
La estadística en la investigación clínica de medicamentos. Estudio de artículos originales procedentes de centros españoles

M.C. Carré Llopis^a, J. Jiménez Villa^b, M. Martín Mateo^c
y F. Jané Carrencà^d

^aGDIS. Barcelona. ^bRegió Sanitària Costa de Ponent. Servei Català de la Salut. L'Hospitalet de Llobregat (Barcelona).

^cDepartament de Bioestadística. Facultat de Medicina. Universitat Autònoma de Barcelona. Bellaterra (Barcelona).

^dDepartament de Farmacologia i Psiquiatria. Universitat Autònoma de Barcelona. Hospital de la Santa Creu i Sant Pau. Barcelona.

El uso de la estadística en la investigación clínica de medicamentos es un fenómeno que se ha ido implantando progresivamente desde que se introdujo el método estadístico en el estudio de la eficacia y seguridad de los fármacos.

En uno de los primeros estudios sobre el uso de la estadística en los artículos originales, Schor y Karten¹ observaron que una gran proporción no usaban las pruebas estadísticas en el análisis de los resultados, o bien, contenían errores importantes. Si bien la complejidad del análisis estadístico ha aumentado desde entonces²⁻⁵, todavía se aprecian errores en un porcentaje importante de artículos⁶⁻⁸. En las publicaciones españolas este fenómeno es más evidente que en las anglosajonas⁹⁻¹¹.

La mayor disponibilidad de paquetes informáticos ha contribuido a la extensión del uso de la estadística y a la utilización, y en ocasiones abuso, de técnicas sofisticadas. Esta tendencia no ha ido pareja a un aumento de su calidad y ha conducido a una disminución de la capacidad de lectura crítica de los profesionales sanitarios¹².

Este estudio pretende cuantificar la frecuencia de utilización de las pruebas estadísticas y evaluar otros aspectos del análisis de los datos en los artículos originales sobre investigación clínica de medicamentos realizados en centros españoles.

Material y métodos

Las unidades de estudio han sido los artículos de investigación clínica de medicamentos realizados en centros españoles. La población de estudio se obtuvo de la base de datos Embase. La estrategia de búsqueda se restringió al apartado de fármacos y productos químicos administrados exógenamente, a España, a estudios realizados

en seres humanos y a artículos publicados en 1975, 1980, 1985 y 1990.

Se consideró como un artículo de investigación a todo estudio que presentara datos originales, publicado en cualquier sección de la revista, con exclusión de las de cartas al director, editoriales, artículos de opinión o de actualización y revisiones. También se excluyeron los artículos que presentaban uno o más casos aislados de un proceso o enfermedad, o reacciones adversas por drogadicción.

Se incluyeron todos los artículos obtenidos en la búsqueda que cumplían los criterios de selección. Para la recogida de datos se utilizó únicamente la información que aparecía en el texto del artículo. Se diseñó un formulario que recogía datos sobre el análisis estadístico.

Se consideró que en un artículo se había usado la estadística cuando se había realizado algún tipo de análisis inferencial. En caso afirmativo, se valoró el grado de su descripción como "adecuado" cuando existía una descripción lo suficientemente detallada para que un lector con conocimientos de estadística pudiera comprender y reproducir el análisis estadístico utilizado. Se valoró como "inadecuado" cuando el análisis no se describía o se hacía de forma insuficiente, hecho que no permitía su replicabilidad.

Los procedimientos estadísticos se clasificaron según las categorías de Emerson y Colditz¹². La clasificación original que constaba de 21 categorías ya había sido modificada por otros autores²⁻⁴. Dado que en esta revisión algunas de las categorías tenían muy pocos representantes, se redujeron a 14 (tabla I). En la categoría de medidas epidemiológicas se incluyó la de estandarización, que se refería a las medidas de incidencia y prevalencia¹². Las categorías de tablas de vida, regresión de supervivencia y otros análisis de supervivencia se integraron en una sola. La categoría de potencia, que incluye la utilización de una diferencia detectable o útil en la determinación

