

Estimación de la variabilidad inicial interobservador de la espirometría forzada en el estudio epidemiológico multicéntrico IBERPOC*

R. Gabriel Sánchez^a, C. Villasante Fernández-Montes^b, J.M. Pino García^b, F. García Río^b, M. Miravittles^c, C.A. Jiménez Ruiz^d y V. Sobradillo Peña^c por el Comité Científico del Estudio IBERPOC

^aUnidad de Epidemiología Clínica. Hospital Universitario de la Princesa. Madrid. ^bServicio de Neumología. Hospital La Paz. Madrid.

^cPharmaConsult Services, S.A. Barcelona. ^dServicio de Neumología. Hospital Universitario de la Princesa. Madrid.

^eServicio de Neumología. Hospital de Cruces. Bilbao.

El objetivo de este estudio ha sido determinar la variabilidad interobservador y el grado de concordancia en los parámetros espirométricos básicos previo a la realización del trabajo de campo del estudio IBERPOC.

Estudio de concordancia entre siete observadores (neumólogos) y un patrón de referencia, mediante esquema de bloques aleatorios balanceados incompletos con un número igual de mediciones espirométricas por paciente ($n = 3$) y observador ($n = 9$) durante la misma sesión. Población de estudio: 14 pacientes con distinto grado de obstrucción al flujo aéreo y 7 sujetos normales.

Se encontraron efectos estadísticamente significativos atribuibles a los sujetos (variabilidad entre pacientes) para las tres variables analizadas y variabilidad atribuible a la condición de observador para la FVC y FEV_1 pero no para el cociente FEV_1/FVC . Las mayores diferencias entre los observadores se encontraron en el FEV_1 , para el que 4 de los 7 observadores mostraron valores significativamente diferentes respecto a la media global ($p < 0,05$). Las diferencias fueron menos llamativas para el FVC y para el cociente FEV_1/FVC , con sólo 2 observadores con diferencias significativas en cada variable.

El alto grado de repetición de la prueba, así como la excelente concordancia conseguida entre los observadores tras la sesión de estandarización, garantizan a priori la validez de las estimaciones espirométricas y descartan la existencia de sesgos diferenciales de información entre las distintas áreas de estudio.

Palabras clave: *Espirometría. Variabilidad interobservador. Concordancia.*

Arch Bronconeumol 1997; 33: 300-305

Estimate of initial interobserver variability in forced spirometry for the IBERPOC multicenter epidemiological study

To estimate the interobserver variability and degree of agreement for basic spirometric parameters before beginning field work for the IBERPOC Project.

Study of agreement between 7 observers (pneumologists) and a gold standard, using a scheme of incomplete balanced and randomized blocks with an equal number of spirometric measurements ($n = 3$) per patient and an equal number of measurements ($n = 9$) per observer performed at the same session. The study population consisted of 14 patients with different degrees of air flow obstruction and 7 normal volunteers.

Statistically significant differences attributable to subjects (inter-patient variability) were found for the three variables analyzed. Variability attributable to the observer was found for FVC and FEV_1 but not for FEV_1/FVC . The greatest interobserver differences were found for FEV_1 , such that 4 of the 7 observers recorded values that were significantly different from the mean ($p < 0.05$). The differences were less marked for FVC and for the FEV_1/FVC ratio, with only 2 observers recording significantly different values for each variable.

The high degree of reproducibility as well as the excellent interobserver agreement found in this standardization session provide an *a priori* guarantee of validity for spirometric measurements and rule out the existence of differential bias in data recorded at the various geographic areas involved in the study.

Key words: *Spirometry. Interobserver variability. Agreement.*

*Institución patrocinadora: Boehringer Ingelheim España, S.A.

Correspondencia: Dr. R. Gabriel Sánchez.
Unidad de Epidemiología Clínica. Hospital Universitario de la Princesa.
Diego de León, 62. 28006 Madrid.

Recibido: 27-2-97; aceptado para su publicación: 4-3-97.

