

Interpretación de la prueba espirométrica

Traducción
Santiago C. Arce

Philip H. Quanjer
Sanja Stanojevic
Janet Stocks
Tim J. Cole

Introducción

Durante los 6 años que llevó a la Global Lung Function Initiative (GLI) finalizar su misión, con el apoyo de seis grandes sociedades respiratorias internacionales, una red colaborativa fue establecida alrededor del mundo. Esta red incluyó médicos asistenciales, investigadores, técnicos, ingenieros informáticos y fabricantes. El objetivo era derivar ecuaciones de referencia para espirometría que cubrieran tantos grupos étnicos como fuera posible, y abarcara un rango de edad desde niños en edad pre-escolar hasta la vejez. Gracias a una cooperación internacional sin precedentes, decenas de miles de mediciones espirométricas de varones y mujeres sanos no fumadores, fueron puestos a disposición por alrededor de 70 centros y organizaciones. Estos datos fueron cotejados y analizados con técnicas estadísticas modernas, y condujeron a las ecuaciones de predicción GLI-2012. Este manuscrito resume los principales resultados previamente presentados en reuniones internacionales y en publicaciones impresas.

Perspectiva histórica

Pasó un largo tiempo desde la introducción del espirómetro por Hutchinson en 1846 [1] hasta su aplicación clínica. Inicialmente, las mediciones estaban limitadas a la determinación de la capacidad "vital" (VC), es decir, la capacidad vital lenta espiratoria (EVC) según la terminología actual. La figura 1 ilustra la subdivisión de la capacidad pulmonar total en EVC y volumen residual en la publicación original de Hutchinson. Tuvo que pasar un siglo hasta que los investigadores franceses Tiffeneau y Pinelli [2] transformaron las mediciones espirométricas a su forma actual, en que el volumen espiratorio forzado en 1 segundo (FEV₁) y la capacidad vital inspiratoria (IVC) o forzada espiratoria (FVC) se tornaron índices diagnósticos pivotaes en la medicina clínica. Yernault resumió concisamente la historia de las mediciones espirométricas en una publicación clara y accesible [3].

Los resultados de la prueba espirométrica están significativamente influenciados por la cooperación del sujeto y se ven afectados por factores técnicos; en consecuencia, las mediciones deben ser administradas de acuerdo a un estricto protocolo. En 1960, la Comunidad Europea para el Carbón y el Acero (ECCS, por sus siglas en inglés) fue la primera organización en emitir recomendaciones [4]. Estas fueron actualizadas en 1971 [5], incluyendo valores predichos para los índices espirométricos, volumen residual,

DIAGRAM 1.

The division of the thoracic movements.

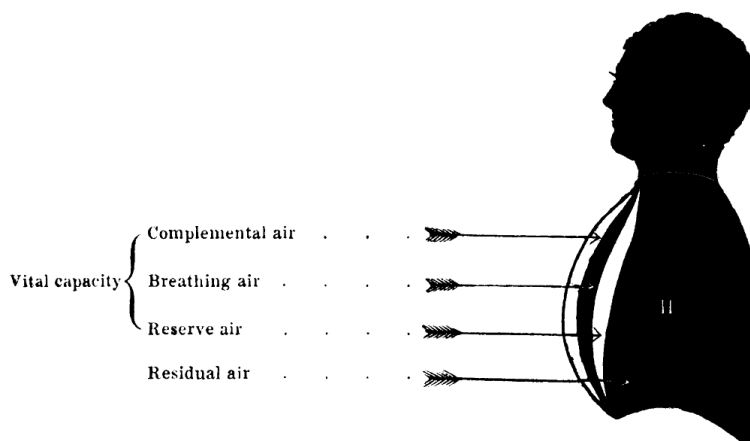


Fig. 1 - Subdivisiones de la capacidad pulmonar total según Hutchinson (1846).

capacidad pulmonar total y capacidad residual funcional. Unos años más tarde, en Estados Unidos, se hicieron los primeros esfuerzos para su estandarización, limitados inicialmente a la espirometría dentro del ámbito epidemiológico [6-7]. Debido a los rápidos desarrollos tecnológicos, un mayor conocimiento de la fisiopatología de las enfermedades pulmonares y un mayor arsenal de pruebas funcionales respiratorias, se impuso la necesidad de revisar el reporte de la ECCS [8]. Desde entonces se emitieron sucesivas revisiones sobre estandarización en Estados Unidos y Europa [9]; los reportes americanos estuvieron enfocados exclusivamente en la espirometría, en tanto las recomendaciones europeas cubrieron un rango más amplio de pruebas funcionales respiratorias e incluyeron, invariablemente, valores de referencia recomendados [9-11].

Valores de referencia para espirometría

El conjunto de valores de referencia publicado por la ECCS [4-5] se basaba en varones trabajando en minas de carbón y acerías. Esta no era una población de referencia representativa y, en la práctica, los valores de referencia eran considerados demasiado altos. Aún cuando no se había estudiado a ninguna mujer, la ECCS emitió valores de referencia para mujeres: se calculaban como el 80% del valor de los hombres. En 1983 la ECCS declinó otorgar fondos para un estudio poblacional que permitiera derivar valores de referencia obtenidos con métodos que cumplieran con los más recientes estándares. Con la visión de combinar recomendaciones técnicas con ecuaciones de predicción apropiadas, y dado que no existía material que hubiese sido obtenido mediante técnicas apropiadas y ante la falta de mejores alternativas, el comité de estandarización decidió adoptar

la técnica previamente utilizada por Polgar [12] al derivar sus ecuaciones de referencia para niños. Esto implicaba la generación de un conjunto de valores predichos por edad, talla y sexo usando las ecuaciones de predicción publicadas y, a su vez, usar este conjunto artificialmente generado para derivar nuevas ecuaciones de regresión. Serías objeciones podrían ser realizadas contra este procedimiento, pero las ecuaciones de regresión resultantes fueron aceptadas con escasas críticas y ampliamente adoptadas.