TABLA I
CATEGORÍAS DE PROCEDIMIENTOS ESTADÍSTICOS UTILIZADAS PARA EVALUAR
EL CONTENIDO ESTADÍSTICO DE LOS ARTÍCULOS

<i>Categoría</i>	<i>Descripción</i>
Sin métodos estadísticos o sólo descriptivos	Sin contenido estadístico, o solamente métodos descriptivos (porcentaje, media, desviación estándar, error estándar...)
t de Student	t de Student para datos independientes y para datos apareados, Z de comparación de medias
Tabla de contingencia	Ji al cuadrado, prueba exacta de Fisher, prueba de McNemar
Pruebas no paramétricas	Prueba de Mann-Whitney, prueba de Wilcoxon, prueba de Kruskal-Wallis...
Medidas epidemiológicas	Riesgo relativo, <i>odds ratio</i> , sensibilidad, especificidad
Correlación de Pearson	r de Pearson
Regresión lineal simple	Regresión lineal con una variable dependiente y una independiente
Análisis de la variancia	Análisis de la variancia de una o varias vías, F de comparación de variancias
Correlación no paramétrica	Correlación de Spearman, tau de Kendall
Regresión lineal múltiple	Regresión lineal con más de una variable independiente
Comparaciones múltiples	Métodos de Bonferroni, Scheffé, Tukey, Duncan, Neumann-Keuls...
Tabla de vida y análisis de supervivencia	Métodos actuarial, de Kaplan-Meier, prueba log-rank, regresión de Cox, regresión logística
Otros procedimientos	Procedimientos no incluidos en las categorías anteriores (pruebas de Kolmogorov-Smirnov, análisis discriminante, índice kappa...)
Procedimiento no especificado	Utilización de un procedimiento estadístico sin identificación explícita

del tamaño de la muestra se eliminó, ya que se analizó como un aspecto específico del diseño del estudio. Se excluyeron las categorías de transformación de variables y de análisis coste-beneficio. En la categoría otros procedimientos, se incluyó la de análisis de la sensibilidad (sensibilidad a pequeños cambios en los parámetros del modelo o en otras asunciones). El resto de categorías no se modificaron. Se añadió una última categoría correspondiente a prueba estadística no identificada.

Cada procedimiento estadístico se contabilizó una única vez por artículo, independientemente del número de ocasiones en que se mencionaba. Cada artículo se contabilizó una sola vez en cada categoría, aunque utilizara más de un procedi-

miento incluido en la misma. Dado que en un mismo trabajo podían coexistir procedimientos de diferentes categorías, un artículo podía estar contabilizado en más de una de ellas.

Emerson y Colditz denominaron accesibilidad estadística al porcentaje de artículos que una persona podría leer (y comprender) con determinados conocimientos¹². Para poder construir este índice se ordenaron los procedimientos según su complejidad, tal como se presenta en la tabla I, de forma que cada artículo se codificó en la categoría más elevada. En su determinación se excluyeron los artículos que utilizaban pruebas no identificadas.

La presentación de los resultados en el artículo se consideró "adecuada" cuando era fácil-

TABLA II
ARTÍCULOS INCLUIDOS EN EL ESTUDIO SEGÚN LA NACIONALIDAD
Y EL AÑO DE PUBLICACIÓN DE LAS REVISTAS

Nacionalidad	Año			Total
	1980	1985	1990	
Española	63 (78,8)	41 (64,1)	73 (50,7)	177 (61,5)
Extranjera	17 (21,3)	23 (35,9)	71 (49,3)	111 (38,5)
Total	80 (27,8)	64 (22,2)	144 (50)	288 (100)

$\chi^2 = 0,001$. Resultados expresados como n.º de artículos (tanto por ciento).

mente comprensible y reflejaba suficientemente la estrategia de análisis y “no adecuada” cuando existían errores o críticas tales como falta de claridad o confusión en los datos presentados, o si éstos aparecían en una secuencia que no permitía seguir el análisis. Se registró si se utilizaban intervalos de confianza (IC) para indicar la magnitud del resultado principal. En los estudios comparativos se recogió la existencia de significación estadística y, en caso afirmativo, al grupo al que favorecía. En los estudios con resultados estadísticamente no significativos (estudios negativos), se registró si se había calculado la potencia estadística o, como mínimo, si se discutía explícitamente la posibilidad de incurrir en un error beta o de que el tamaño de la muestra hubiera sido insuficiente para detectar el efecto de interés.

La recogida de datos se realizó por un solo observador, que procedió a la lectura del artículo identificando la existencia y el grado de cumplimiento de cada uno de los ítems. Previamente a la recogida de datos se realizó un análisis de concordancia entre dos observadores de forma enmascarada, con 30 artículos de investigación clínica de medicamentos indizados en Embase en años distintos a los revisados. Se obtuvo un índice kappa que variaba, según los ítems, entre 0,82 y 0,93.

Para el análisis, las revistas se consideraron españolas cuando eran editadas en España, aunque lo fueran en inglés, y extranjeras en caso contrario.

El análisis estadístico se efectuó con el programa SPSS/PC + 4.0. La comparación entre las revistas españolas y las extranjeras se realizó mediante la prueba de la ji al cuadrado o la prueba exacta de Fisher cuando se trataba de proporciones y la prueba de la t de Student-Fisher cuando eran medias. La comparación entre los años de publicación se realizó mediante la prueba de tendencia lineal de Mantel para las proporciones y el análisis de la variancia para las me-

dias. En todos los contrastes de hipótesis, el nivel de significación se fijó en el 5 %.