Introducción

El Consenso de la Sociedad Europea de Aparato Respiratorio¹ define operativamente el término enfermedad pulmonar obstructiva crónica (EPOC) como un síndrome caracterizado por la reducción del flujo aéreo máximo espiratorio y enlentecimiento del vaciamiento pul-

monar forzado de los pulmones, características que no cambian significativamente a lo largo de varios meses y en ausencia de asma u otras enfermedades respiratorias que pueden provocar obstrucción reversible¹. Para hablar de EPOC es necesario por tanto la existencia de un flujo espiratorio máximo reducido, mantenido en el tiempo y objetivado mediante espirometría forzada.

Fuera del ambiente clínico, son pocos los estudios epidemiológicos realizados hasta la fecha sobre prevalencia de EPOC en la comunidad basados en la utilización conjunta de criterios clínicos y espirométricos estandarizados². Por otro lado, resulta imprescindible la estimación del grado de precisión y de variabilidad en los índices espirométricos para garantizar la validez de las estimaciones que se obtengan en un estudio. Recientemente, Enright et al han mostrado cómo un programa intensivo de control de calidad tiene un impacto positivo en los resultados de la espirometría³.

En términos epidemiológicos, para la correcta interpretación de la espirometría es preciso clasificar los valores individuales obtenidos respecto a una población de referencia y asegurar la repetición del dato. La estimación de la variabilidad intrapaciente (tanto a corto como a largo plazo) es especialmente relevante en los estudios prospectivos y en los ensayos clínicos, donde se realizan medidas repetidas y donde la probabilidad de variaciones biológicas o cambios en ciertos factores asociados, tabaquismo por ejemplo, pueden modificar sustancialmente los parámetros espirométricos a lo largo del tiempo⁴.

Por otro lado, la variabilidad interobservador requiere una atención particular sobre todo en estudios multicéntricos con distintos técnicos en espirometría. A pesar de que este aspecto es reconocido como una importante fuente de error en las estimaciones, no existen datos publicados a este respecto procedentes de estudios epidemiológicos multicéntricos⁵.

Los objetivos generales del estudio IBERPOC son estimar la prevalencia específica de obstrucción al flujo aéreo mediante espirometría según edad, hábitat y sexo en la población general adulta entre 40 y 69 años de diferentes áreas geográficas españolas. Describir los valores de la distribución de la función pulmonar (FEV₁, FVC y cociente FEV₁/FVC) en la misma población por edad y sexo, y estimar su grado de asociación con la exposición al tabaco⁵.

El objetivo concreto de este trabajo es determinar la variabilidad interobservador y el grado de concordancia en los distintos parámetros espirométricos mencionados como paso previo a la realización del estudio de campo en los centros participantes (anexo 1).

Métodos

Se trata de un estudio de concordancia entre siete observadores (neumólogos) participantes en el estudio de campo y entre ellos y un patrón de referencia, para lo cual se diseñó un esquema de bloques aleatorios balanceados incompletos en el que cada observador realizó a un número igual al azar de participantes el mismo número de mediciones espirométricas⁶.

TABLA I
Parejas de pacientes (y sujetos normales) asignados de forma aleatoria a cada observador para examen

Pacientes (sujetos normales)	Observadores							PR	Media
	1	2	3	4	5	6	7		
1,8	a	b		c				a	
(1)	x	x			x			X1	
2,9		d	e		f			β	
(2)		x	x			x		X2	
3,10			g	h		i		g	
(3)				x	x		x	X3	
4,11				j	k		l	d	
(4)	x				x	x		X4	
5,12	m				n	o		e	
(5)		x				x	x	X5	
6,13		p				q	r	t	
(6)	x		x					X6	
7,14	s		t				u	p	
(7)	x	x		x				X7	

PR: patrón de referencia.

Las condiciones del modelo fueron⁷:

– Número de observadores (sin contar el patrón de referencia): $g = 7$.

– Número de pacientes o bloques de observación: $n = 14$.

– Número de pacientes examinados por cada observador: $r = 6$.

– Número de observadores que examinan a cada paciente: $k = 3$.

