ, Pocock SJ, 1983; International Committee of Medical Journal Editors, 1993). Todas estas normas coinciden en aconsejar utilizar pruebas de inferencia en contraposición a las de hipótesis al referirse a la selección, descripción y presentación de análisis estadísticos en los artículos científicos. A pesar de este consenso --alcanzado hace ya algunos años--, ni la política editorial de *Atención Primaria* ni la de *Medicina Clínica* parecen seguir de cerca dichas recomendaciones. Ello se desprende de los resultados de la revisión de los artículos publicados analizados en esta memoria. Además, los artículos rechazados que se evaluaron también presentaron exclusivamente pruebas de hipótesis. Por consiguiente, los autores tampoco observan las recomendaciones de los comités internacionales. Esto es presuntamente atribuible, por un lado, a que no lo exigen las revistas. Y, por otro, a que en la formación académica en análisis estadístico todavía prevalece o ha prevalecido durante muchos años el enfoque de las pruebas de hipótesis; baste con revisar los libros y materiales de formación utilizados desde los años setenta.

Expresado en términos epidemiológicos, los resultados obtenidos traducen un efecto de cohorte: los investigadores que en la actualidad envían artículos a las revistas recibieron formación en métodos de investigación en el pasado.

En conclusión, la falta de observancia puesta de manifiesto es una responsabilidad compartida tanto por editores, revisores como por autores, aunque ello no es óbice para afirmar que, a la postre, la observancia de las normas internacionales es una función que deben desempeñar los editores y por la cual han de velar los revisores de las revistas (Feinstein AR, 1983; Silva G, Campillo C, 1991; CBE Peer Review Retreat Consensus Group, 1995).

El segundo dato que sobresale –valga decir de forma abrumadora– es que de los 146 artículos en los que se aplicaron pruebas estadísticas, 143, es decir, el 97,9%, presentaron resultados estadísticamente significativos, es decir, fueron estudios positivos según los términos empleados en la investigación de sesgos de publicación (Dickersin K, 1990). Sólo 3, por tanto, fueron negativos. Por añadidura, de esos 146 artículos, 79 se publicaron y 67 se rechazaron. De los 3 estudios negativos, 2 fueron publicados por *Medicina Clínica* y 1, rechazado por *Atención Primaria*. De este resultado diáfano y concluyente se desprenden las siguientes reflexiones.

Primero, aunque el tamaño de la muestra de esta memoria de investigación pudo haber sido mayor, permite estimar claramente los elementos que vertebran el sesgo de publicación en las dos muestras submuestras estudiadas.

Segundo, sin lugar a dudas, la validez externa de cualquier conclusión extraída de los datos analizados es tributaria en gran medida de la representatividad de la muestra estudiada en esta memoria de investigación.

Tercero, suponiendo que su representatividad es aceptable, se puede afirmar que, obviamente, los editores de las dos revistas no han introducido sesgo de publicación en la

muestra estudiada, habida cuenta de que el 97,9% de los artículos publicados y rechazados presentaron resultados estadísticamente significativos.

Cuarto, sobre la base de estos resultados también es lícito afirmar que, claramente, los revisores no han introducido sesgo de publicación. El hecho de que casi todos los estudios fuesen positivos impide cualquier posibilidad de introducir este sesgo.

Quinto, los resultados indican, por consiguiente, que, si en cualquiera de las dos revistas estudiadas existe realmente sesgo de publicación, ese sesgo fue introducido por los autores. Es decir, de acuerdo con la definición de este sesgo²⁰, si existe sesgo de publicación, los autores son los responsables de introducirlo, al decidir enviar a publicación o no un estudio en función del nivel de significación estadístico alcanzado. En el caso concreto de este estudio, si existiese el sesgo, los autores sólo envían artículos a publicación, o lo hacen mayormente, cuando los resultados son estadísticamente significativos.

En este estudio no se formuló como objetivo confirmar, de sugerirlo los resultados, la existencia de este sesgo como consecuencia del problema del archivero (Rosenthal R, 1979). Ello obligaría a realizar otro tipo de investigación que remedara las efectuadas para investigar dicho problema, tales como la realización de entrevistas a los autores.

En cualquier caso, los resultados obtenidos son totalmente coherentes y contribuyen a confirmar una de las hipótesis barajadas en el campo de investigación de los sesgos de publicación (Dickersin K, 1993): de existir este sesgo de publicación, los autores, y no los editores, son responsables de su introducción. Obviamente, ellos, los revisores, los editores y la comunidad científica comparten la responsabilidad de las consecuencias de este sesgo en la investigación y en la práctica de las ciencias de la salud.

²⁰

Este sesgo se definió en la Introducción como la tendencia de autores, revisores y editores a enviar a publicación, recomendar la aceptación y aceptar para publicación, respectivamente, los estudios sobre la base de la magnitud y dirección de los resultados, entendidas desde el prisma de la significación estadística.

5.10 Análisis multivariante

Como se desprende de los resultados de los análisis multivariantes, la calidad –tal como se definió al inicio de esta memoria– fue la única variable que se asoció con el estado de publicación en *Atención Primaria y Medicina Clínica*. Esta asociación fue más fuerte en los artículos recibidos para publicación en la segunda revista que en la primera. Por tanto, en la muestra analizada, el nivel de calidad de los artículos se discierne tanto en *Medicina Clínica* como en *Atención Primaria*.

La media de la calidad de los artículos publicados fue mayor que la de los rechazados tanto en una como en otra revista. Por ello, cabe esperar que la fuerza de la asociación estimada para las dos revistas debiera haber sido mucho más similar que las estimadas en este caso. Ahora bien, los intervalos de confianza de ambas estimaciones puntuales de las *RP* comparten bastantes valores, por lo cual no es posible descartar, desde un punto de vista inferencial, que sean similares o idénticas. No hay que olvidar el grado de variabilidad de las estimaciones inherente al tamaño muestral alcanzado en los análisis de subgrupos. Consiguientemente, los datos observados son, en definitiva, congruentes con la posibilidad de que en ambas revistas se distingan con la misma efectividad los artículos en virtud de su calidad y, por ende, que se seleccionen para publicación en función del mismo criterio o atributo y con la misma eficiencia.

Este resultado cardinal del análisis multivariante –y de esta memoria de investigación– es además coherente con los resultados arrojados por el análisis bivariante. La única variable claramente asociada con el estado de publicación en este último tipo de análisis en ambas revistas fue la calidad de los artículos, lo cual confiere mayor rigor y solidez al hallazgo.

A este resultado sólo cabe añadir una salvedad. Los datos observados en *Atención Primaria* también son compatibles con la existencia de una asociación entre el estado de publicación y el tipo o diseño del estudio notificado en el artículo. La fuerza de esta asociación fue menor que la estimada entre el estado de publicación y la calidad en las dos revistas. No obstante, como indica el intervalo de confianza, los valores del parámetro *RP* compatibles con los datos observados con una confianza del 95% son congruentes con la presencia de una asociación fuerte: la que corresponde a los valores mayores de 10,0. Este resultado traduce los mismos datos obtenidos en el análisis bivariante: la asociación entre tipo de estudio y estado de publicación en esa revista es atribuible a los estudios de prevalencia y descriptivos, y es independiente de la calidad, pues la *RP* está ajustada.

Otro comentario hace referencia a la variabilidad de las estimaciones. Como ha ocurrido en otros análisis de subgrupos, la precisión de las estimaciones de las *RP* derivadas de los análisis multivariantes no son menospreciables, especialmente la de la *RP* correspondiente a *Medicina Clínica*, que estuvo comprendida entre 7,14 y 125,0. Como se comentará en el apartado de limitaciones del estudio, estos intervalos de confianza amplios son atribuibles por entero al bajo tamaño de las submuestras en que se fundamentan los análisis de subgrupos y estratificados y, por tanto, la solución postrera a este problema consistiría en aumentar el tamaño de las muestras (Donner A, 1979; Donner A, 1982). El bajo tamaño muestral es, por otra parte, el contrapunto o el precio que debe pagarse por la alta especificidad alcanzada con las categorizaciones de diversas variables estudiadas en esta memoria. Cuanto mayor es el número de categorías de las variables, mayor especificidad se alcanza en las clasificaciones y más información se obtiene, pero a expensas de reducir la precisión de las estimaciones. Lo mismo ocurre cuando en un estudio se intentan detectar interacciones entre variables con varias categorías (Schlesselman JJ, 1982; Kahn HA, Sempos CT, 1989).