Una alternativa que el grupo de estandarización de la ECCS ha acogido como una buena alternativa para un nuevo estudio poblacional fue derivar nuevas ecuaciones a partir de mediciones de comprobada buena calidad, que cumplieran con las recomendaciones contemporáneas; dichos datos no estaban disponibles. El primer uso de conjuntos de datos cotejados para derivar valores predichos para niños se basaba en 6 bases de datos provenientes de 5 países europeos [13]. Este estudio demostró que los valores de referencia resultantes ajustaban con 5 de los 6 conjuntos de datos; esto evidenciaba que el sexto conjunto había sido afectado por problemas técnicos. De esa manera, este enfoque fue validado y llevó a la American Thoracic Society (ATS) y la European Respiratory Society (ERS) a recomendar esta técnica para la derivación de valores de referencia basados en grandes grupos poblacionales con un amplio rango de edad [14].

En 2005 la tradición europea de combinar reportes de estandarización con conjuntos de valores predichos recomendados llegó a su fin: un comité conjunto de ATS y ERS [5] recomendó valores predichos para los Estados Unidos y Canadá, dejando al resto del mundo sin cobertura. En 2006, uno de nosotros (PHQ) comenzó a remediar esta de-

ficiencia, con el objetivo de cubrir un rango de edad tan amplio como fuera posible, así como varios grupos étnicos. En 2008, más de 30.000 registros provenientes de todo el mundo habían sido puestos a disposición, y un manuscrito estaba en preparación, pero esto fue suspendido porque se fundó un grupo de trabajo de la ERS con los mismos objetivos. Este grupo posteriormente adquirió el status de “Task Force” de la ERS en 2010, y consiguió el apoyo de 6 grandes sociedades internacionales [15]. El 2008 fue también el año de la innovadora publicación de Stanojevic *et al.* [16], aplicando una nueva y muy poderosa técnica estadística sobre datos espirométricos cotejados provenientes de sujetos blancos entre 3 y 80 años de edad.

El trabajo colaborativo en el grupo, que fue llamado “Global Lung Function Initiative” (Iniciativa Global para la Función Pulmonar, en español) [15], fue un privilegio gracias a la cooperación efectiva y cordial, basada en el respeto mutuo y la confianza, con alrededor de 70 grupos de todo el mundo. El trabajo analítico fue realizado por el “Equipo Analítico” (fig. 2).

Situación en 2006

La comparación de los valores predichos de FEV₁ en hombres blancos según 30 autores diferentes revela una imagen preocupante (fig. 3). Para la misma talla y edad, los valores predichos pueden diferir en 1 litro o más. Los valores predichos para niños y adolescentes no tienen continuidad con los de los adultos. ¡Estas ecuaciones de predicción fueron usadas en muchas partes del mundo para propósitos diagnósticos! Un preocupante estado de las cosas.



Fig. 2 - El Equipo Analítico de la Global Lung Function Initiative. De izquierda a derecha: Prof. Tim Cole, Prof. Janet Stocks, Prof. Philip Quanjer, Dr. Sanja Stanojevic.

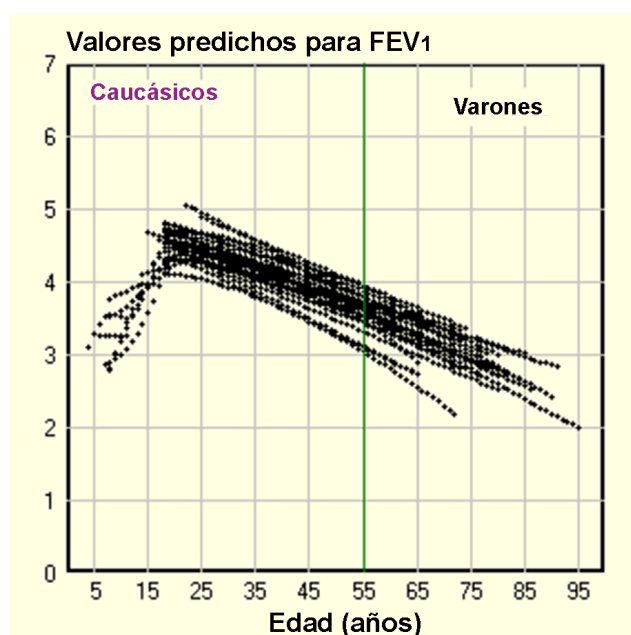


Fig. 3 - Valores predichos de FEV1 en varones blancos. Derivado del software descargable de www.spirxpert.com/GOLD.html.

Modelando la función pulmonar

Hasta recientemente, las ecuaciones de regresión para predecir la función pulmonar se basaban en técnicas de regresión lineal simple aditivas. Los modelos más populares tenían la siguiente forma:

$$Y = a + b \cdot \text{talla} + c \cdot \text{edad} + \text{error (adultos)}$$

$$\log(Y) = a + b \cdot \log(\text{talla}) + \text{error (niño)}$$

“Y” es el valor predicho, por ejemplo, FEV1. El “error”, también llamado residuo, es la diferencia entre los valores medido y predicho. Para niños y adolescentes los índices reciben usualmente una transformación logarítmica, y la edad es raramente considerada. Cuando se utilizan los modelos lineales arriba presentados, habitualmente se asume que los residuos son los mismos a cualquier combinación de edad y talla.

La figura 4 muestra FEV1 en función de la edad en un gran número de mujeres sanas entre 3 y 95 años de edad, e ilustra algunos aspectos:

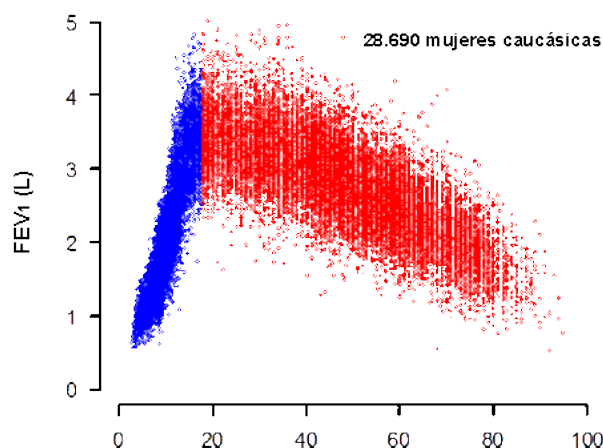


Fig. 4 - Relación entre la edad y el FEV1 en 28.690 mujeres blancas sanas. Alrededor de la mitad de la dispersión es debida a diferencias en la talla.