Resultados

Se incluyeron 288 artículos de investigación clínica de medicamentos realizados en centros españoles. En 1975 no se halló ninguno, mientras que la mitad correspondían a 1990. Ciento setenta y siete artículos (61,5 %) se publicaron en revistas españolas y 111 (38,5 %) en revistas extranjeras. El porcentaje de artículos en revistas extranjeras aumentó significativamente en los años evaluados (tabla II).

En 211 artículos (73,3 %; IC 95 %: 68,2-78,4) se utilizaba algún procedimiento estadístico, con un aumento significativo ($p < 0,001$) a lo largo de los años revisados (tabla III). Dicho porcentaje fue superior en las revistas extranjeras que en las españolas (84,7 frente al 66,1 %, $p < 0,001$).

En 121 artículos (57,3 %; IC 95 %: 50,6-64,0) se describían de forma adecuada las pruebas estadísticas usadas en el análisis de resultados y no se observaron diferencias estadísticamente significativas entre revistas españolas y extranjeras ($p = 0,25$), aunque se aprecia una mejora significativa entre los años estudiados ($p < 0,001$), principalmente entre 1980 (27,3 %) y 1985 (58 %) (tabla III).

La categoría utilizada con mayor frecuencia fue la de la t de Student (33,7 %; IC 95 %: 28,2-39,2), seguida de la de tablas de contingencia (28,8 %; IC 95 %: 23,6-34,0) y de la falta de uso de la estadística o sólo de métodos descriptivos (27,8 %; IC 95 %: 22,6-33,0) (tabla IV). En relación con las revistas españolas, en las extranjeras utilizaron con más frecuencia las categorías de tablas de contingencia (43,2 %; IC 95 %: 37,5-48,9), análisis de supervivencia (18,9 %, IC 95 %: 14,4-23,4) y otros procedimientos (4,5 %; IC 95 %: 2,1-6,9). En cambio, en las revistas españolas fue más frecuente la categoría de falta de uso de estadística o sólo de

TABLA III
USO DE PROCEDIMIENTOS ESTADÍSTICOS Y SU GRADO DE DESCRIPCIÓN
EN EL ARTÍCULO EN EL ANÁLISIS DE RESULTADOS, SEGÚN LA NACIONALIDAD
Y AÑO DE PUBLICACIÓN DE LA REVISTA

	Nacionalidad			Año			<i>p</i>	Total (<i>n</i> = 288)
	Española (<i>n</i> = 177)	Extranjera (<i>n</i> = 111)	<i>p</i>	1980 (<i>n</i> = 80)	1985 (<i>n</i> = 64)	1990 (<i>n</i> = 144)		
Usan estadística	117 (66,1)	94 (84,7)	< 0,001	44 (55,0)	50 (76,9)	117 (81,8)	< 0,001	211 (73,3)
Descripción adecuada de la estadística ^a	63 (53,4)	58 (61,7)	0,25	12 (27,3)	29 (58,0)	80 (68,4)	< 0,001	121 (57,3)

^aEn estudios que usan estadística.

Resultados expresados como n.º de artículos (tanto por ciento).

procedimientos descriptivos (35,6 %, IC 95 %: 30,1-41,1). El 8,7 % (IC 95 %: 5,4-12,0) de artículos no identificaban las pruebas estadísticas utilizadas, sin observarse diferencias entre revistas españolas y extranjeras en este aspecto.

En los años estudiados se observó un aumento estadísticamente significativo en la utilización de las categorías de *t* de Student, tablas de contingencia, pruebas no paramétricas, análisis de la variancia, comparaciones múltiples y análisis de supervivencia (tabla IV). Tan sólo se observó un descenso estadísticamente significativo en la categoría de falta de uso de estadística o sólo de técnicas descriptivas.

La complejidad del análisis estadístico fue superior en las revistas extranjeras. Mientras que en las revistas españolas con el conocimiento de las siete primeras categorías se podía comprender el 85 % de los artículos, en las extranjeras ello ocurría únicamente con el 60,6 % (tabla V). La complejidad también aumentó con los años valorados. El conocimiento de las pruebas incluidas en las siete primeras categorías permitía comprender el 95,7 % de los artículos publicados en 1980, el 80,4 % en 1985 y únicamente el 63,2 % en 1990. Cabe destacar que en 1980 el conocimiento de las tablas de contingencia y de la *t* de Student permitía comprender el 90 % de los ar-