Esta limitación que dimana del tamaño muestral es minimizada por la congruencia de los resultados de los distintos análisis (mentada más arriba al aludir a los resultados del análisis bivalente) y por la validez de los modelos multivariantes finales. Así lo indican las pruebas de bondad del ajuste realizadas de los dos modelos.

Como complemento a la interpretación de los análisis multivariantes es menester indicar que con el diseño de casos y controles no es posible ni correcto, desde el punto de vista del diseño y el análisis estadístico, calcular la probabilidad de que se publique o rechace un artículo a partir de los valores de las variables independientes consideradas en esta memoria de investigación (Breslow NE, Day NE, 1980; Schlesselman JJ, 1982; Kahn HA, Sempos CT, 1989). Para ello sería necesario emplear un diseño de naturaleza prospectiva, como el de cohortes. La variante concurrente de este diseño presenta como inconveniente el dilatado período que debería esperarse a fin de disponer de datos suficientes para llevar a cabo el estudio, al igual que ocurre con las enfermedades con largos períodos de latencia (Breslow NE, Day NE, 1987; Kahn HA, Sempos CT, 1989). No obstante, su variante no concurrente permitiría analizar la validez de los modelos multivariantes que se construyeran, muy similares a los presentados en el presente estudio. Esto es así porque con dicho estudio sería posible utilizar la regresión logística como modelo predictivo. Es decir, este modelo lineal generalizado permitiría calcular las probabilidades mencionadas anteriormente y contrastarlas con observaciones reales en períodos de tiempo prudenciales. De ahí la insistencia en disuadir el cálculo de dichas probabilidades con el diseño retrospectivo empleado en esta memoria de investigación. En cualquier caso, el cálculo de dichas probabilidades no fue el objetivo de este estudio.

5. 11 Limitaciones del estudio

Las limitaciones de la presente memoria de investigación hacen referencia a la validez interna y externa del estudio, y pueden resumirse en tres apartados:

representatividad de la muestra, valoración de la calidad metodológica de los artículos y tamaño muestral.

Las muestras procedentes de las dos revistas fueron consecutivas. Los responsables de *Atención Primaria* y de *Medicina Clínica* señalaron que durante el período en que se extrajeron ambas muestras y en los años inmediatamente precedentes a este estudio no se había introducido cambio alguno en las políticas editoriales de revisión y selección de manuscritos. Asimismo, estas personas afirmaron que en los artículos recibidos durante dicho período no se habían observado cambios ni tendencias en las variables consideradas en el diseño de esta memoria de investigación. Además, los artículos evaluados constituyen un porcentaje muy elevado de los recibidos por las dos revistas durante el lapso que cubre este estudio. Por consiguiente, y a pesar de que en rigor la muestra estudiada no puede considerarse aleatoria, todos los datos disponibles son congruentes con que es representativa de la población de artículos recibidos por ambas revistas en los años que anteceden a la realización de este estudio. A juzgar por los métodos de muestreo empleados en las dos revistas y por las afirmaciones de sus responsables, la muestra obtenida no parece haberse configurado como resultado de los efectos ejercidos por sesgos de selección.

La ausencia de cambios en las políticas editoriales de revisión y selección de manuscritos es un garante de que los resultados obtenidos son atribuibles a las políticas editoriales vigentes en el período mencionado y no a los efectos que podrían ejercer modificaciones de estas políticas que se hubiesen introducido durante el período durante el cual se conformaron las muestras. Ello también se contrapone a la posibilidad de que se hayan introducido sesgos de selección en el muestreo tributarios de esos cambios. Por todo ello, se concluye que los resultados pueden extrapolarse al período a que pertenecen los artículos analizados de ambas revistas.

La segunda limitación hace referencia a la valoración de la calidad científica de los artículos incluidos en el estudio. Como se ha mencionado en varios pasajes de esta memoria, la calidad es un concepto intrincado y multidimensional. Para hacer frente a este escollo, se adoptó una definición de calidad consensuada por tres evaluadores y parsimoniosa. Asimismo, para medirla se construyó una escala sencilla, lineal y no ponderada, que fue validada en dos estudios. En el primero, esos tres evaluadores elaboraron y validaron el cuestionario (Campillo C, de los Ríos R, Flores R, 1996). Y en el segundo, se comprobó la elevada confiabilidad de la escala construida para medir la calidad.

Por otra parte, la aceptación del método de valoración de la calidad utilizado y del cuestionario empleado para obtener los datos de este estudio se basa en cuatro hechos. Primero, en la validez de contenido, de constructo y de criterio del cuestionario y la escala utilizados para medir la calidad. Segundo, en la reproducibilidad o concordancia interobservador del cuestionario (la *kappa* osciló entre 45 y 60%) (Campillo C, de los Ríos R, Flores R, 1996). La ausencia de datos publicados sobre otros cuestionarios validados destinados a los mismos propósitos que los de este estudio obligó por otra parte a construir este instrumento de medición de la calidad. Aún más, la elaboración y aplicación de este instrumento de medida se contempla como una aportación innovadora y útil en este tipo de estudios, así como uno de los principales pilares en que descansa la originalidad de esta memoria de investigación. Los resultados en que se basa la aceptación del cuestionario son muy similares a los notificados por los responsables de la revisión de manuscritos en la revista *Annals of Internal Medicine* (Goodman S, comunicación personal 1993), la única revista prestigiosa de la que se dispone de datos --aunque no publicados-- sobre este particular. Tercero, en la alta confiabilidad de la escala de calidad utilizada, tal como lo indican los resultados de los análisis de componentes principales y factoriales y el

coeficiente *theta* de Armor ($theta = 0,969$) (Carmines EG, Zeller RA, 1981). Por último es preciso destacar que la evaluación de los artículos y la aplicación de la escala de calidad se efectuaron a ciegas del estado de publicación, la procedencia y la revista, lo cual reduce notablemente la posibilidad de que en su aplicación hayan podido intervenir sesgos de valoración.

La tercera limitación es la relativa a las restricciones que el tamaño de la muestra estudiada impone en los análisis y a sus consecuencias en la validez interna y externa del estudio. De acuerdo con los objetivos e hipótesis de esta memoria, se pretendía, en síntesis, conocer los perfiles editoriales de las dos revistas (en lo que atañe exclusivamente a los artículos originales que reciben en los cuales se notifican estudios observacionales, cuasiexperimentales y experimentales), así como a averiguar qué factores se asocian con el estado de publicación y a detectar sesgos de publicación y de procedencia. Por lo tanto, el presente estudio sólo intenta describir perfiles editoriales y aportar un modelo general que permita detectar asociaciones entre el estado de publicación y diversas variables y sesgos de publicación. No fue su propósito estimar la magnitud de estos sesgos, lo cual redujo muy notablemente el tamaño de la muestra que se precisó para llevarlo a cabo. Por estos motivos, los tamaños de las muestras estudiadas se consideraron suficientes para alcanzar los objetivos propuestos y contrastar las hipótesis formuladas.

A pesar de ello, el tamaño muestral se redujo considerablemente al estratificar el estudio de algunas asociaciones según la revista y algunas variables de interés. Aunque dichos análisis no permiten extraer conclusiones definitivas, han permitido detectar las asociaciones de mayor peso específico en el estudio, a expensas, no obstante, de reducir la precisión de algunas estimaciones. Si se acepta la representatividad de las muestras, los resultados obtenidos, aunque no ofrecen estimaciones muy precisas en todos los casos,

conforman un modelo que permite responder en la práctica a los objetivos marcados en este estudio y contrastar todas las hipótesis formuladas.

CAPÍTULO 6. CONCLUSIONES

Teniendo en cuenta los resultados obtenidos en esta memoria de investigación, se formulan las siguientes conclusiones.