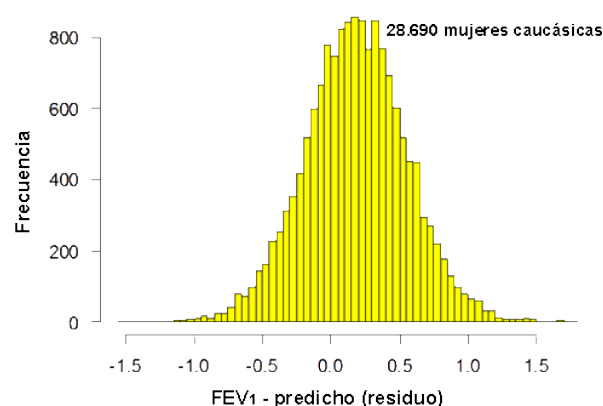


Fig. 5 - Diferencia entre FEV1 medido y predicho en mujeres blancas sanas al usar las ecuaciones de predicción ECCS/ERS.

- 1 La relación no puede ser representada por líneas rectas.
- 2 La dispersión (“error”) no es constante.
- 3 La dispersión no es proporcional al valor predicho.

Podemos calcular los valores predichos de FEV1 para las mujeres de la figura 4 usando las ampliamente difundidas ecuaciones de predicción de ECCS/ERS [10]. La diferencia media entre los valores medidos y predichos de FEV1 debería ser 0 si la ecuación ajustara a los datos perfectamente. La figura 5 muestra que hay una diferencia sistemática: el FEV1 medido en, es promedio, 180 ml mayor que el predicho. Los valores predichos por ECCS/ERS son, por lo tanto, sistemáticamente bajos.

Esta breve introducción lleva a las siguientes conclusiones:

- 1 La separación de niños/adolescentes de adultos es artificial y lleva a discontinuidades en los valores predichos en la transición de la adolescencia a la adultez.
- 2 Los modelos ajustan pobremente a los valores medidos, especialmente en niños.
- 3 Las diferencias en los valores predichos por varios autores son muy grandes.

El uso del porcentaje del predicho

Al interpretar datos espirométricos, un hábito arraigado en medicina respiratoria es expresar los valores medidos

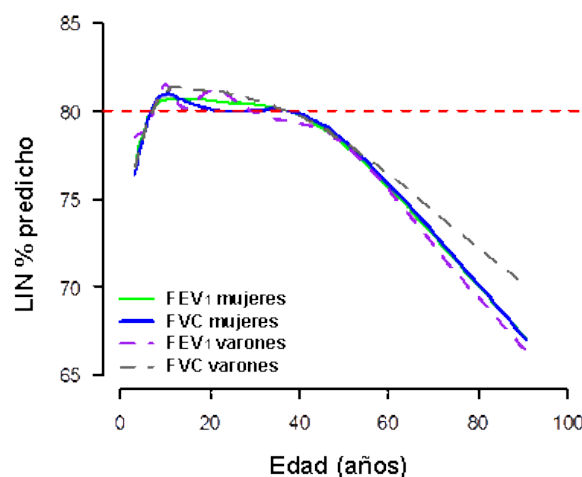


Fig. 6 - El límite inferior de la normalidad (LIN) para FEV1 y FVC expresado como porcentaje de los valores predichos por GLI-2012 en el rango de edad de 3-95 años.

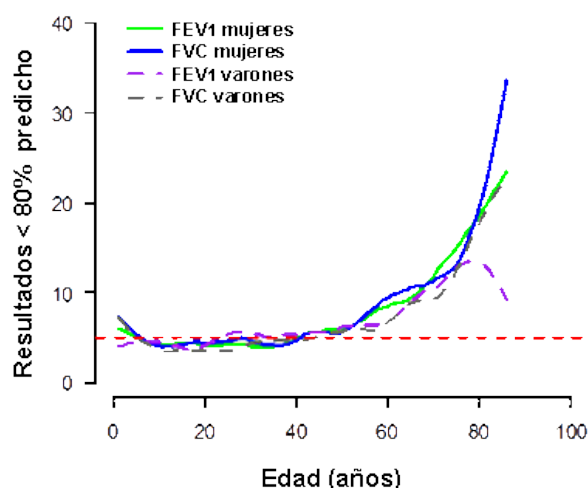


Fig. 7 - Porcentaje de varones y mujeres sanos en quienes el FEV1 o FVC medidos son <80% del predicho.

como porcentaje del predicho. Esta tradición fue originalmente propuesta por Gramprey [17] y más tarde adoptada por Bates y Christie [18]: “una regla general útil es que un desvío del 20% respecto al valor normal predicho es probablemente significativo”. Esto lleva a considerar al 80% del predicho como el “límite inferior de la normalidad” (LIN). Esta regla de oro fue adoptada acríticamente. La regla es válida sólo si la dispersión alrededor del valor predicho es proporcional a ese valor; en consecuencia, es grande si el valor predicho es grande, y proporcionalmente más pequeña si el valor predicho es menor. Como se ve en la figura 4, no existe proporcionalidad, de modo que el uso del porcentaje del predicho llevará inevitablemente a una interpretación errónea de los resultados, como ha sido explicado en decenas de publicaciones [10,16,19-24]. De hecho, Sobol escribió [20]: “En ningún otro lado en la medicina existe una visión tan ingenua del límite de la normalidad”. Dado que el grupo GLI tenía decenas de miles de datos disponibles, esto proveyó una oportunidad de estimar el límite inferior de la normalidad más precisamente (ver luego). Expresar el límite inferior de la normalidad como % del predicho lleva a la imagen de la figura 6: a lo largo de un amplio rango de

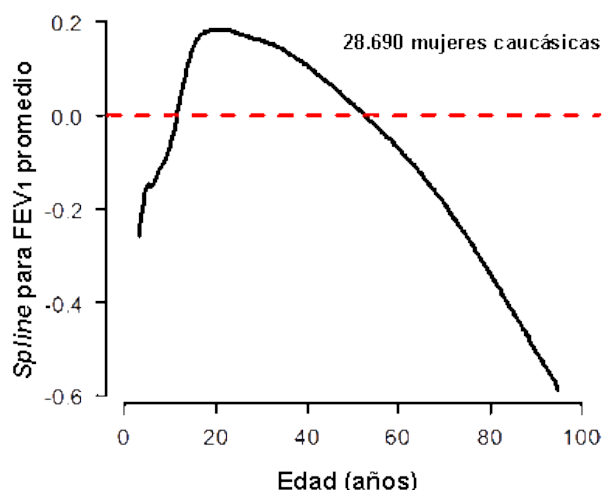


Fig. 8 - El spline, que agrega un término específico según la edad al valor predicho. Nótese que las ecuaciones de predicción usan una escala logarítmica.