– Número de pacientes vistos por cada par de observadores: $\xi = 2$.

donde:

$$(1) \quad g \cdot r = n \cdot k \quad , \text{o sea,} \quad 7 \cdot 6 = 14 \cdot 3$$

$$(2) \quad g \leq n \quad \text{o} \quad 7 \leq 14$$

$$(3) \quad \xi(g-1) = r(k-1) \quad \text{o} \quad 2(7-1) = 6(3-1)$$

La población de estudio consistió en 14 pacientes con distinto grado de obstrucción espiratoria al flujo aéreo elegidos consecutivamente en un día concreto de la consulta externa de neumología de un hospital de tercer nivel y 7 personas (neumólogos participantes en el estudio) sin patología pulmonar conocida y función pulmonar normal.

La ejecución del proceso de estandarización siguió los pasos que se detallan a continuación. En primer lugar, se asignó aleatoriamente un número al azar (del 1 al 7) a cada observador. Cada paciente fue examinado consecutivamente y siguiendo el orden establecido por tres observadores (entre paréntesis el número de paciente):

(1) 1,2,4 (2) 2,3,5 (3) 3,4,6 (4) 4,5,7 (5) 5,6,1 (6) 6,7,2 (7) 7,1,3, (8) 1,2,4 (9) 2,3,5 (10) 3,4,6 (11) 4,5,7 (12) 5,6,1 (13) 6,7,2 (14) 7,1,3.

A su vez, el neumólogo que actuó como patrón de referencia realizó una espirometría a cada uno de los 14 pacientes y a cada uno de los observadores (total: 21 espirometrías), para poder comparar con los resultados de los observadores.

Este modelo (cada paciente es examinado en tres ocasiones) permite, así mismo, obtener un estimador medio adecuado (basado en tres pruebas) para cada uno de los parámetros espirométricos considerados en cada paciente. Esta estrategia permite además estimar el grado de repetición de cada parámetro y corregir potenciales sesgos de dilución de la regresión debidos a mediciones únicas del FEV₁ en los análisis de asociación que se realicen en el estudio real.

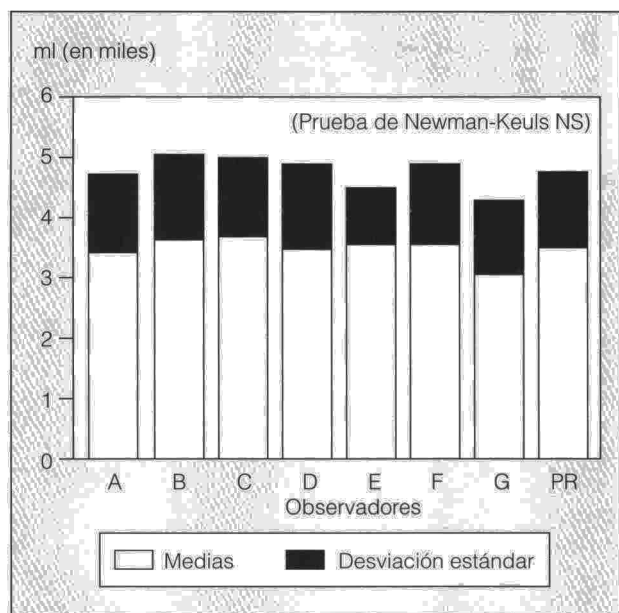


Fig. 1. Valores medios y desviación estándar de FVC obtenidos por cada observador (n = 9). PR: patrón de referencia.

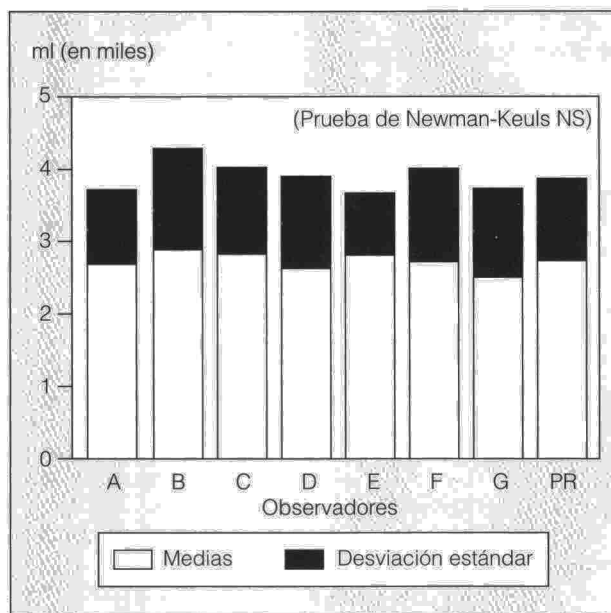


Fig. 2. Valores medios y desviación estándar de FEV₁ obtenidos por cada observador (n = 9). PR: patrón de referencia.