1. La calidad global de los artículos evaluados osciló entre intermedia y baja, y la de las seis secciones de los artículos publicados fue mayor que la de los rechazados.
2. La Introducción fue la sección de calidad más alta de los artículos recibidos por ambas revistas. El apartado de menor calidad de esta sección fue la formulación de objetivos e hipótesis.
3. En *Atención Primaria* los Resultados ocuparon el segundo lugar en calidad, seguidos por Materiales y Métodos y Discusión. En *Medicina Clínica*, el rango de calidad fue: Materiales y Métodos, Discusión y Resultados.
4. En Materiales y Métodos los apartados de menor calidad fueron la descripción de la recolección, control de calidad y análisis de los datos, del diseño utilizado, de los métodos de análisis y del muestreo. La idoneidad del diseño utilizado para los objetivos e hipótesis del estudio fue el apartado de mayor calidad.
5. En Resultados la construcción y corrección de los modelos de análisis fue el apartado de calidad más alta.

-
6. Las principales deficiencias de la Discusión fueron las conclusiones, que trasgreden los límites de extrapolación, la sobrestimación de la validez interna y externa del estudio y la notificación de sus limitaciones.
 7. La calidad de la Presentación fue intermedia en las dos revistas.
 8. *Atención Primaria* recibió más artículos de Cataluña que de otras regiones que *Medicina Clínica*, pero la procedencia no se asoció con su publicación o rechazo en ninguna de las dos revistas.
 9. En *Medicina Clínica* la calidad de los artículos de Cataluña fue menor que la de los de otra región.
 10. Es posible que *Medicina Clínica* rechace más artículos de Cataluña, no porque introduzca un sesgo de región de procedencia, sino porque la calidad de los artículos de Cataluña es en conjunto menor.
 11. Ni *Atención Primaria* ni *Medicina Clínica* introducen sesgo de procedencia en la selección de los artículos que reciben.
 12. En casi todos los artículos recibidos se estudiaron muestras. El tamaño muestral no se asoció con el estado de publicación ni con la calidad.
 13. El perfil temático de *Atención Primaria* da mayor cabida a investigaciones de servicios sanitarios y a estudios de alcance poblacional y comunitario que

Medicina Clínica. *Medicina Clínica* publica y recibe más artículos de naturaleza clínica.

14. El objeto de estudio no se asoció con el estado de publicación ni con la calidad.
15. El tipo de estudio no se asoció con el estado de publicación, exceptuando a *Atención Primaria*, en la cual se rechazaron más estudios descriptivos de los que se publicaron y se publicaron más de prevalencia de los que se rechazaron.
16. Los perfiles de ambas revistas no difieren sustancialmente en cuanto al tipo de estudio sino en relación con su objeto.
17. Sólo en *Atención Primaria* se detectó una asociación (débil) entre el tipo de estudio y la calidad, atribuible a los estudios de prevalencia.
18. En casi todos los artículos recibidos por las dos revistas se realizaron análisis estadísticos. Los de complejidad intermedia fueron los más frecuentes, seguidos por los univariantes y los multivariantes.
19. Ni el estado de publicación ni la calidad se asociaron con la complejidad en ninguna de las dos revistas.
20. En ningún artículo en el que se presentaron análisis estadísticos las pruebas de inferencia estadística fueron prominentes.

21. Las políticas editoriales de las dos revistas no siguen de cerca las recomendaciones internacionales de presentación del diseño y de los datos estadísticos.
22. Casi todos los artículos recibidos notificaron estudios positivos.
23. Los editores y revisores de las dos revistas no introdujeron sesgo de publicación en la muestra estudiada. Si en cualquiera de las dos revistas estudiadas existe sesgo de publicación, éste fue introducido por los autores, lo cual confirma la principal hipótesis formulada en la investigación de los sesgos de publicación.
24. A nivel multivariante, la calidad fue la única variable que se asoció con el estado de publicación en las dos revistas. Por consiguiente, sus sistemas de revisión parecen ser efectivos a este respecto, ya que la calidad rige la selección de artículos.
25. Los resultados obtenidos en *Atención Primaria* también son compatibles con la existencia de una asociación entre el estado de publicación y el tipo de estudio, lo cual es atribuible a los estudios de prevalencia y descriptivos.
26. El modelo de análisis presentado en este estudio es útil para monitorizar la revisión y selección editoriales de manuscritos, así como para detectar si en ellas intervienen sesgos de publicación.

CAPÍTULO 7. REFERENCIAS

Agency for Health Care Policy and Research, US Department of Health and Human Services. Publication bias examined. *Research Activities* 1990;(130):5-6.

Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J* 1983;286:1489-1493.

Álvarez-Dardet C, Mur P, Gascón E, Nolasco A, Bolumar F. La investigación clínica en España: diseños utilizados. *Med Clin (Barc)* 1987;89:221-223.

American Journal of Epidemiology. *Suggestions for referees*. Baltimore, MD: The Johns Hopkins University, School of Hygiene and Public Health; 1992. (Documento no publicado.)

American Journal of Epidemiology. Comments and summary of referee. Baltimore, MD: The Johns Hopkins University School of Hygiene and Public Health; 1992. (Documento no publicado.)

Angell M. Negative studies. *N Engl J Med* 1989;321:464-466.

Angell M. [letter]. *N Engl J Med* 1990;322:1085.

Annals of Internal Medicine. Readers questionnaire. Philadelphia: *Ann Intern Med*; 1992. (Documento no publicado.)

Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell Scientific Publications; 1986.

Atkinson DR, Furlong MJ, Wampold BE. Statistical significance, reviewer evaluations, and the scientific process: is there a (statistically) significant relationship? *J Consulting Psychol* 1982;29:189-194.

Audet N, Gagnon R, Ladouceur R, Marcil M. L'enseignement de l'analyse critique des publications scientifiques médicales est-il efficace? Revision des études et de leur qualité méthodologique. *Can Med Assoc J* 1993;148:945-952.

Bacan D. The test of significance in psychological research. *Psychol Bull* 1966;66:432-437.

Bailar III JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals: amplifications and explanations. *Ann Intern Med* 1988;108:317-332.

Bailar III JC, Mosteller F. *Medical uses of statistics*. 2nd ed. Boston: NEJM Books; 1992.

Banta HD, Luce BR. *Health care technology assessment. An international perspective*. Oxford: Oxford University Press; 1993.

-
- Barker DJP. *Practical epidemiology*. 2nd. ed. London: Churchill Livingstone: 1976.
- Beeler MF, Sappenfield RW. *How to analyze clinical research reports*. Chicago: American Society of Clinical Pathologists Press; 1986.
- Begg CB. A measure to aid in the interpretation of published clinical trials. *Statistics Med* 1985; 4:1-9.
- Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J Am Stat Assoc* 1988;151:419-463.
- Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989;81:107-114.
- Begg CB. Publication bias. In: Cooper H, Hedges L, eds. *Handbook of research synthesis*. New York: Sage Publications; 1994:399-409.
- Benn-Shlomo Y. "Place of publication" bias [Letter]. *Br Med J* 1994;309:274.
- Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Une Stat Assoc* 1989;84:381-392.
- Berlin JA. Will publication bias vanish in an age of online journals. *Online J Curr Clin Trials* 1992;doc No. 12.
- Bero L, Rennie D. Publication bias and public policy. *Presentado en Second International Congress on Peer Review in Biomedical Publication, Chicago, September 9-11, 1993*. Chicago: American Medical Association; 1993:17.
- Bettinger C. [Letter]. *JAMA* 1992;267:2891.
- Black D. *Lancet* 1958;1.
- Black D. Bias in reporting. *Lancet* 1993;336:1587.
- Breslow NE, Day NE. *Statistical methods in cancer research*. Vol. I: the analysis of case-control studies. Lyon: IARC/WHO; 1980.
- Breslow NE, Day NE. *Statistical methods in cancer research*. Vol. II: the design and analysis of cohort studies. Lyon: IARC/WHO; 1987.
- Briggs A, Sculpher M. Sensitivity analysis in economic evaluation: a review of published studies. *Health Economics* 1995;4:355-371.
- Burris JF. Publication bias [Letter]. *Clin Pharmacol Ther* 1993;53:495.
- Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company; 1963.

Campillo C, Alexander D, Ibrahim I. A case control study of publication status of articles submitted to *Boletín de la Oficina Sanitaria Panamericana*. Baltimore: Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University; 1991. (Estudio no publicado.)

Campillo C. Standardizing criteria for logistic regression models. *Ann Intern Med* 1993;119:540-541.

Campillo C. Interpretaciones erróneas de los valores de p. *Atención Primaria* (Barc) 1996;17:221-224.