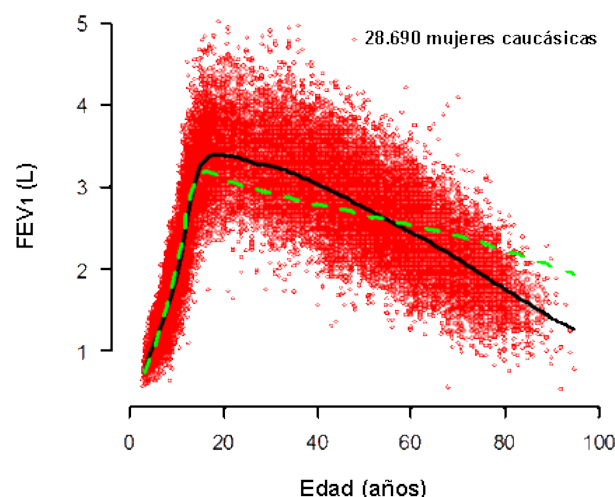


Fig. 9 - El FEV1 predicho sin el uso de la función de *spline* (línea verde-amarillenta) provee un mal ajuste; la que incluye el *spline* (línea negra) ajusta bien.

edad el límite inferior de la normalidad se encuentra muy por debajo de la línea del 80% del predicho (fig. 7).

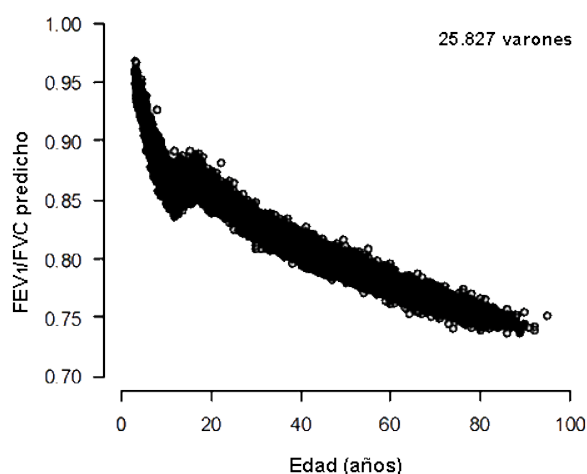
Global Lung Function Initiative: que hay de nuevo??

Expresar la relación no lineal de los índices espirométricos con la edad y talla, usando técnicas estándar de regresión lineal, no es posible. Ocasionalmente, esta situación era resuelta mediante la separación del rango de edad en dos partes: adultos y niños/adolescentes, y derivando dos conjuntos de ecuaciones que coincidieran en sus extremos, como en Hankinson *et al.* [25]. Antes de eso, la función pulmonar en la infancia era expresada mediante un modelo más complejo [13], o por el uso de una gran cantidad de ecuaciones de regresión, cada una abarcando un año [26]. Modelos más sofisticados fueron usados de manera similar en el rango adulto, prestando especial atención a definir exactamente el límite inferior de la normalidad [27-28]. Un método elegante para expresar curvas no lineales es el agregado de un *spline* a una regresión lineal:

$$\log(Y) = a + b \cdot \log(\text{talla}) + c \cdot \log(\text{edad}) + \text{spline} + \text{error}$$

Este enfoque fue adoptado por Pistelli *et al.* [29-30]. Sin embargo, el paquete estadístico GAMLSS [31], usado por primera vez con este propósito por Stanojevic *et al.* [16], ofrece métodos más avanzados para modelar la función pulmonar. En la práctica, este suavizamiento de la curva es modelado como una función de la edad. Es mejor imaginarlo como un ajuste específico para la edad del valor predicho: una corrección que varía con la edad a lo largo del rango de 3 a 95 años (figura 8). Esto opera en una escala logarítmica e implica, por ejemplo en una mujer de 20 años, que el FEV1 predicho calculado usando los coeficientes lineales (a , b y c en la ecuación de más arriba) debería ser multiplicado por $\exp(0,19) = 1,21$; es decir, un incremento del 21%. En una mujer de 85 años, multiplicamos por $\exp(-0,40) = 0,67$; y esto corrige al FEV1 en un 33%.

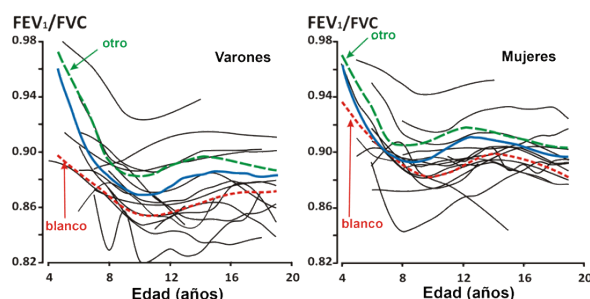
La diferencia entre el valor predicho sin y con el suavizamiento es ilustrada en la figura p. La línea punteada repre-

Fig. 10 - FEV₁/FVC predicho en varones blancos.

senta el valor predicho sin el *spline*. En niños y adolescentes, ajusta de manera aceptable, pero en adultos, el ajuste es deficiente. Por el contrario, la línea negra, que representa el valor predicho al agregar la función de *spline*, ajusta a los valores reales a lo largo de todo el rango de edad.