TABLA IV
PRUEBAS ESTADÍSTICAS UTILIZADAS SEGÚN LA NACIONALIDAD Y AÑO
DE PUBLICACIÓN DE LA REVISTA

Categoría	Nacionalidad			Año			<i>p</i>	Total (<i>n</i> = 288)
	Española (<i>n</i> = 177)	Extranjera (<i>n</i> = 111)	<i>p</i>	1980 (<i>n</i> = 80)	1985 (<i>n</i> = 64)	1990 (<i>n</i> = 144)		
No estadística o sólo descriptiva	63 (35,6)	17 (15,3)	0,0002	37 (46,3)	15 (23,1)	28 (19,6)	< 0,0000	80 (27,8)
<i>t</i> de Student	61 (34,5)	36 (32,4)	0,72	18 (22,5)	23 (35,4)	56 (39,2)	0,04	97 (33,7)
Tablas de contingencia	35 (19,8)	48 (43,2)	< 0,0000	11 (13,8)	16 (24,6)	56 (39,6)	0,0002	83 (28,8)
Pruebas no paramétricas	15 (8,5)	11 (9,9)	0,68	1 (1,3)	8 (12,3)	17 (11,9)	0,017	26 (9,0)
Medidas epidemiológicas	-	2 (1,8)	0,08	-	-	2 (1,4)	0,36	2 (0,7)
Correlación de Pearson	6 (3,4)	3 (2,7)	0,74	3 (3,8)	1 (1,5)	5 (3,5)	0,70	9 (3,1)
Regresión lineal simple	8 (4,5)	2 (1,8)	0,22	2 (2,5)	4 (6,2)	4 (2,8)	0,40	10 (3,5)
Análisis de la variancia	13 (7,3)	14 (12,6)	0,14	-	5 (7,7)	22 (15,4)	0,0007	27 (9,4)
Correlación no paramétrica	-	1 (0,9)	0,21	-	1 (1,5)	-	0,18	1 (0,3)
Regresión lineal múltiple	2 (1,1)	2 (1,8)	0,64	-	1 (1,5)	3 (2,1)	0,44	4 (1,4)
Comparaciones múltiples	5 (2,8)	4 (3,6)	0,71-	-	-	9 (6,3)	0,009	9 (3,1)
Análisis de supervivencia	9 (5,1)	21 (18,9)	0,0002	2 (2,5)	3 (4,6)	25 (17,5)	0,0005	30 (10,4)
Otros procedimientos	1 (0,6)	5 (4,5)	0,02	-	2 (3,1)	4 (2,8)	0,3	6 (2,1)
Prueba no identificada	17 (7,2)	12 (6,7)	0,88	10 (11,9)	8 (9,0)	11 (4,5)	0,32	29 (7,1)
Total de pruebas utilizadas	235	178		84	89	242		413

Resultados expresados como n.º de artículos (tanto por ciento).

TABLA V
ACCESIBILIDAD ESTADÍSTICA SEGÚN LA NACIONALIDAD Y AÑO DE PUBLICACIÓN
DE LA REVISTA

Categoría	Nacionalidad		Año			Total (n = 259)
	Española (n = 160)	Extranjera (n = 99)	1980 (n = 70)	1985 (n = 56)	1990 (n = 133)	
No estadística o sólo descriptiva	61 (38,1)	16 (16,1)	37 (52,9)	13 (23,2)	27 (20,3)	77 (29,7)
t de Student	28 (55,6)	12 (28,3)	13 (71,4)	13 (46,4)	14 (30,8)	40 (45,2)
Tablas de contingencia	29 (73,8)	23 (51,5)	13 (90,0)	13 (69,6)	26 (50,4)	52 (65,3)
Pruebas no paramétricas	12 (81,3)	7 (58,6)	- (90,0)	5 (78,6)	14 (60,9)	19 (72,6)
Medidas epidemiológicas	- (81,3)	- (58,6)	- (90,0)	- (78,6)	- (60,9)	- (72,6)
Correlación de Pearson	2 (82,5)	2 (60,6)	2 (92,9)	- (78,6)	2 (62,4)	4 (74,1)
Regresión lineal simple	4 (85,0)	- (60,6)	2 (95,7)	1 (80,4)	1 (63,2)	4 (75,7)
Análisis de la variancia	9 (90,6)	10 (70,7)	- (95,7)	6 (91,1)	13 (72,9)	19 (83,0)
Correlación no paramétrica	- (90,6)	- (70,0)	- (95,7)	- (91,1)	- (72,9)	- (83,0)
Regresión lineal múltiple	1 (91,3)	- (70,0)	- (95,7)	- (91,1)	1 (73,7)	1 (83,4)
Comparaciones múltiples	4 (93,8)	3 (73,7)	- (95,7)	- (91,1)	7 (78,9)	7 (86,1)
Análisis de supervivencia	9 (99,4)	21 (94,9)	3 (100)	3 (96,4)	24 (97,0)	30 (97,7)
Otros procedimientos	1 (100)	5 (100)	- (100)	2 (100)	4 (100)	6 (100)

^aSe excluyeron los artículos que utilizaban pruebas estadísticas no identificadas.