Campillo C, de los Ríos R, Flores R. *The methodological quality of medical research financed by PAHO/WHO. 1981-1994*. Washington, DC: Pan American Health Organization/World Health Organization; 1996.

Carmines EG, Zeller RA. *Reliability and validity assessment*. Beverly Hills: Sage Publications; 1981

CBE Peer Review Retreat Consensus Group. Peer review guidelines: a working draft. *CBE News* (Chicago) 1995;18:79-81.

Chalmers I. Proposal to outlaw the term "negative trial". *Br Med J* 1985;290:1002.

Chalmers I, Hetherington J, Newdick M, Mutch L, Grant A, Enkin M, et al. The Oxford Database of Perinatal Trials: developing a register of published reports of controlled trials. *Controlled Clin Trials* 1986;7:306-324.

Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. *JAMA* 1990;263:1401-1404.

Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990;263:1405-1408.

Chalmers I, et al. Getting to grips with Archie Cochrane's agenda. *Br Med J* 1992;305:786-788.

Chalmers I, Collins RE, Dickersin K. Controlled trials and meta-analysis can help resolve disagreements among orthopaedic surgeons. *J Bone Joint Surgery* 1992;74-B:641-643.

Chalmers I. Publication bias [Letter]. *Lancet* 1993;342:1116.

Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I: control of bias and comparison with large co-operative trials. *Stat Med* 1987;6:315-325.

Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. II: replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987;6:733-744.

Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990;263:1392-1395.

Chan SS, Sacks HS, Chalmers TC. The epidemiology of unpublished randomized control trials [abstract]. *Clin Res* 1982;30:234A.

Charny M. [Letter]. *Lancet* 1991;337:1102.

Cholesterol Treatment Trialist' (CTT) Collaboration. Protocol for a prospective collaborative overview of all current and planned randomized trials of cholesterol treatment regimens. *Am J Cardiol* 1995;75:1130-1134.

Cicchetti DV, Conn HO. A statistical analysis of reviewer agreement and bias in evaluating medical abstracts. *Yale J Biol Med* 1976;49:373-383.

Clarke MJ, Stewart LA. Obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analyses? *Br Med J* 1994;309:1007-1010.

Cohen LH. Clinical psychologists's judgements of the scientific merit and clinical relevance of psychotherapy outcome research. *J Consulting Clin Psychol* 1979;47:421-423.

Collins R, Gray R, Godwin J, Peto R. Avoidance of large biases and large random errors in the assessment of moderate treatment effects: the need for systematic overviews. *Stat Med* 1987;6:245-250.

Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariate models. *Ann Intern Med* 1993;118:201-210.

Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, et al. Should nonpublished data be included in meta-analysis? Current convictions and controversies. *JAMA* 1993;269:2749-2753.

Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Postdam consultation on meta-analysis. 1995;48:167-171.

Cooper H, Hedges LV, eds. *The handbook of research synthesis*. New York: Russell Sage; 1994.

Coursol A, Wagner EE. Effect of positive findings on submission and acceptance rates: a note on meta-analysis bias. *Professional Psychol* 1986;17:137.

Daniel HD. An evaluation of the peer review process at *Angewandte Chemie*. *Eur Sci Editing* 1993;50:4-8.

Davidson RA. Source of funding and outcome of clinical trials. *J Gen Internal Med* 1986;1:155-158.

Day R. *Cómo escribir y publicar trabajos científicos*. Washington, DC: Organización Panamericana de la Salud; 1995.

Dear KGB, Begg CB. An approach for assessing publication bias prior to performing a meta-analysis. *Statistical Sci* 1992;7:237-245.

de Melker HE, Rosendaal FR, Vandenbroucke JP. Is publication bias a medical problem? [Letter]. *Lancet* 1993;342:621.

DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *New Engl J Med* 1982;306:1332-1337.

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;7:177-188.

Detsky AS, Sackett DL. When was a "negative" clinical trial big enough? *Arch Intern Med* 1985;145:709-712.

Dever GEA. *Epidemiology in health services management*. Gaithersburg, MD: Aspen Publishers; 1984.

Dickersin K, Hewitt P, Mutch, Chalmers I, Chalmers TC. Perusing the literature: comparison of MEDLINE searching with a perinatal trials database. *Controlled Clin Trials* 1985;6:306-317.

Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith Jr H. Publication bias and clinical trials. *Controlled Clin Trials* 1987;8:343-353.

Dickersin K. Reference bias in reports of clinical trials. *Br Med J* 1987;295:1066-1067.

Dickersin K. Report from the Panel on The Case for Registers on Clinical Trials at the Eighth Annual Meeting fo the Society for Clinical Trials. *Controlled Clin Trials* 1988;9:76-81.

Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385-1390.

Dickersin K. Confusion about "negative" studies [letter]. *N Engl J Med* 1990;322:1084.

Dickersin K, Meinert CL. Risk factors for publication bias: results of a follow-up study [Abstract]. *Controlled Clin Trials* 1990;11:255.

Dickersin K, Min YI, Meinert CL. Follow-up trials funded by NIH in 1979: evidence of publication bias. *Controlled Clin Trials* 1991;12:634.

Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267:374-378.

Dickersin K, Meinert CL, Min YI. [Letter]. *JAMA* 1992;267:2891-2892.

-
- Dickersin K. Why register clinical trials?--Revisited. *Controlled Clin Trials* 1992;13:170-177.
- Dickersin K, Min YI. Publication bias: the problem that won't go away. In: Warren KS, Mosteller F, eds. *Doing more good than harm. The evaluation of health care interventions.* *Ann NY Acad Sci* 1993;703:135-148.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *Br Med J* 1994;309:1286-1291.
- Donner A. The design of a clinical trial with several patients categories. *Can J Stat* 1979;7:169.
- Donner A. A bayesian approach to subgroup analysis. *J Chron Dis* 1982;35:429-435.
- Drummond MF, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programs.* Oxford: Oxford University Press; 1987.
- Drummond M, Brandt A, Luce B, Rovira J. Standardizing methodologies for economic evaluation in health care. *Int J Technol Assess Health Care* 1993;9:26-36.
- Dudley HAF. Surgical research: master or servant. *Am J Surg* 1978;135:458-460.
- Earleywine M. The file drawer problem in the meta-analysis of subjective responses to alcohol [Letter]. *Am J Psychiatry* 1993;150:1435-1436.
- Easterbrook PJ. Reducing publication bias. *Br Med J* 1987;295:1347.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-872.
- Easterbrook PJ. Directory of registries of clinical trials. *Stat Med* 1992;11:345-423.
- Eaton CB. [Reply to the letter by Katerndahl DA]. *J Am Board Fam Practice* 1992;5:450.
- Ector H, Aubert A, Stroobandt R. Review of the reviewer. *PACE* 1995;18:1215-1217.
- Emerson JD, Colditz GA. Use of statistical analysis in the *New Engl J Med.* *New Engl J Med* 1983;309:709-713.
- Emerson JD, McPeck B, Mosteller F. Reporting clinical trials in general surgery journal. *Surgery* 1984;95:572-579.
- Epidemiology Program Office, Centers for Disease Control, World Health Organization, Global Program on AIDS. *Epi Info, Version 5.00.* Atlanta, GA; Geneva: CDC, WHO; 1990.

Erill S. La ética de la publicación: el caso de los estudios negativos. *Med Clin (Barc)* 1992;98:308-309.

Ernst E, Resch KL. Reviewer bias: a blinded experimental study. *J Lab Clin Med* 1994;124:178-182.

Feinstein AR. Clinical biostatistics: a survey of the statistical procedures in general medical journals. *Clin Pharmacol Ther* 1974;15:97-107.

Feinstein AR. Prophylactic and remedial therapy for the intellectual ailments of <<biostatistics>>. *Pediatrics* 1983;72:131-133.

Feinstein AR. *Clinimetrics*. New Haven: Yale University Press; 1987.

Fleiss JL. *The design and analysis of clinical experiments*. New York: John Wiley and Sons, INC; 1986.

Fleming TR. Historical controls, data bank, and randomized control trial in clinical research: a review. *Cancer Treatment Rep* 1982;66:1101-1105.

Fletcher RH, Fletcher SW. Clinical research in general medical journals: a 30 years perspective. *N Engl J Med* 1979;301:180-183.