FEV₁/FVC: una sorpresa

El análisis del cociente FEV₁/FVC derivó en un resultado inesperado. El valor predicho decayó rápidamente entre los 3 y aproximadamente 10 años de edad, seguido por un pe-

Fig. 11 - Los datos de 15 centros, abarcando diferentes grupos étnicos, generalmente exhiben el mismo patrón: declinación rápida del cociente FEV₁/FVC hasta el comienzo de la adolescencia, luego un pequeño crecimiento seguido por una declinación.

queño incremento hasta los 16 años, y luego una gradual declinación no lineal en adultos (figura 10). Como este patrón nunca había sido descrito antes, el primer pensamiento fue que estábamos lidiando con un artefacto proveniente de la recopilación de tantos conjuntos de datos. Después de todo, si un centro hubiera aportado datos con un cociente FEV₁/FVC inusualmente bajo alrededor de los

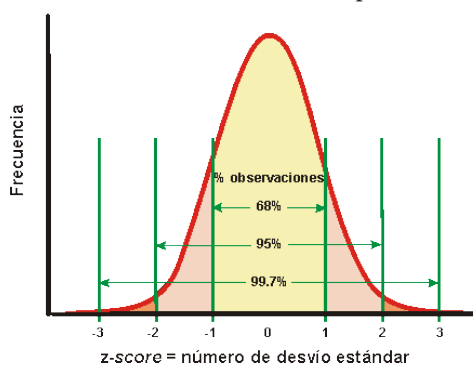


Fig. 12 - Relación entre el desvío estándar y el porcentaje de datos bajo la curva en una muestra con distribución normal

10 años de edad, esto podría haber explicado los hallazgos. Sin embargo, ningún centro había aportado un grupo de niños limitado a este relativamente estrecho rango de edad.

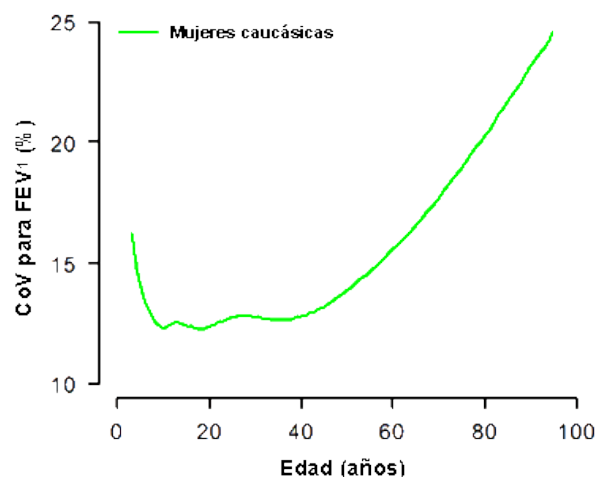
La evidencia de que los hallazgos no eran un artefacto provino del análisis de datos de niños y niñas de 15 centros diferentes, comprendiendo diferentes grupos étnicos (figura 11, [32]). Dado que los determinantes de FEV₁ y VC no son los mismos, se desprende que luego del nacimiento la capacidad vital crece proporcionalmente más rápido que el FEV₁, y que este patrón es temporalmente revertido durante el crecimiento acelerado de la adolescencia [32].

“Límite inferior de la normalidad”

En la medicina clínica, el “rango normal” es generalmente definido como el rango de valores que comprende al 95% de una población sana. El límite inferior de la normalidad (LIN) es el valor de corte por debajo del cual se halla sólo el 2,5% de los individuos sanos, mientras que el límite superior de la normalidad (LSN) representa el umbral por encima del cual serán hallados los resultados de sólo el 2,5% de la población. En consecuencia, se considera que el 95% de la población sana tiene resultados “normales”, en tanto que en 2,5% son “demasiado bajos” y en 2,5% son “demasiado altos”, resultando en un 5% de falsos positivos. Los resultados de las espirometrías típicamente presentan valores de FEV₁ y VC que son demasiado bajos en presencia de enfermedad. El hallazgo de valores demasiado altos es más raro, y suele responder a errores técnicos o de procedimiento. Esto probablemente explique por qué en medicina respiratoria el límite inferior de la normalidad es definido como el valor que identifica al 5^{to} percentilo más bajo de una población de sujetos sanos no fumadores.

Existen varios métodos para derivar el límite inferior de la normalidad. El más elegante se basa en una “distribución normal” de los resultados de la medición. En ese caso (figura 12), 68% de las observaciones se encuentran entre +1 y -1 desvío estándar (DE) de la distribución, 90% entre +1,64 y -1,64 DE; 95% entre +1,96 y -1,96 DE y 99,7% entre +3 y -3 DE.

En un sujeto sano, los datos espirométricos varían con la edad, talla, sexo y grupo étnico. Luego de tomar esto en

Fig. 13 - El coeficiente de variación (CoV) para FEV₁ en mujeres blancas sanas varía con la edad.

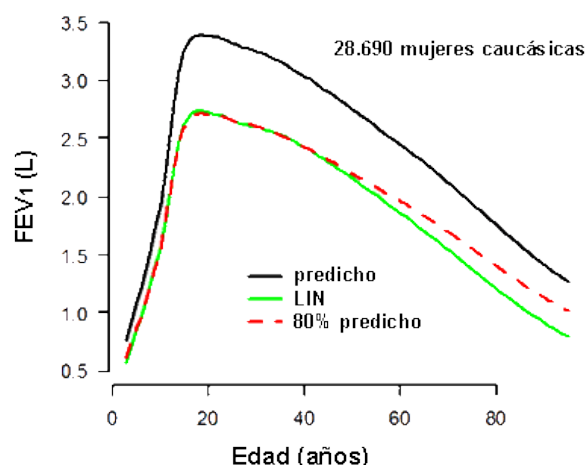


Fig. 14 - Valores predichos de FEV₁ y LIN en mujeres blancas sanas, y 80% del predicho, en función de la edad.

cuenta, nos resta el residuo (valor medido - valor predicho). Si el residuo se halla normalmente distribuido, el promedio de los residuos es 0. Dividir los residuos por el DE de la distribución $[(\text{medido}-\text{predicho})/\text{DE}]$ arroja un número adimensional, el z-score. En el caso de una distribución normal, el promedio de todos los z-scores es 0, y el DE es 1 (fig. 12).