Resultados expresados como n° de artículos (tanto por ciento).

tículos publicados, mientras que sólo permitía el acceso al 50,4 % de los publicados en 1990.

El 49 % (IC 95 %: 43,2-54,8) de los artículos presentaban de forma adecuada los resultados. Se ha observado una mejoría estadísticamente significativa en los años estudiados (tabla VI). Sólo el 22,2 % (IC 95 %: 17,4-27,0) de los artículos revisados utilizaban IC, porcentaje que fue superior en las revistas extranjeras (32,4 frente al 15,8 %, $p = 0,002$) y que presentó una mejoría significativa en los años estudiados, especialmente entre 1980 y 1985 (5 y 28,1 %, respectivamente). El 46,6 % (IC 95 %: 40,8-52,4) de los artículos publicados no encontraron diferencias estadísticamente significativas, mientras que en el 49,1 % de los casos las diferencias favorecían al grupo al que se administró la terapia en estudio. Estos porcentajes no mostraron diferencias significativas en función de la nacionalidad de las revistas ni de los años estudiados. De los 74 estudios negativos, únicamente 7 (9,5 %; IC 95 %: 6,1-12,9) habían calculado o discutido la potencia del estudio. Se ha observado una tendencia estadísticamente significativa a la disminución de dicho porcentaje en los años estudiados.

Discusión

En este estudio se evaluó la descripción de los procedimientos estadísticos utilizados que constan en el artículo. No se pretendía evaluar la ade-

cuación de las pruebas, ni el cumplimiento de las condiciones para su aplicación.

Se utilizó Embase como base de datos, ya que desde sus inicios ha utilizado al país como descriptor, su cobertura está focalizada en las revistas de farmacología y europeas, de las que indiza todos los artículos sobre fármacos. Sin embargo, evidentemente, los resultados de este estudio no reflejan exhaustivamente la situación española en la investigación de medicamentos, ya que no todos los estudios son publicados ni todos los que lo son se incluyen en las bases de datos bibliográficas.

De forma semejante a los estudios de Feinstein⁶ y Pocock et al¹³, los artículos han sido revisados por un único evaluador que, aunque está avezado en la lectura crítica de la bibliografía biomédica, podría haber introducido un sesgo, especialmente en los ítems que requerían una cierta subjetividad en su evaluación. Sin embargo, se observaron concordancias elevadas con un segundo observador "ciego" antes del inicio del estudio. Por otro lado, la evaluación por un único revisor podría disminuir la variabilidad de los resultados.

En un estudio realizado entre 1966 y 1969, Lionel y Herxheimer¹⁴ observaron que sólo se utilizaba la estadística en el 30 % de los artículos. Este porcentaje ha aumentado y actualmente más del 75 % de las publicaciones usan algún tipo de estadística inferencial^{5,15-22}, porcentaje que, en al-

TABLA VI
USO DE TÉCNICAS ESTADÍSTICAS EN LA PRESENTACIÓN DE RESULTADOS
SEGÚN LA NACIONALIDAD DE LA REVISTA

	Nacionalidad			Año			p	Total (n = 288)
	Española (n = 177)	Extranjera (n = 111)	p	1980 (n = 80)	1985 (n = 64)	1990 (n = 144)		
Presentación adecuada de resultados	82 (46,3)	59 (53,1)	< 0,001	32 (40,0)	33 (51,6)	76 (52,8)	< 0,000	141 (49,0)
Uso de intervalos de confianza	28 (15,8)	36 (32,4)	0,002	4 (5,0)	18 (28,1)	42 (29,2)	0,002	64 (22,2)
Grupo favorecido por las diferencias ^a			0,49				0,88	
Intervención	43 (53,0)	36 (45,0)		15 (51,7)	17 (47,2)	46 (47,9)		79 (49,1)
Control	4 (4,9)	3 (3,8)		2 (6,9)	2 (5,6)	3 (3,2)		7 (4,3)
Resultados sin diferencias significativas ^a	34 (42,0)	41 (51,3)		12 (41,3)	17 (47,2)	47 (50,0)		75 (46,6)
Total de estudios comparativos	81	80		29	36	96		161
Cálculo de la potencia ^b	5 (14,7)	2 (4,9)	0,15	4 (33,3)	2 (11,8)	1 (2,1)	0,009	7 (9,3)

^aSobre estudios comparativos. ^bSobre estudios negativos.

Resultados expresados como n° de artículos (tanto por ciento).

gunos casos, llega al 90 %²³. En este estudio, se ha observado que el 73,3 % de los artículos utilizaban la estadística, si bien, esta cifra era significativamente inferior en las revistas españolas (66,1 %) frente a las extranjeras (84,7 %).