Freiman JA, Chalmers TC, Smith R, Keubler RR. The importance of β , the tipe II error and the sample size in the design and interpretation of the randomized control trial: a survey of '71 negative' trials. *New Engl J Med* 1978;299:690-694.

Ganiats TG, Wong AF. Evaluation of cost-effectiveness research: a survey of recent publications. *Family Med* 1991;23:457-462.

Gardner MJ, Altman DG. Confidence intervals rather than P values. *Br Med J* 1986;292:746-750.

Gardner MJ, Machin D, Campbell MJ. Use of checklist for assessing the statistical content of medical studies. *Br Med J* 1986;292:810-812.

Garfunkel JM, Hamrick HJ, Lawson EE, Ulshen MH. Effect of institutional prestige on reviewers' recommendations and editorial decisions. *Presentado en Second International Congress on Peer Review in Biomedical Publication, Chicago, September 9-11, 1993*. Chicago: American Medical Association; 1993:17-18.

Gilbert J, Williams E, Lundberg GD. *Presentado en Second International Congress on Peer Review in Biomedical Publication, Chicago, September 9-11, 1993*. Chicago: American Medical Association; 1993:18.

Ginzberg E. *Health services research. Key to health policy*. Cambridge, MA: Harvard University Press; 1991.

Goldman L, Loscalzo AF. Fate of cardiology research originally published in abstract form. *N Engl J Med* 1980;303:255-259.

Goldstein DJ. [Letter]. *Clin Pharmacol Ther* 1993;53:496.

Gómez C, et al. *La producción científica española en biomedicina y ciencias de la salud a través de las bases de datos SCI y SSCI. Estudio del período 1990-93 y comparación con el cuatrienio 1986-89*. Madrid, Barcelona: Consejo Superior de Investigaciones Científicas, Institut Municipal d'Investigació Mèdica; 1996.

González JC, Pulido M, Sanz F. Evaluación del uso de procedimientos estadísticos en los artículos originales publicados en MEDICINA CLÍNICA (1962-1992). *Med Clin (Barc)* 1995;104:448-452.

Goodman SN. [Reply to Campillo C, 1993]. *Ann Intern Med* 1993;119:541.

Gotzsche PC. Methodology and overt and hidden bias in reports in 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clin Trials* 1989;10:31-56.

Gotzsche PC, Lange B. Comparison of search strategies for recalling double-blind trials from Medline. *Dan Med Bull* 1991;38:476-478.

Grant D. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychol Rev* 1962;69:54-61.

Greenwald AG. Consequences and prejudices against the null hypothesis. *Psychol Bull* 1975;82:1-20.

Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publication in a meta-analysis: is there a tower of Babel bias? *J Clin Epidemiol* 1995;48:159-163.

Haynes RB, Mulrow CD, Huth J, Altman DG, Gardner MJ. More informative abstracts revisited. *Ann Intern Med* 1990;113:69-76.

Hedges LV. Estimation of effect size under nonrandom sampling: the effects of censoring studies yielding statistical insignificant mean differences. *J Educ Stat* 1984;9:61-85.

Hedges LV, Olkin I. *Statistical methods for meta-analysis*. Orlando: Academic Press; 1985.

Hedges LV. Modelling publication selection effects in meta analysis. *Statistical Sci* 1992;7:246-255.

Hemminki E. Study of information submitted by drug companies to licesing drugs. *Br Med J* 1980;280:833-836.

Herting LL. A new approach to referees' assessments of manuscripts. *Soc Sci Res* 1990;19:1-16.

Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989;84:374-380.

Higginson J. Publication of "negative" epidemiological studies [Editorial]. *J Chron Dis* 1987;40:371-372.

Hill AB. The environment and disease: association or causation. *Proc Royal Acad Méd* 1965;58:295-300.

Hillman AL, Eisenberg JM, Pauly MV, et al. Avoiding bias in the conduct and reporting of cost-effectiveness research sponsored by pharmaceutical companies. *N Engl J Med* 1991;324:1362-1365.

Home PD. Techniques for ensuring that your next paper is quite unsuitable for publication. *J Royal College Physicians London* 1988;22:48-50.

Hosmer DW, Lomeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 1991;81:1630-1635.

Huston P. Information for peer reviewers. *Can Med Assoc J* 1994;150:1211-1216.

International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to medical journals. *JAMA* 1993;269:2282-2286.

Iyengar S, Greenhouse JBY. Selection models and the file-drawer problem. *Statistical Sci* 1988;3:109-117.

Jadad AR, McQuay HJ. Be systematic in your searching. *Br Med J* 1993;307:66.

Jefferson T, Demicheli V. Are guidelines for peer-reviewing economic evaluations necessary? A survey of current editorial practice. *Health Economics* 1995;4:383-388.

Johnson JD, Breimer LH. [Letter]. *Lancet* 1991;337:1102.

Juhl E, Christensen E, Tygstrup N. The epidemiology of the gastrointestinal randomized clinical trial. *N Engl J Med* 1977;296:20-22.

Kadar N. Laparoscopic surgery: publication bias and its perils [Letter]. *Am J Obstet Gynecol* 1995;172:1636-1637.

Kahn HA, Sempos CT. *Statistical methods in epidemiology*. New York: Oxford University Press; 1989.

Kahtan S. Publication bias [Letter]. *Lancet* 1991;337:1102.

Katerndahl DA. Physical activity and coronary heart disease [Letter]. *J Am Board Fam Practice* 1992;5:448-449.

Kaufman JL. Publication bias and the editorial process [Letter]. *JAMA* 1992;267:2891.

-
- Kelsey JL, Thompson WD, Evans AS. *Methods in observational epidemiology*. New York: Oxford University Press; 1986.
- Kepple SR. Why journal manuscripts are edited. *Am J Hospital Pharmacy* 1991;48:909-910.
- Kleijnen J, Knipschild P. Review articles and publication bias. *Arzneim Forsch (Drug Res)* 1992;42:587-591.
- Kleinbaum DG, Kupper LL, Muller Keith E. *Applied regression analysis and other multivariate methods*. Boston: PWS-Kent Publishing Co.; 1988.
- Knox FG. Relationship between signing and recommendation in a review. *J Lab Clin Med* 1981;97:1-3.
- Kochar MS. The peer review of manuscripts: in need for improvement. *J Chron Dis* 1986;39:147-149.
- Koren G, Shear H, Graham K, Einarson T. Bias against the null hypothesis: the reproductive hazards of cocaine. *Lancet* 1989;2:1440-1442.
- Kroch CL. A checklist system for critical review of medical literature. *Med Educ* 1985;19:392-395.
- Kuhn T. *The structure of scientific revolutions*. 2nd ed. Chicago: The University of Chicago Press; 1970.
- Landi G, Ciccone A. Publication bias via suppressed criticism [Letter]. *Lancet* 1993;341:697-698.
- Lane DM, Dunlap WP. Estimating effect size: bias resulting from the significance criterion in editorial decisions. *Br J Math Stat Psychol* 1978;31:107-112.
- Last JM. *A dictionary of epidemiology*. New York: Oxford University Press; 1988.
- Lauritsen K, Havelund T, Laursen LS, Rask-Madsen J. Withholding unfavourable results in drug company sponsored trials. *Lancet* 1987;1:1091.
- Le Fanu J. Misleading meta-analysis. Public policy is based on results of epidemiological meta-analysis that contradict common sense [Letter]. *Br Med J* 1995;310:1603-1604.
- Leizorovicz A, Haugh MC, Boissel JP. Meta-analysis and multiple publication of clinical trial reports [Letter]. *Lancet* 1992;340:1102-1103.
- Levois ME, Layard MW. Publication bias in the environmental tobacco smoke coronary heart disease epidemiologic literature. *Regul Toxicol Pharmacol* 1995; 21:184-191.