El DE (o coeficiente de variación: $\text{CoV} = 100 \times \text{DE}/\text{predicho}$) varía con la edad [16,24]. Por lo tanto, el coeficiente de variación debe ser modelado de modo de obtener una distribución normal, es decir, independiente de la edad. Nuevamente, un *spline* puede ser usado para un modelado óptimo:

$$\log(\text{CoV}) = a + b \cdot \log(\text{edad}) + \text{spline} + \text{error}$$

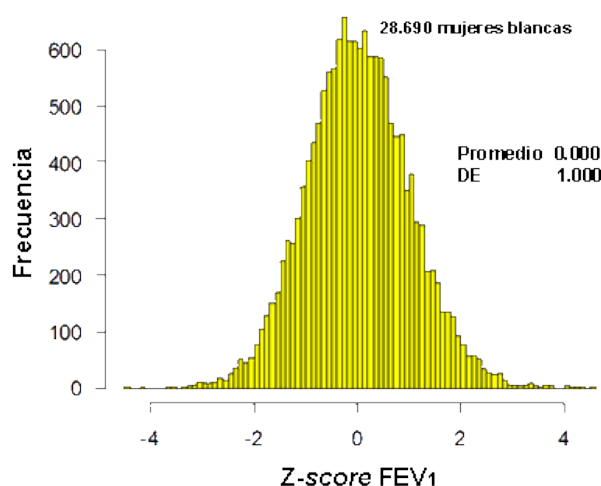


Fig. 15 - Distribución de z-scores para FEV₁ en mujeres blancas sanas.

El coeficiente de variación para FEV₁ en mujeres blancas varía entre 12,5% y 25% (fig. 13). Cómo afecta esto al límite inferior de la normalidad? A edades de 3, 20 y 80 años el coeficiente de variación es aproximadamente 16%; 12,5% y 21% respectivamente. El límite inferior de la normalidad en medicina respiratoria es el 5^{to} percentil, donde el z-score es -1.64, es decir, el valor predicho menos 1,64 veces el coeficiente de variación. De esto se desprende que el límite inferior de la norma-

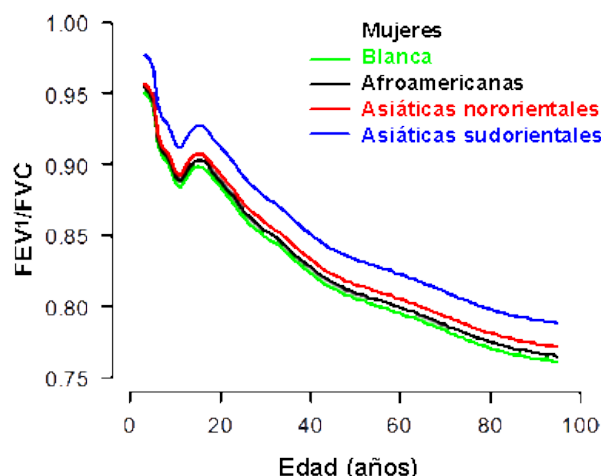


Fig. 16 - Cociente FEV₁/FVC en mujeres sanas de diferente origen étnico.

lidad para el FEV₁ en mujeres blancas de 3; 20 y 80 años de edad es de 74%, 80% y 66% del valor predicho. Esto confirma, una vez más, que no deberíamos utilizar el 80% del valor predicho como límite inferior de la normalidad.

Para poner esto en mayor perspectiva, podemos describir el valor predicho y el límite inferior de la normalidad para FEV₁ (según GLI-2012) en mujeres blancas como una función de la edad. Al agregar la línea que representa el 80% del predicho se observa que, particularmente en adultos, esta línea se desplaza progresivamente más alto dentro del rango normal, conduciendo a una proporción progresivamente mayor de resultados falsos-positivos (figura 14).

Como fuera explicado más arriba, el procedimiento adoptado debería conducir a una distribución normal de los residuos, de modo que los z-scores tuvieran un promedio entre 0 y DE. La figura 15 demuestra que es posible alcanzar esto con el paquete estadístico GAMLSS [31]. Esto se halla asociado con marcados beneficios: el z-score es completamente independiente de la edad, talla y sexo. Por ejemplo, si el z-score para cualquier índice es -1.64, significa en varones y mujeres, niños o adultos, que el valor medido se encuentra en el 5^{to} percentil; en función pulmonar, esto es considerado como el límite inferior de la normalidad.

Etnicidad

Es bien conocido que la función pulmonar difiere entre grupos étnicos. En el pasado, se usaban “factores de corrección étnica”, implicando que los valores predichos para una variable respiratoria de, por ejemplo, sujetos negros,

Table 1 - Porcentaje de diferencia en la función pulmonar respecto a sujetos blancos, por edad y sexo [24].

	Varones			Mujeres		
	FEV ₁	FVC	FEV ₁ /FVC	FEV ₁	FVC	FEV ₁ /FVC
Afroamericanos	-13.8	-14.4	0.6	-14.7	-15.5	0.8
Asiáticas nororientales	-0.7	-2.1	1.1	-2.7	-3.6	0.9
Asiáticas sudorientales	-13.0	-15.7	2.9	-9.7	-12.3	2.8
Combinados/otros	-6.8	-7.9	1.1	-6.8	-7.9	1.1

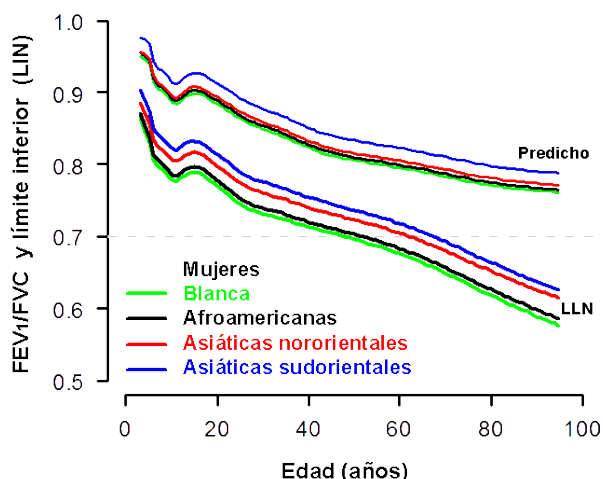


Fig. 17 - Valor predicho del cociente FEV₁/FVC y límite inferior de la normalidad en mujeres sanas de diferentes etnias.