Para calcular el grado de variabilidad interobservador de los parámetros espirométricos básicos en el rango de la normalidad (en sujetos sin patología respiratoria), se complementó el estudio utilizando un diseño similar al anterior en el que cada uno de los observadores participantes realizó una espirometría a tres de sus compañeros, y el "patrón de referencia" (PR) una a cada uno de ellos, según el esquema que se muestra en la tabla I.

Cada observador utilizó un único espirómetro, el mismo a utilizar durante el trabajo de campo. Los siete espirómetros eran portátiles del modelo DATOSPIR-200 Sibel, S.A.

Para aumentar la precisión y disminuir al máximo posible la variabilidad intraindividual del FEV₁ y FVC se siguieron todas las recomendaciones de la normativa SEPAR⁸ para la realización de la espirometría estandarizada, que se describen en detalle en el artículo del Comité Científico del Estudio IBERPOC⁵. En resumen, la espiración forzada se realizó con el paciente sentado y con la nariz ocluida por una pinza; la maniobra consistió en una espiración máxima desde la posición de inspiración máxima, con expulsión violenta y forzada al máximo de todo el aire contenido en los pulmones hasta alcanzar su volumen residual; se realizaron un mínimo de tres maniobras satisfactorias y se eligió como más representativa

la de mayor capacidad vital y FEV₁. Además, se vigiló que el esfuerzo hubiera sido máximo y que no se produjera tos ni maniobra de Valsalva por cierre de glotis.

Análisis estadístico

Se han calculado las medias globales por paciente (obtenidas por todos los observadores juntos) y para cada bloque de pacientes (n = 9) por observador para el FVC, FEV₁ y cociente FEV₁/FVC. La comparación de medias entre observadores (todas las combinaciones posibles) se ha realizado mediante la prueba de Newman-Keuls, descartándose la hipótesis nula de igualdad con un nivel de significación estadística p < 0,05. Posteriormente, se ha efectuado un análisis de la varianza (ANOVA) para dos factores (variabilidad interpacientes y variabilidad interobservadores) y se han calculado los correspondientes coeficientes de correlación intraclass (CCI) para cada una de las tres variables analizadas. Todos los cálculos han sido realizados con el paquete estadístico PRESTA⁹.

TABLA II
Valores medios de FVC, FEV₁ y cociente FEV₁/FVC obtenidos por cada observador sobre los 9 sujetos (3 pacientes y 6 normales) examinados y diferencias (efectos) respecto al valor medio global

Observador	FVC			FEV ₁			FEV ₁ /FVC		
	\bar{X}	Diferencia	p	\bar{X}	Diferencia	p	\bar{X}	Diferencia	p
1	3.461	-38	0,202	2.678	-25	0,65	76,7	0,49	0,563
2	3.655	110	0,001	2.912	56	0,038	75,4	-0,75	0,627
3	3.683	54	0,074	2.836	60	0,027	75,4	0,74	0,621
4	3.463	48	0,112	2.617	41	0,121	73,2	-0,17	0,827
5	3.551	-24	0,572	2.801	-95	0,001	77,7	-2,1	0,015
6	3.533	-116	0,001	2.692	-87	0,002	71,3	-0,73	0,615
7	3.046	-33	0,272	2.485	49	0,065	77,9	2,52	0,004