-
- Levy G. Publication bias: its implications for clinical pharmacology. *Clin Pharmacol Ther* 1992;52:115-119.
- Levy PS, Lemeshow S. Sampling of populations. *Methods and applications*. New York: John Wiley and Sons, Inc.; 1991.
- Liberati A. [Letter]. *JAMA* 1992;267:2891.
- Lichtenstein MJ, Murlow CD, Elwood PC. Guidelines for reading case-control studies. *J Chron Dis* 1987;40:893-903.
- Light RJ, Pillemer DB. *Summing up. The science of reviewing research*. Cambridge, MA: Harvard University Press; 1984.
- Lilienfeld AM, Lilienfeld DE. *Foundations of epidemiology*. 2nd. ed. New York: Oxford University Press; 1980.
- Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J* 1970;3:637-640.
- Lock S. *A difficult balance. Editorial peer review in medicine*. Philadelphia: ISI Press; 1985.
- Luce BR, Elixhauser A. *Standards for the economic evaluation of health care services*. Berlin: Springer-Verlag; 1990.
- Lwanga SK, Lomeshow S. *Sample size determination in health studies. A practical manual*. Geneva: World Health Organization; 1991.
- Mahon WA, Daniel EE. A method for the assessment of reports of drug trials. *Can Med Ass J* 1964;90:565-569.
- Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in peer review system. *Cog Ther Res* 1977;1:161-175.
- Mahoney MJ, Kimper TP. From ethics to logic: a survey of scientists. In: Mahoney MJ. *Scientists as subject*. Cambridge, MA: Ballinger; 1976:187-193.
- Majchrzak A. *Methods for policy research*. Newbury Park, CA: Sage Publications, Inc.; 1984.
- Makris M, Greaves M. Bias due to publication of positive results with treatments for thrombotic thrombocytopenic purpura [Letter]. *Lancet* 1991;337:1106-1107.
- Mantel N. Lung cancer and passive smoking. *Br Med J* 1987;294:440-441.
- Mausner JS, Kramer S. *Epidemiology. An introductory text*. Philadelphia: W. B. Saunders Co.; 1985.

Maxwell C. Clinical trials, reviews, and the journal of negative results. *Br J Clin Pharmacol* 1981;1:15-18.

McCormick MC, Holmes JH. Publication of reserach presented at the pediatric meetings. *AJDC* 1985;139:122-126.

McDowell I, Newell C. *Measuring health. A guide to rating scales and questionnaires*. Oxford: Oxford Univrsity Press; 1987.

McNamee R. Outcome of retrospective cohort studies and study size: a publication bias? *Br J Industrial Med* 1989;46:143.

McNutt M, et al. Effect of blinding in the *Journal of General Internal Medicine*. *JAMA* 1990;263:1371.

Mehta CR, Patel NR, Gray R. *J Am Stat Assoc* 1985;78:969-973.

Meinert CL. *Clinical trials. Design, conduct, and analysis*. Oxford: Oxford University Press; 1986.

Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Controlled Clin Trials* 1984;5:328-347.

Meinert CL. Toward prospective registration of clinical trials [Editorial]. *Controlled Clin Trials* 1988;9:1-5.

Meranze J, Ellison N, Greenhow DE. Publications resulting from anesthesia meeting abstracts. *Anesth Analg* 1982;61:445-448.

Michael M, Boyce WT, Wilcox AJ. *Biomedical bestiary: an epidemiologic guide to flaws and fallacies in the medical literature*. Boston: Little and Brown Co; 1984.

Mills JL, Rhoads GG, Simpson JL, et al. The absence of a relation between the periconceptional use of vitamins and neural-tube defects. *N Engl J Med* 1989;321:430-435.

Mindel JS. Failure of controlled clinical trial data to reach the literature. *Clin Pharmacol Ther* 1992;52:4-5.

Mindel JS. [Letter]. *Clin Phahrmacol Ther* 1993;53:497.

Minerva. Views. *Br Med J* 1983;287:1886.

Molloy Hubbard S, Henney JE, DeVita VT. A computer database for information on cancer treatment. *New Engl J Med* 1987;316:315-321.

Morris RD. Meta-analysis in cancer epidemiology. *Environ Health Perspectives* 1994;102:61-66.

Mosteller F, Gilbert JP, McPeck B. Reporting standards and research strategies for controlled trials. *Controlled Clin Trials* 1980;1:37-58.

-
- Muir Gray JA. Registers for clinical trials already exist [Letter]. *Br Med J* 1995;310:1604.
- Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;106:484-488.
- Mulrow CD, Thacker SB, Pugh JA. A proposal for more informative abstracts of review articles. *Ann Intern Med* 1988;108:613-615.
- Newcombe RG. Towards a reduction in publication bias. *Br Med J* 1987;295:656-659.
- Nicholson PA. Information for drug trials participants. *Lancet* 1987;2:396.
- Nolasco A, Gascón E, Mur P, Ferrándiz E, Álvarez-Dardet C. Utilización de la estadística en publicaciones médicas: una comparación internacional. *Med Clin (Barc)* 1986;86:841-844.
- Novack GD. Industrial publication of controlled clinical trial data [Letter] *Clin Pharmacol Ther* 1993;5:495-496.
- O'Brian Smith E. Interpreting results of a "negative" study. *J Pediatr* 1986;109:828-829.
- O'Fallon JR, Dubey SD, Salsburg DS, Edmonson JH, Soffer A, Colton T. Should there be statistical guidelines for medical research papers? *Biometrics* 1978;34:687-695.
- Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;138:697-703.
- Palmas W, Denton TA, Diamond TR. Publication criteria for statistical prediction models. *Ann Intern Med* 1993;118:231-232.
- Parmley WW. Publication bias. *J Am College Cardiol* 1994;24:1424-1425.
- Pathmanatan I, Nik-Safiah NI. *Training of trainers for health systems research*. Ottawa: International Development Research Centre; 1991. (Health systems research training series, Vol. 5.)
- Piantadosi S, Byar DP. A proposal for registering clinical trials. *Controlled Clin Trials* 1988;9:82-84.
- Pockock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. *New Engl J Med* 1987;317:426-432.
- Pollock VE. [Reply to the letter by Earleywine M]. *Am J Psychiatry* 1993;150:1436.
- Poynard T, Conn HO. The retrieval of randomized clinical trials in liver disease from the medical literature. A comparison of MEDLARS and manual methods. *Controlled Clin Trials* 1985;6:271-279.

Poynard T. Evaluation de la qualité méthodologique des essais thérapeutiques randomisés. *La Presse Médicale* 1988;17:315-318.

Powe NR, Turner JA, Maklan CW, Ersek M. Alternative methods for formal literature review and meta-analysis in AHCPR patient outcomes research teams. *Med Care* (Suppl.) 1994;32:JS22-JS37.

Preston NW. Publication bias via suppressed expert conflict [Letter]. *Lancet* 1993;341:1031.

Programa de Publicaciones, Organización Panamericana de la Salud/Organización Mundial de la Salud. *Guía para la revisión de manuscritos*. Washington, DC: OPS/OMS; 1993. (Documento no publicado.)

Relman AS. Dealing with conflict of interest. *N Engl J Med* 1984;310:1183-1184.

Relman AS. Dealing with conflict of interest [Correspondence]. *N Engl J Med* 1984;311:405.

Rennie D, Flanagan A. The triumph of hope over experience. *JAMA* 1992;267:411-412.

Rennie D. [Reply]. *JAMA* 1992;267:2892.

Riegelman RK. *Studying a study and testing a test. How to read the medical literature*. Boston: Little, Brown and Co.; 1981.

Rogers EM. *Diffusion of innovations*. 3rd ed. New York: The Free Press; 1983.

Rose G. *The strategy of preventive medicine*. Oxford: Oxford University Press; 1993.

Rosenthal R. The "file drawer problem" and tolerance for null results. *Psychol Bull* 1979;86:638-641.

Rothman KJ. *Modern epidemiology*. Boston: Little, Brown and Co.; 1986.

Rubin DB. A new perspective. In: Wachter KW, Straf ML, eds. *The future of meta-analysis*. New York: The Russell Sage Foundation; 1990:955-963.

Sackett DL, Spitzer WO. Guidelines for improving meta-analysis [Letter] *Lancet* 1994;343:910.

San Martín H, Martín AC, Carrasco JL. *Epidemiología. Teoría, investigación, práctica*. Madrid: Díaz de Santos; 1986.

SAS Institute. *SAS Release 5.18*. Cary, NC: SAS Institute; 1986

Schlesselman JJ. *Case-control studies. Design, conduct, analysis*. Oxford: Oxford University Press; 1982.

Scott WA. Interreferee agreement on some characteristics of manuscripts submitted to the *Journal of Personality and Social Psychology*. *Am Psychologist* 1974;29:698-702.

Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966;195:1123-1128.

Shadish WR, Doherty M, Montgomery LM. How many studies are in the file drawer? An estimate from the family/marital psychotherapy literature. *Clin Psychol Rev* 1989;9:589:603.

Shapiro S. The decision to publish. Ethical dilemmas. *J Chron Dis* 1985;38:365-372.

Sharp DW. What can and should be done to reduce publication bias? The perspective of an editor. *JAMA* 1990;263:1390-1391.

Shear SL. Publication bias in pharmaceutical industry-sponsored research. *J Am Acad Dermatol* 1993;28:1024.

Silva LC. *Muestreo para la investigación en ciencias de la salud*. Madrid: Díaz de Santos; 1993.

Silva LC, Pérez-Nieves C, Cuellar L. Uso de métodos estadísticos en dos revistas médicas con alto factor de impacto. *Gac Sanit* 1995;9:189-195.

Silva G, Campillo C. Cómo se deben evaluar los artículos presentados para publicación. *Med Clin (Barc)* 1991;97:744-748.

Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529-1541.

Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;6:11-29.

Simon R. Randomized clinical trials and research strategy. *Cancer Tratament Rep* 1982;60:1083-1087.

Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995;48:9-18.

Smart RG. The importance of negative results in psychological research. *Can Psychologist* 1964;5:225-232.

Smith ML. Publication bias and meta-analysis. *Eval Educ* 1980;4:22-24.

Smith DC, Donohue LR, Waszak SJ. A hospital review of advanced gynecologic endoscopic procedures. *Am J Obstet Gynecol* 1994;170:1635-1642.

Smith DR. [Reply to the letter by Kadar N]. *Am J Obstet Gynecol* 1995;172:1636-1637.

Sommer B. The file drawer effect and publication rates in menstrual cycle research. *Psychol Women Quarterly* 1987;11:233-242.

Staquet MJ, Rozenewig M, Von Hoff DD, Maggia FM. The delta and epsilon errors in the assessment of cancer trials. *Cancer Treatment Rep* 1979;63:1917-1921.

Starfield B. *The effectiveness of medical care*. Baltimore, MD: The Johns Hopkins University Press; 1985.

Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance --or viceversa. *Am Stat Assoc J* 1959;54:30-34.

Stiller MJ, Mehrel T. Multicenter clinical drug trials: should single center publish their results independently? *J Am Acad Dermatol* 1992;27:646.

Stouffer SA, Suchman EA, Devinney LC, Star SA, Williams RA. *The American soldier: adjustment during army life*. Vol. I. Princeton, NY: Princeton University Press; 1949.

Strayhorn J, McDermott JF, Tanguay P. An intervention to improve reliability of manuscript reviews for the *Journal of the American Academy of Child and Adolescent Psychiatry*. *Am J Psychiatry* 1993;150:6:947-952.

Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use*. Oxford: Oxford University Press; 1989.

Sugita M, Kanamori M, Izuno T, Miyakawa M. Estimating a summarized odds ratio whilst eliminating publication bias in meta-analysis. *Jpn J Clin Oncol* 1992;22:354-358.

Swaen GMH, Meijers JMM. Influence of design characteristics on the outcome of retrospective cohort studies. *Br J Ind Med* 1988;45:624-629.

Szklo M. Issues in publication and interpretation of research findings. *J Clin Epidemiol* 1991;44(Suppl 1):109S-113S.

Thijs C, Knipschild P. Oral contraceptives and the risks of gallbladder disease: a meta-analysis. *Am J Public Health* 1993;83:1113-1120.

Thompson SG, Clarke MJ. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994;309:1351-1355.

Tunstall-Pedoe H. [Letter]. *N Engl J Med* 1990;322:1984.

Udvarhelyi S, Colditz GA, Epstein AM. Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly? *Ann Intern Med* 1992;116:238-244.

UK Cancer Trials Register. *Lancet* 1982;1:293.

US National Library of Medicine, Department of Health and Human Services. *Grateful Med. Version 6.6*. Bethesda, MD: Department of Health and Human Services; 1995.

US Preventive Services Task Force. *Guide to clinical preventive services*. Baltimore, MD: Williams and Wilkins; 1989.

Vandenbroucke JP. Passive smoking and lung cancer: a publication bias? *Br Med J* 1988;296:391-392.

Vandenbroucke JP, Rosendaal FR. Publication bias [Letter]. *Lancet* 1994;343:119.

Verstraete M. Registry of prospective clinical trials. Sixth report. *Thromb Haemostas* 1984;51:283-290.

Wald NJ, Nanchackal K, Thompson SG, et al. Does breathing other people's tobacco cause lung cancer? *Br Med J* 1986;293:1217-1222.

Walter SD. Methods for reporting statistical results from medical research studies. *Am J Epidemiol* 1995;141:896-906.

Warren KS, Mosteller F, eds. *Doing more good than harm. The evaluation of health care interventions*. *Ann NY Acad Sci* 1993;703.

White KR. The relationship between socioeconomic status and academic achievement. *Psychol Bull* 1982;91:461-481.

Wilmschurst P. Publication bias [Letter]. *Lancet* 1991;337:1419.

Wilson WR, Miller HA. A note on the inconclusiveness of accepting the null hypothesis. *Psychol Rev* 1964;71:238-242.

Wilson A, Henry DA. 10. Meta-analysis. Part 2: assessing the quality of published meta-analysis. *Med J Australia* 1992;156:173-187.

Young MJ, Bresnitz EA, Strom BL. Sample size nomograms for interpreting negatives clinical studies. *Ann Intern Med* 1983;99:248-251.

Zelen M. Strategy and alternate randomized designs in cancer clinical trials. *Cancer Treatment Rep* 1982;66:1095-1100.

Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *J Clin Oncol* 1983;1:164-169.

CAPÍTULO 8. RESUMEN

INTRODUCCIÓN

La validez de los procesos editoriales de revisión y publicación depende fundamentalmente de que se rijan por la calidad científica de los estudios que reciben las revistas para publicación. Desde los años cincuenta, numerosos estudios –inicialmente realizados en el campo de la psicología– han demostrado que la calidad no siempre predomina como criterio de selección editorial de artículos. Se ha comprobado que la probabilidad de que se publique un estudio que notifica resultados estadísticamente significativos es más alta que la de un estudio que no alcanza el nivel de significación prefijado. Esta selección sesgada se ha denominado sesgo de publicación. Los primeros estudios realizados en medicina se centraron en descubrir la existencia de este sesgo en distintas especialidades médicas. Posteriormente, se desarrollaron métodos matemáticos para demostrar su existencia, cuantificar su magnitud y corregir las estimaciones de los estudios publicados en función de la magnitud del sesgo. La aplicación de diversos diseños a la investigación de este sesgo reveló la existencia de otros factores de riesgo además de la significación estadística, tales como el tamaño muestral, diseño del estudio, idioma en que se escribió el artículo, prestigio de los autores o fuente de financiación. Los vínculos de los sesgos de publicación con las revisiones de la literatura, los registros de estudios y el metanálisis, aunados con sus consecuencias en la práctica médica, explican la importancia que se les ha otorgado en la última década. Para minimizar sus efectos se ha propuesto registrar los estudios desde su etapa de protocolo, incluir estudios en metanálisis y revisiones sistemáticas en función de su calidad y educar a los autores –principales responsables del sesgo– para que utilicen este criterio y no otros al decidir publicar o no sus estudios. Los editores de las revistas médicas han de vigilar periódicamente la introducción

de estos sesgos en la selección de manuscritos para publicación. Los objetivos de esta memoria de investigación son conocer qué factores están asociados con la publicación y el rechazo de los artículos originales que reciben dos de las revistas españolas de mayor difusión: *Atención Primaria* y *Medicina Clínica*; investigar en ellas la presencia de sesgos de publicación y, en caso afirmativo, estimar si son introducidos por autores o editores; conocer su perfil editorial en relación con los artículos que publican, y evaluar la validez de sus procesos de revisión y selección de artículos originales.