era calculada como un 15% por debajo de los valores de sujetos blancos. Estos “factores de corrección” se hallaban determinados empíricamente en adultos. La disponibilidad de una gran número de registros de sujetos de 3 a 95 años de edad de diferentes procedencias étnicas permitió a la Global Lung Function Initiative estudiar esas diferencias étnicas en mayor profundidad. La fig. 16 ilustra una importante observación: con la excepción de los asiáticos del sudeste de Asia (sur de China, Tailandia, Corea), el cociente FEV₁/FVC es el mismo en todos los grupos étnicos a una determinada edad y talla. Esto implica que las diferencias en FEV₁ y FVC entre grupos étnicos son proporcionales e independientes de la edad. Biológicamente, esto tiene sentido. Después de todo, todos los grupos étnicos pertenecen al género *Homo sapiens*; es decir, son subgrupos de mamíferos que se adaptaron a diferentes condiciones locales y socio-económicas. En un proceso evolucionario que abarcó millones de años, los mamíferos fueron provistos de un diseño pulmonar escalable, de modo que puede adaptarse tanto a animales pequeños como grandes, abasteciendo sus necesidades metabólicas, entre otras, bajo circunstancias muy diferentes [33]. Las diferencias en variables respiratorias entre grupos étnicos no son más que una cuestión de diferentes escalas. Basados en este hallazgo de diferencias proporcionales, podemos ahora agregar grupos étnicos a nuestro modelo, según:

$$\log(Y) = a + b \cdot \log(\text{talla}) + c \cdot \log(\text{edad}) + d \cdot \text{Etn} + \text{spline} + \text{error}$$

La etnia (Etn) es ahora un cofactor. Las diferencias medias en la función pulmonar de una cierta cantidad de grupos étnicos, en relación a los blancos, son mostradas en la tabla. El grupo “combinados/otros” se refiere a personas de origen étnico combinado; las figuras en la tabla son un estimado para este grupo. Se aguardan más estudios al respecto.

Lo antedicho representa un importante avance, ya que todos los grupos étnicos pueden ahora ser incluidos en la ecuación de regresión. Esto no resuelve todos los problemas, ya que parece haber diferencias en la dispersión de datos alrededor de las medias predichas. Esto implica que es necesario ajustar el modelo al coeficiente de variación,

mostrado previamente, del siguiente modo:

$$\log(\text{CoV}) = a + b \cdot \log(\text{edad}) + d \cdot \text{Etn} + \text{spline} + \text{error}$$

El cociente FEV₁/FVC es un índice objetivo fundamental en el diagnóstico de la patología de la vía aérea. Mientras que los valores predichos para este cociente varían escasamente entre los diferentes grupos étnicos, el límite inferior de la normalidad es claramente diferente. El grupo GOLD consideró como muy dificultoso el cálculo del límite inferior de la normalidad para FEV₁/FVC y decidió que era mucho más sencillo adoptar un límite fijo de 0,70. Muchas críticas han sido publicadas respecto a este abordaje científicamente incorrecto, así como sobre la falta de evidencia de que la patología obstructiva de la vía aérea pueda de este modo ser apropiadamente diagnosticada. Al respecto es posible consultar la Carta Abierta a los Miembros de GOLD, firmada por un amplio número de reputados investigadores y clínicos [34]. La figura 17 también muestra que el criterio GOLD puede llevar al hallazgo espúreo de que la EPOC (enfermedad pulmonar obstructiva crónica) es menos prevalente en Asiáticos Orientales, ya que el límite inferior de la normalidad (5^o percentilo) para FEV₁/FVC permanece por encima del límite de 0,70 hasta una edad más avanzada que para blancos y negros.

El “límite inferior de normalidad” para espirometría, una vez más

Sin duda, la distribución de los índices de función pulmonar de sujetos sanos y aquellos con patología pulmonar se superponen. Por lo tanto, es arriesgado concluir que el resultado de una prueba en valores superiores al límite inferior de la normalidad (5^o percentil) descarta enfermedad; no hace falta decir que el juicio clínico es fundamental. En ese sentido, se ha sugerido que un cociente FEV₁/FVC < 0,70 pero > límite inferior de la normalidad, por consiguiente dentro del rango normal y denominado “zona incierta”, representa enfermedad pulmonar. No hay evidencia que apoye esto. Si los sujetos en la “zona incierta” desarrollasen signos y síntomas respiratorios con el paso de los años, esto puede llegar a sustentar esta postura. Sin embargo, no se ha encontrado evidencia a favor en estudios longitudinales: El estadio 1 de GOLD (FEV₁/FVC < 0,70 y FEV₁ > 80% del

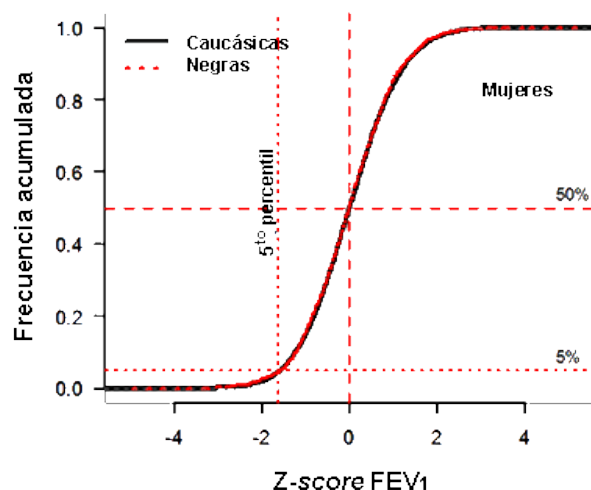


Fig. 18 - Distribución de las frecuencias acumuladas de z-scores para FEV₁ en mujeres blancas y negras, sanas y no fumadoras.