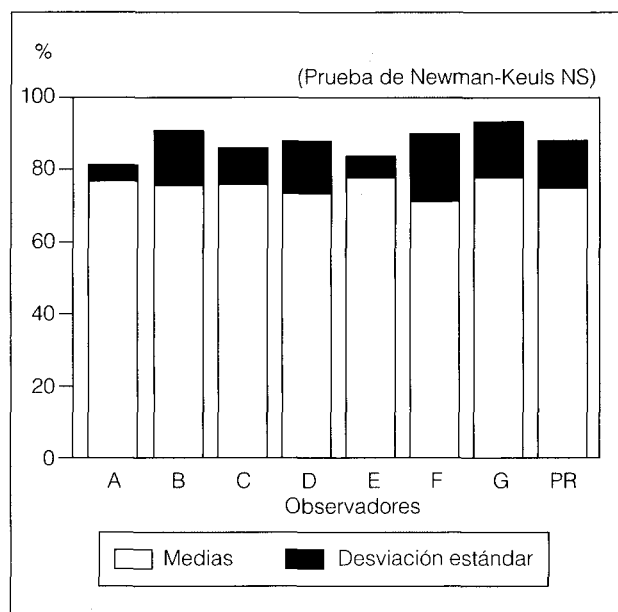


Fig. 3. Valores medios y desviación estándar de FVE₁/FVC obtenidos por cada observador (n = 9). PR: patrón de referencia.

Resultados

En la tabla I se muestra el esquema de asignación de parejas de pacientes a cada observador (bloque) utilizado para el cálculo de las medias en el modelo descrito.

En la tabla II se presentan los valores medios y diferencias (efectos) del FVC, FEV₁ y cociente entre cada observador y el valor medio global. Las mayores diferencias entre los valores medios observados y los globales se encontraron para el FEV₁, para el que 4 de los 7 observadores mostraron valores significativamente diferentes (efectos significativos, p < 0,05). Sin embargo, las diferencias fueron menos llamativas para el FVC y para el cociente FEV₁/FVC, con sólo 2 observadores con diferencias significativas en cada variable.

En las figuras 1-3 se muestran gráficamente los valores obtenidos por cada observador para cada una de las tres medidas. No se han encontrado diferencias estadísticamente significativas entre ningún observador para ninguna de las tres variables analizadas. Del mismo modo, el coeficiente de variación (CV%) intraindividual global considerando todas las mediciones realizadas por cada observador fue del 4%, no superándose en ningún caso en términos absolutos diferencias mayores a 200 ml (límite máximo aceptado) entre las dos mediciones más altas para el FEV₁ en un mismo sujeto.

En la tabla III se resumen los principales hallazgos del análisis de la varianza para las tres variables consideradas. Para el cociente FEV₁/FVC no se encontraron efectos estadísticamente significativos atribuibles a los observadores (variabilidad entre espirometristas), aunque la variabilidad atribuible a la condición de observador fue significativa cuando estas variables (FVC y FEV₁) se consideraron aisladamente. Los coeficientes de correlación intraclase (CCI) hallados para los distin-

TABLA III
ANOVA para dos factores (observadores y sujetos) y coeficientes de correlación intraclase para las principales variables espirométricas al inicio del estudio

Variable	Efectos observadores		Efectos sujetos		CCI
	F	p	F	p	
FVC	5,44	0,0004	627,03	0,00000	0,993
FEV ₁	5,75	0,0003	687,17	0,00000	0,993
FEV ₁ /FVC	2,65	0,0309	78,05	0,00000	0,958

CCI: coeficiente de correlación intraclase.

tos pares de observadores fueron, respectivamente, 0,99 para el FVC y FVEI y 0,958 para el cociente FEV₁/FVC.

Discusión

La medición de la imprecisión se considera en general como un fenómeno aleatorio en el que se acepta que los estimadores obtenidos se distribuyen alrededor del parámetro real del sujeto. Las desviaciones de su valor real se asume que son independientes y se distribuyen de forma idéntica dentro del mismo sujeto que entre sujetos con media cero e igual varianza. El contraste de la homogeneidad de la varianza consiste en comprobar si ésta aumenta cuando lo hace el parámetro real.