MATERIALES Y MÉTODOS

El estudio se realizó con una muestra consecutiva y representativa de 200 artículos originales, 100 de *Atención Primaria* y 100 de *Medicina Clínica*. De los 100 artículos de cada revista, 50 habían sido publicados y 50, rechazados. Los estudios debían presentar diseños observacionales, cuasiexperimentales y experimentales. Todos los artículos debían haber sido revisados en las revistas durante un período exento de cambios en las políticas de revisión y selección de manuscritos. El diseño utilizado fue el de casos (artículos publicados) y controles (artículos rechazados). La variable de resultado fue el estado de publicación (aprobado o rechazado). La exposición se definió como el conjunto de variables que pueden determinar la publicación o el rechazo de un artículo: región de procedencia de los autores, tamaño muestral, objeto y tipo de estudio, complejidad de los análisis, significación estadística y calidad global del artículo, todas ellas consideradas como posibles factores de riesgo del sesgo de publicación. Todos los artículos se evaluaron con un cuestionario previamente validado y a ciegas del estado de publicación, la revista y la procedencia de los autores. La calidad se midió a ciegas de esas tres variables con una escala de 1 a 5 previamente validada. En los análisis se calcularon las distribuciones de frecuencias de las variables de interés. Además, se averiguó la existencia de asociaciones

entre variables y se estimó su fuerza por medio de la razón de posibilidades. Con objeto de analizar la asociación entre el estado de publicación y las restantes variables se construyeron modelos logísticos de la covarianza. La búsqueda de sesgos de publicación y procedencia se basó en la detección de asociaciones ajustadas entre el estado de publicación y, respectivamente, la significación estadística y la procedencia.

RESULTADOS

De 200 artículos, se excluyeron 4 (3 de *Atención Primaria* y 1 de *Medicina Clínica*), porque no correspondían al tipo de estudio seleccionado. La mediana de la calidad fue 2,0 (recorrido: 1,0 - 4,8). Las medias de la calidad total de los artículos publicados fueron más altas que las de los rechazados ([*Atención Primaria*: $\tilde{X}_{\text{publicados}} = 2,34$; IC95%: 2,20 - 2,47, $\tilde{X}_{\text{rechazados}} = 1,60$; IC95%: 1,48 - 1,72]. [*Medicina Clínica*: $\tilde{X}_{\text{publicados}} = 2,70$; IC95%: 2,55 - 2,92, y $\tilde{X}_{\text{rechazados}} = 1,80$; IC95%: 1,68 - 1,89].), al igual que la de todas las secciones y subpartados de los artículos publicados, excepto la calidad de los resúmenes de los artículos publicados en *Medicina Clínica*, que no difirió de la de los rechazados. La calidad se asoció con el estado de publicación: $RP_{\text{Atención Primaria}} = 17,28$; IC95%: 5,68 - 55,07, y $RP_{\text{Medicina Clínica}} = 43,38$; IC95%: 10,60 - 241,59. De los 196 artículos evaluados, se pudo conocer la procedencia de 173 (88,26%): 129 (74,6%) procedían de Cataluña y 44 (25,4%), de otras regiones del territorio español. La posibilidad de que un artículo recibido por *Atención Primaria* procediera de Cataluña fue más alta que la un artículo recibido por *Medicina Clínica* ($RP = 2,68$; IC95%: 1,25 - 5,78). La procedencia no se asoció con el estado de publicación en ninguna de las dos revistas. La posibilidad de que la calidad de un artículo de otra región fuese intermedia (frente a baja) fue más alta que la de un artículo proveniente de Cataluña sólo en *Medicina Clínica* ($RP = 7,41$; IC95%: 1,85 - 42,52). De los 196 artículos revisados, en 182 (92,86%; IC95%: 89,26% - 96,46%) se analizaron

muestras y en los 14 (7,14%) restantes, poblaciones. En 154 el tamaño muestral fue mayor de 50 y en los restantes, menor, variable que no se asoció con el estado de publicación ($RP = 1,04$; $IC95\%: 0,43 - 2,51$). En *Atención Primaria* el objeto de estudio más frecuente fue servicios, programas o proyectos, seguido por poblaciones, comunidades y personas, y por desarrollo, utilización y evaluación de tecnología. El menos frecuente fue <<otro>>. En *Medicina Clínica*, poblaciones, comunidades y personas ocupó el primer lugar, seguido por desarrollo, utilización y evaluación de tecnología, y por servicios, programas o proyectos. Las frecuencias de los objetos de estudio de los artículos publicados por las dos revistas fue muy similar a la de los artículos rechazados, excepto en *Medicina Clínica*, donde el porcentaje de artículos publicados sobre servicios, programas o proyectos fue mayor que el de rechazados. El objeto de estudio no se asoció con la calidad. En *Atención Primaria* los tres tipos de estudios más frecuentes fueron, por orden decreciente, los de prevalencia, los comparativos y los descriptivos y en *Medicina Clínica*, comparativos, de prevalencia y descriptivos. Los porcentajes que representaron estos tres tipos de estudios en conjunto en cada revista no difirieron. Sólo se encontraron diferencias entre los descriptivos y los de prevalencia en *Atención Primaria*, en la cual se rechazaron más estudios descriptivos de los que se publicaron y se publicaron más estudios de prevalencia de los que se rechazaron. Por este motivo, el tipo de estudio se asoció con el estado de publicación solamente en *Atención Primaria*, no en *Medicina Clínica*. De los 196 artículos, 191 presentaban análisis estadísticos. En 128 (67,0%; $IC95\%: 60\% - 74\%$) la complejidad fue intermedia, en 40, baja (20,9%; $IC95\%: 14,9\% - 26,9\%$) y en 23, alta (12,9%; $IC95\%: 7,9\% - 17,9\%$). Se rechazaron más artículos con análisis de complejidad baja y alta de los que se publicaron, y se publicaron más de complejidad intermedia de los que se rechazaron. La complejidad no se asoció con la calidad en ninguna revista. De los 196 artículos revisados, sólo en 146 (62 de *Atención Primaria* y 84 de *Medicina Clínica*) se aplicaron pruebas de hipótesis: 143

(97,9%) fueron positivos y 3 (2,1%), negativos. De esos 146, 79 se publicaron y 67 se rechazaron. A nivel multivariante, en *Atención Primaria* el tipo de estudio ($RP = 4,30$; $IC95\%: 1,45 - 12,50$) y la calidad ($RP = 10,42$; $IC95\%: 3,45 - 31,25$) se asociaron con el estado de publicación, y en *Medicina Clínica* sólo la calidad se asoció a con el estado de publicación ($RP = 30,3$ ($IC95\%: 7,14 - 125,0$)).

DISCUSIÓN

Los resultados revelan que los procesos de revisión y selección de artículos para publicación de *Atención Primaria* y *Medicina Clínica* son efectivos, ya que el criterio cardinal que gobierna la publicación de artículos originales es su calidad científica, a pesar de que tanto la de los artículos publicados como rechazados por ellas sea baja. Las deficiencias más frecuentes encontradas en los diferentes apartados de los artículos coinciden con las notificadas en otros estudios y responden fundamentalmente a limitaciones en la capacidad investigadora de los autores y a la débil adhesión de las revistas a los requisitos de edición de artículos originales formulados por comités internacionales de editores de revistas biomédicas. En ninguna de las dos revistas se detectó sesgo de procedencia. Ni sus editores ni revisores introducen sesgo de publicación en la selección de manuscritos y los resultados indican que su existencia es posible y que los autores son enteramente responsables de su introducción, lo cual coincide con las hipótesis barajadas en la literatura internacional sobre este tema. Ambas revistas difieren en cuanto a sus perfiles temáticos y a la distribución de frecuencias de los objetos de estudios, no con respecto al tipo de estudio. Aunque los resultados son compatibles con la existencia de asociaciones entre la complejidad del diseño y de los análisis estadísticos y la calidad de los artículos, estas asociaciones, de existir, son muy débiles, lo cual subraya la necesidad de mejorar la calidad metodológica y la notificación de los estudios que se realizan en

medicina en las zonas y áreas temáticas que representan estas dos revistas en España. El modelo presentado en este estudio ofrece a los editores de revistas médicas una herramienta novedosa y útil para analizar y monitorizar los factores que rigen la publicación y el rechazo de artículos científicos originales, estimar la calidad de los artículos recibidos por las revistas para publicación, así como estimar la existencia de sesgos de publicación; es decir, supervisar aspectos vinculados con la validez de sus procesos de publicación de artículos originales.