El porcentaje de artículos en que la descripción del análisis estadístico puede considerarse adecuada varía ampliamente entre los distintos estudios y se ha observado en el 6 %²⁴, el 26 %²⁵, del 70 al 75 %¹³, el 85 %²³, o hasta el 92,9 %²⁶. En este estudio se ha considerado correcta en el 57,3 % de los casos, cifra muy similar al 60 % observado por Evans y Pollock²⁷. Este porcentaje ha ido mejorando progresivamente en los años evaluados, aunque sin apreciarse diferencias entre revistas españolas y extranjeras.

Existe una gran diversidad de pruebas estadísticas que pueden utilizarse en investigación clínica, dependiendo del objetivo, del diseño del estudio y de las variables implicadas. Esta diversidad hace que la frecuencia de su uso varíe ampliamente en función de los temas de interés o de la especialidad médica^{2-4,16}. En este estudio se aprecia que todavía el 8,7 % de los artículos no identifican las pruebas estadísticas que utilizan, semejante al 8,1 % observado por Jiménez Villa et al²⁶ y que el porcentaje de artículos que usan procedimientos estadísticos más allá de los puramente descriptivos es únicamente del 72,2 %. Sin embargo, se observa un incremento hasta el 80,4 % en 1990, porcentaje similar al 81,1 % observado en *Medicina Clínica* en 1991 y 1992²⁸ o

en *Atención Primaria* en 1990²⁶. Las pruebas estadísticas más utilizadas en este estudio son las bivariantes, la t de Student (33,7 %) y las tablas de contingencia (28,8 %). Los porcentajes observados en 1990 son similares a los encontrados en *Medicina Clínica* en 1991 y 1992, con el 33,3 y el 49,2 %, respectivamente y al 29 % observado en la t de Student en *Atención Primaria* en 1990²⁶, aunque en este mismo estudio se observa el 69,4 % de pruebas de contingencia. Probablemente las pequeñas diferencias observadas en relación con el estudio de Mora et al²⁸ y de Jiménez Villa et al²⁶ se deban a que únicamente se han evaluado artículos de investigación de medicamentos.

En este estudio, los procedimientos estadísticos se han agrupado en categorías y se han ordenado siguiendo las recomendaciones de Emerson y Colditz. Aunque puede ser discutible, ya que incluye cierto grado de subjetividad, estos autores justifican dicha ordenación por la cadencia y el nivel de adquisición de conocimientos estadísticos de un alumno que inicia su formación partiendo de cero, de forma que en las siete primeras categorías se incluyen los métodos bivariantes que corresponden a los contenidos tradicionales de un curso de estadística básica. Además, el índice de accesibilidad permite comparar el nivel de complejidad de las pruebas estadísticas utilizadas.

Aunque sólo se sobrepasa el nivel de la estadística bivariante en un escaso porcentaje de ar-

tículos, el grado de complejidad es superior al observado por Emerson y Colditz. Estos autores determinaron una accesibilidad del 73 % para las tres primeras categorías y del 83 % para las siete primeras¹², mientras que en ese estudio se han observado del 65,3 y del 75,7 %, respectivamente. Sin embargo, la accesibilidad habría disminuido en 1990 hasta el 50,4 y el 63,3 % para las tres y las siete primeras categorías respectivamente, porcentajes que se asemejan al 40,2 y al 61,7 % observados por Mora et al²⁸ en *Medicina Clínica* entre 1991 y 1992. Estas diferencias podrían deberse al distinto período de estudio, ya que Emerson y Colditz evaluaron artículos publicados en 1978 y 1979 y Mora et al en 1991 y 1992.

En este estudio se ha observado que, globalmente, la accesibilidad estadística es menor en los artículos procedentes de revistas extranjeras, de acuerdo con los hallazgos de otros autores^{17,28}. El uso de pruebas estadísticas más allá de las descriptivas es del 83,9 % en las revistas extranjeras, superior al 60 % observado en la revista *The New England Journal of Medicine* en 1878-1879¹² y al 70 % de *Atención Primaria* en 1989-1991¹⁰, y similar al 88-89 % observado en un estudio de diversas revistas médicas en la década de los ochenta¹⁶.

El índice de accesibilidad ha ido disminuyendo progresivamente en los años estudiados, tanto en las revistas extranjeras como en las españolas, aunque puede apreciarse un desfase de estas últimas en relación a las extranjeras. La consecuencia de este aumento de la complejidad del análisis estadístico, que según diversos autores no se ha acompañado de un aumento paralelo de los conocimientos de estadística de los profesionales²⁹⁻³², es la disminución de la capacidad de los lectores para comprender y evaluar los resultados de los estudios de investigación. A pesar de la insistencia sobre este fenómeno en los últimos años, el interés de los investigadores en estas materias todavía es escaso, lo que hace necesario dar una mayor importancia a la formación de los profesionales en metodología y estadística^{11,14}. Hubiera sido interesante evaluar la adecuación del análisis estadístico para poder comprobar si este aumento de complejidad obedece a un mejor uso de la estadística o tan sólo refleja la mayor facilidad de utilización de programas informáticos sofisticados.