En nuestro estudio se observó la existencia de variabilidad interobservador significativa, fundamentalmente para el FVC y FEV₁. Esta variabilidad podría ser debida en parte a las condiciones del modelo utilizado, que no garantiza al 100% una distribución aleatoria de los pacientes entre observadores, al escaso número de sujetos analizados por observador (n = 6) y a la gran variabilidad inherente a los pacientes en comparación con la atribuible a los observadores. No obstante, esta variabilidad interobservador tiende a disminuir hasta situarse en el límite de la significación estadística (F = 2,65; p = 0,03) cuando se analiza el cociente FEV₁/FVC, un estimador más preciso de obstrucción al flujo aéreo, en lugar de sus componentes por separado.

Sin embargo, los excelentes coeficientes de correlación intraclase (CCI) obtenidos en los tres parámetros considerados para cada bloque de observadores (k) y por paciente (tanto en los casos con EPOC como en los sujetos normales) garantizan la fiabilidad de la concordancia en las mediciones que se realicen en el futuro.

Una limitación importante de los estudios transversales es la imposibilidad de estimar la variabilidad intraindividual (en diferentes días, en distintas condiciones). En la bibliografía, los datos procedentes de estudios epidemiológicos similares al presente muestran que el coeficiente de variación intraindividual (CV%) para el FEV₁ es mucho menor entre mediciones realizadas en la misma sesión (menos de 15 o 30 min) que tras largos intervalos de tiempo (meses o años) debido al descenso paulatino de la capacidad pulmonar con el tiempo y/o al cambio en múltiples factores externos. Para estudios epidemiológicos a gran escala la ATS acepta que un buen índice de reproducibilidad para el FEV₁ es una di-

ANEXO 1

Organización del proyecto

<p>1. Institución organizadora Sociedad Española de Neumología y Cirugía Torácica (SEPAR). Área de Trabajo sobre Insuficiencia Respiratoria y Trastornos del Sueño (IRTS).</p>	<p>Hospital General de Vic. Barcelona. Hospital de Cruces. Bilbao. Unidad de Epidemiología Clínica. Hospital Universitario de la Princesa (centro de análisis de datos). PharmaConsult Services, S.A. Barcelona (Oficina del Proyecto).</p>
<p>2. Institución patrocinadora Boehringer Ingelheim, S.A.</p>	<p>5. Instituciones colaboradoras Sibel, S.A. Unidad de Investigación en Servicios Sanitarios IMIM. Barcelona Soikos S.L.</p>
<p>3. Comité científico <i>SEPAR</i> Fernández Fau L. Servicio de Cirugía Torácica. Hospital Universitario de la Princesa. Madrid</p> <p>Sobradillo Peña V. (coordinador). Servicio de Neumología. Hospital de Cruces. Baracaldo (Vizcaya).</p> <p>Jiménez Ruiz C. Servicio de Neumología. Hospital Universitario de la Princesa. Madrid.</p> <p>Masa Jiménez JF. Servicio de Neumología. Hospital San Pedro de Alcántara. Cáceres.</p> <p>Viejo Bañuelos JL. Sección de Neumología. Hospital General Yagüe. Burgos.</p> <p>Villasante Fernández-Montes C. Servicio de Neumología. Hospital La Paz. Madrid.</p> <p><i>Asociación Española para el Desarrollo de la Epidemiología Clínica (AEDEC)</i> Gabriel Sánchez R. Unidad de Epidemiología Clínica. Hospital Universitario de la Princesa. Madrid.</p> <p><i>Atención primaria</i> Baos Vicente V. Centro de Salud de Villalba. Madrid.</p> <p>Álvarez Guisasola F. Centro de Salud de la Calzada. Gijón.</p> <p>Salgado Saez JI. Centro de Salud Usánsolo. Galdácano (Vizcaya).</p> <p>Simonet Aineto PJ. Centro de Salud ABS Viladecans 2. Barcelona.</p>	<p>6. Responsables e investigadores locales del estudio <i>Asturias</i> Coordinador SEPAR: Dr. Jaime Martínez González del Río. Servicio de Neumología. Hospital de Asturias (Oviedo). Investigador: Dr. José Antonio Gullón Blanco.</p> <p><i>Burgos</i> Coordinador SEPAR: Dr. Jose Luis Viejo Bañuelos. Investigador: Dra. Lourdes Lázaro Asegurado.</p> <p><i>Cáceres</i> Coordinador SEPAR: Dr. Juan Fernando Masa Jiménez. Investigador: Dra. Lourdes Ramos Casado.</p> <p><i>Madrid</i> Coordinador SEPAR: Dr. Carlos A. Jiménez Ruiz. Dr. Carlos Villasante Fernández-Montes. Investigador: Dr. Ali Dorgham.</p> <p><i>Sevilla</i> Coordinador SEPAR: Dr. José Castillo Gómez. Servicio de Neumología. Hospital Virgen del Rocío. Investigador: Dr. José Fernández Guerra.</p> <p><i>Manlleu</i> Coordinador SEPAR: Dr. Joan Serra Batlles. Sección de Neumología. Hospital General de Vic. Investigador: Dr. Joan Casadevall Escayola.</p> <p><i>Vizcaya</i> Coordinador SEPAR: Dr. Víctor Sobradillo Peña. Investigador: Dra. Pilar Gil Alaña.</p>
<p>4. Instituciones participantes Hospital General de Asturias. Oviedo. Hospital General Yagüe. Burgos. Hospital S. Pedro de Alcántara. Cáceres. Hospital La Paz. Madrid. Hospital Universitario de la Princesa. Madrid. Hospital Universitario Virgen del Rocío. Sevilla.</p>	<p>7. Coordinador del trabajo de campo Dr. Marc Miravittles. PharmaConsult Services, S.A. Barcelona.</p>