En un artículo científico, la presentación de los resultados debe reflejar los objetivos de la investigación. En este estudio se ha observado que el 51 % de los artículos presentan los resultados de forma adecuada. Probablemente este porcenta-

je hubiera sido menor si se hubieran evaluado otros aspectos de la presentación estadística de resultados, como la presentación del valor exacto de la p, o la existencia de una p huérfana. De hecho, Mora et al³³ al analizar estos aspectos encontraron incorrecciones mayores en el 15 % y menores en el 82 % de los artículos publicados en *Medicina Clínica* en 1993.

El uso de los intervalos de confianza ofrece grandes ventajas en relación con la significación estadística en las publicaciones biomédicas, ya que permite conocer la estimación de la magnitud del efecto observado. Por este motivo, diversos autores y comités editoriales de revistas han recomendado su uso en las publicaciones²¹. Sin embargo, el porcentaje de artículos revisados que utilizaban IC es tan sólo del 22,2 % en el global de la muestra y del 29,2 % en el año 1990. Estas cifras son similares a las observadas por Liberati et al³⁴, e intermedias entre las encontradas por otros autores que oscilan desde el 3,6²⁷ y el 16,1 %²⁶ hasta el 51 %²⁴. Recientemente, Mora et al³³ encontraron que el 43 % de los artículos publicados en *Medicina Clínica* en 1993 presentaban intervalos de confianza.

En este estudio se ha observado que en el 49,1 % de los artículos se observan diferencias estadísticamente significativas a favor de la terapia de estudio, porcentaje semejante al de Jiménez Villa et al³⁵, mientras que en el 46,6 % no se aprecian diferencias estadísticamente significativas. Aunque este porcentaje puede ser considerado elevado, es similar al informado por otros autores^{36,37}, y podría atribuirse tanto al insuficiente número de sujetos estudiados, como al incremento de estudios que comparan fármacos de la misma familia, que buscan demostrar su equivalencia en vez de su diferencia. Sin embargo, en cualquier caso, en un estudio negativo debe determinarse su potencia estadística para poder interpretarlo adecuadamente. En este estudio se ha encontrado que tan sólo el 9 % de los artículos realizan este cálculo, corroborando los datos de otros autores^{12,23,37}.

En este estudio se ha puesto de manifiesto la insuficiente información sobre la metodología estadística en los artículos de investigación clínica de medicamentos realizados en centros españoles, especialmente cuando son publicados en revistas españolas. Aunque se aprecia una mejoría en los años evaluados en la mayoría de los ítems, la situación todavía dista mucho de ser la adecuada. Si se tiene en cuenta que la literatura biomédica es la principal vía de difusión de los avances terapéuticos, es preciso que tanto los investigadores como los comités editoriales de las

revistas adopten estrategias para mejorar no sólo la calidad de los estudios sino también de las publicaciones, proporcionando toda la información necesaria para que los lectores puedan evaluar la validez científica y la utilidad de los resultados. Así mismo, los profesionales sanitarios deberían tener la suficiente formación para realizar esta lectura crítica y evaluar la aplicabilidad de los resultados a los propios pacientes. Estudios como el presente pueden ayudar a prever a los departamentos de bioestadística de las facultades de medicina las necesidades de la formación continuada en estos temas para los profesionales de las ciencias de la salud.

BIBLIOGRAFÍA

- Schor S, Karten I. Statistical evaluation of medical journals manuscripts. *JAMA* 1966; 195: 145-115.
- Hokanson JA, Luttmann DJ, Weiss GB. Frequency and diversity of use statistical techniques in *Oncology Journals*. *Cancer Treat Rep* 1986; 70: 589-594.
- Hokanson JA, Stiernberg CM, McCracken MS, Quinn FB. The reporting of statistical techniques in *Otorrinolaryngology Journals*. *Arch Otolaryngol Head Neck Surg* 1987; 113: 45-50.
- Hokanson JA, Ladoulis CT, Quinn FB, Bienkowski AC. Statistical techniques reported in *Pathology Journals* during 1983-1985: implications for *Pathology Educators*. *Arch Pathol Lab Med* 1987; 111: 202-207.
- Meinert CL, Tonascia S, Higgins K. Content of reports on *Clinical Trials*: A critical Review. *Control Clin Trials* 1984; 5: 328-347.
- Feinstein AR. Clinical bioestatistics. XXV. A survey of the statistical procedures in in general medical journals. *Clin Pharmacol Ther* 1974; 15: 97-107.
- Altman DG. Statistics in medical journals. *Stat Med* 1982; 1: 59-71.
- Murray GD. Statistical guidelines for the *British Journal of Surgery*. *Br J Surg* 1991; 78: 782-784.
- 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.
- Farrús M, De la Fuente JA, Iglesias M, Borrell C. Revisión de los artículos originales publicados en *Atención Primaria* en Cataluña. *Aten Primaria* 1993; 12: 325-332.
- Martín M, Sanz F, Andreu D. Efecto de la introducción de la Bioestadística en el curriculum de los estudios de Medicina. Análisis de una década en la revista *Medicina Clínica (Barcelona)*. *Med Clin (Barc)* 1982; 79: 273-276.
- Emerson JD, Colditz GA. Use of statistical analysis in the *New England Journal of Medicine*. *N Engl J Med* 1983; 309: 709-713.
- Pocock SJ, Hugges MD, Lee RJ. Statistical problems in reporting of clinical trials. *N Engl J Med* 1987; 317: 426-432.
- Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J* 1970; 3: 637-664.
- Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines or contributors to medical journals. *Br Med J* 1983; 286: 1.489-1.493.
- Altman DG. Statistics in Medical Journals: developments in the 1980's. *Stat Med* 1991; 10: 1.887-1.913.
- Álvarez-Dardet C, Gascón E, Mur P, Nolasco A. 10 years trends in the *Journal's* publications. *N Engl J Med* 1985; 312: 1.521-1.522.
- Bombardier C, Tugwell P, Sinclair A, Dok C, Anderso G, Buchanan WW. Preference for endpoint measures in clinical trials: results of structured workshops. *J Rheumatol* 1982; 9: 798-801.
- Smith DG, Clemens J, Crede W, Harvey M, Gracely EJ. Impact of multiple comparisons in randomized clinical trials. *Am J Med* 1987; 83: 545-550.
- Emrich LJ. Common problems with statistical aspects of periodontal research papers. *J Periodontol* 1990; 61: 206-208.
- Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 1986; 292: 746-750.
- González JC, Pulido M, Sanz F. Evaluación del uso de procedimientos estadísticos en los artículos originales publicados en *Medicina Clínica* durante tres décadas. *Med Clin (Barc)* 1995; 12: 448-452.
- Der Simonian R, Charette J, McPeck B, Mosteller F. Reporting methods in clinical trials. *N Engl J Med* 1982; 306: 1.332-1.337.
- Powe NR, Kinnison ML, Steinberg EP. Quality assessment of randomized controlled trials of contrast media. *Radiology* 1989; 170: 377-380.
- Tyson JE, Furzan JA, Reisch JS, Mize SG. An evaluation of the quality of therapeutic studies in perinatal medicine. *J Pediatrics* 1983; 102: 10-13.
- Jiménez Villa J, Carré Llopis MC, Argimón Pallás JM. Información sobre los aspectos metodológicos en los artículos de la revista *Atención Primaria*. *Aten Primaria* 1994; 14: 1.113-1.117.
- Evans M, Pollock AV. Trials on trials. A review of trials of antibiotic prophylaxis. *Arch Surg* 1984; 119: 109-113.
- Mora Ripoll R, Ascaso Terrén C, Sentís Vilalta J. Tendencias actuales en la utilización en la estadística en medicina. Estudios de artículos originales publicados en *Medicina Clínica (1991-1992)*. *Med Clin (Barc)* 1995; 105: 444-447.
- Weiss ST, Samet JM. An assessment of physician knowledge of epidemiology and biostatistics. *Am J Med Educ* 1980; 55: 692-697.
- Haynes SJ. Randomized clinical trials in neurosurgery. *Neurosurgery* 1983; 12: 259-264.
- Reznick RK, Dawson-Saunders E, Folse Jr. A rationale for the teaching of stistics to surgical residents. *Surgery* 1987; 101: 611-617.
- Pérez Mateo M, Boix V, Martínez A, Latour J. Análisis crítico de los ensayos clínicos en pancreatitis aguda publicados entre 1983 y 1989. *Med Clin (Barc)* 1991; 96: 371-373.

33. Mora Ripoll R, Ascaso Terren C, Sentís Vilalta J. Uso y presentación de la metodología estadística en los artículos originales publicados en Medicina Clínica durante 1993. *Med Clin (Barc)* 1995; 105: 9-12.
34. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; 4: 942-951.
35. Jiménez Villa J, Carré Llopis MC, Argimón JM. Tipos de estudios publicados en la revista Atención Primaria (1984-1990). *Aten Primaria* 1993; 12: 9-14.
36. Reed III F, Slaichert W. Statistical proof in inconclusive "negative" trials. *Ann Intern Med* 1981; 141: 1.307-1.310.
37. Friedman JA, Chalmers TC, Smith H, Kuebler RR. The importance of Beta, the type II error and sample size in the design and interpretation of the randomized control trial. Surveys of 71 "negative" trials. *N Engl J Med* 1978; 299: 690-694.