ferencia no superior a 200 ml entre la segunda medida más alta o un CV% menor al 5%. En nuestro estudio el CV% intraindividual global considerando todas las mediciones realizadas fue del 4% y en ningún caso se superó la diferencia de los 200 ml, límite máximo aceptado en términos absolutos³. No obstante, en el estudio de campo se comprobará la espirometría de forma ciega (sin conocer el resultado obtenido por el observador) en el 10% de los participantes. Esta estrategia permitirá corregir los valores que se obtengan para cada variable según su grado de repetición y controlar posibles sesgos de dilución de la regresión entre las variables independientes y niveles de FEV₁.

En conclusión, el alto grado de repetición de la prueba, así como la excelente concordancia conseguida entre los observadores tras la sesión de estandarización, garantizan a priori la validez de las estimaciones espirométricas y descartan la existencia de sesgos diferenciales de información entre las distintas áreas de estudio.

BIBLIOGRAFÍA

1. Siafakas NM, Vermeire NB, Pride P, Paoletti J, Gibson P, Howard P et al. Optimal assessment and management of chronic obstructive pulmonary disease (COPD). A consensus statement of the European Respiratory Society (ERS). *Eur Respir J* 1995; 8: 1.398-1.420.
2. Manfreda J, Mas Y, Litven W. Morbidity and mortality from chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1989; 140: 19-26.
3. Enright EP, Johnson LR, Connett JE, Voelker H, Buist S. Spirometry in the Lung Health Study. 1 Methods and Quality Control. *Am Rev Respir Dis* 1991; 143: 1.215-1.223.
4. Camilli AE, Burrows B, Knudson RJ, Lyle SK, Lebowitz MD. Longitudinal changes in FEV₁ in adults. Effects of smoking and smoking cessation. *Am Rev Respir Dis* 1987; 135: 794-799.
5. Comité Científico del Estudio IBERPOC. Proyecto IBERPOC: un estudio epidemiológico de la EPOC en España. *Arch Bronco Neumol* 1997; 33: 293-299.
6. Joseph L. Fleiss. *The Design and analysis of clinical experiments*. Nueva York: John Wiley and Sons, 1986.
7. Cochran WG, Cox GM. *Experimental designs* (2.ª ed.). Nueva York: John Wiley and Sons, 1957.
8. Sanchis J y Grupo de Trabajo de la SEPAR para la práctica de la espirometría en clínica. Normativa para la espirometría forzada. *Arch Bronconeumol* 1989; 25: 132-142.
9. Abraira V, Zaplana J. PRESTA paquete de procesamientos estadísticos. Versión 2.2 para PC. Madrid: Fondo de Investigación Sanitaria, 1992.