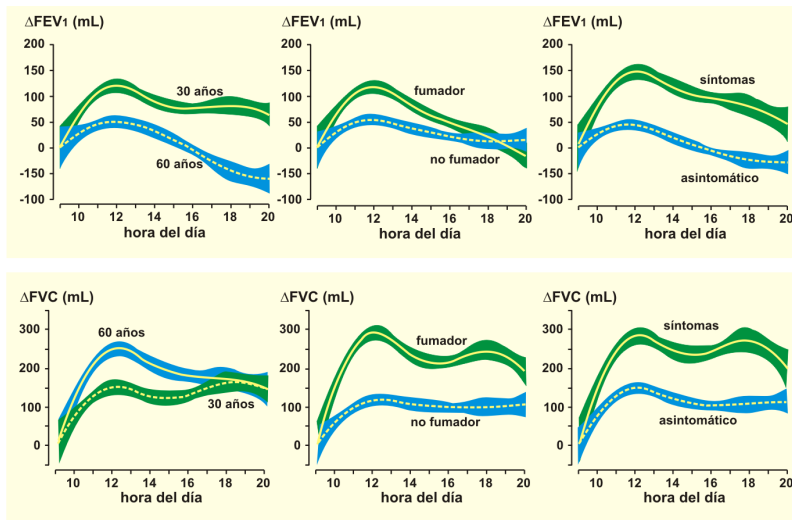


Fig. 19 - Variación circadiana y estacional de la función pulmonar. Datos derivados de una población normal, de mediciones realizadas con intervalos de 3 años por hasta 12 años [43].

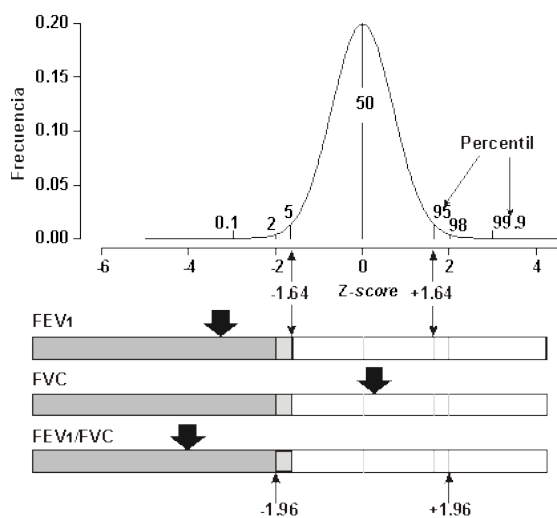


Fig. 20 - Relación entre el percentil y el z-score, y su uso en un pictograma para facilitar la interpretación de los resultados espirométricos.

predicho) en sujetos asintomáticos no está asociado a

- Muerte prematura [35-39]
- Declinación acelerada del FEV₁, desarrollo de síntomas respiratorios, incremento en el uso de recursos sanitarios, deterioro de la “calidad de vida” [40].

FEV₁/FVC < límite inferior de la normalidad *se* asocia a

- Muerte prematura [36,41]
- Desarrollo de síntomas respiratorios [42].

Conclusión: el criterio GOLD carece de fundamento científico o clínico, y el uso de FEV₁/FVC < 0,70 como criterio para diagnosticar obstrucción de la vía aérea debería ser desaconsejado en vista del subdiagnóstico en sujetos jóvenes y extenso sobrediagnóstico en adultos mayores [34]; en cambio, el límite inferior de la normalidad (LIN) debería ser usado.

Etnicidad y z-score

No hace ningún daño ilustrar la utilidad del z-score en las pruebas de función pulmonar desde una perspectiva adicional. En la figura 18, de derecha a izquierda, los z-scores se relacionan a una proporción de la población cada vez más grande. Remplace el conteo absoluto con el porcentaje acumulado en el eje Y para obtener la figura de abajo. La escala es de 0 (cero sujetos) a 1 (todos los sujetos, 100% de la

Fig. 21 - La gran cantidad de datos no permite una fácil interpretación de los estudios de función pulmonar. El uso de pictogramas, que resumen los hallazgos (abajo izquierda), permite la interpretación con sólo un vistazo.

	FEV ₁ /FVC	FEV ₁	FVC	FEF _{25-75%}	FEF _{75%}
Pre-BD	0.65	2.23	3.45		
Post-BD	0.70	2.57	3.69		
Predicted	0.80	4.21	5.26		
LLN	0.70	3.33	4.17		
Z-score (pre)	-2.36	-3.58	-2.76		
post	-1.69	-3.00	-2.39		
% predicted		53.0	65.6		
post		61.1	70.2		
Percentile	0.9	0.02	0.3		
post	5.0	0.1	0.8		

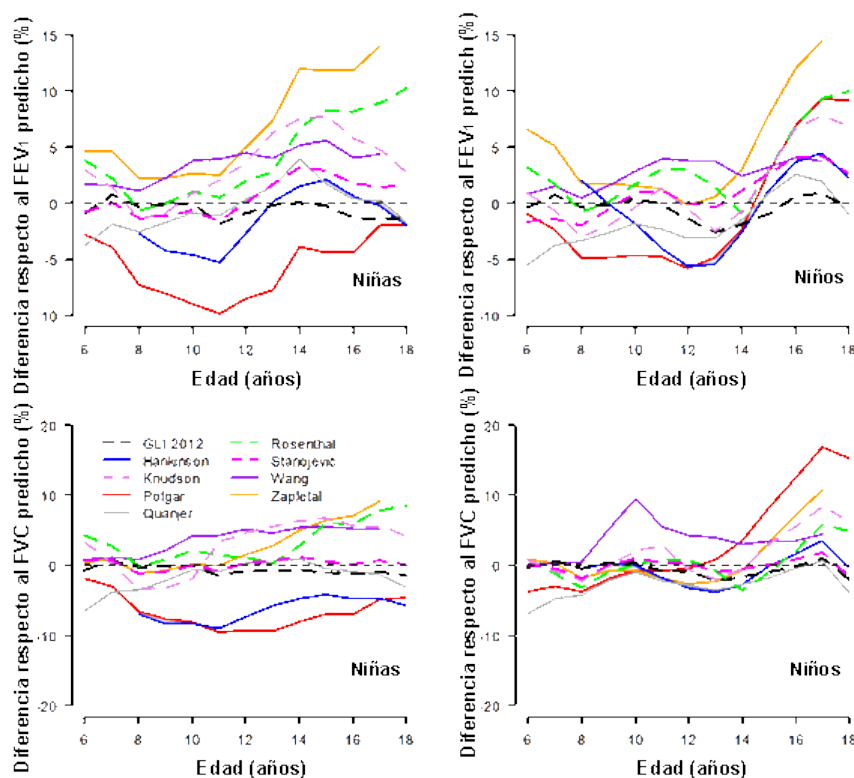


Fig. 22 - Comparación de los valores de FEV1 y FVC en niños y niñas sanos, predichos por GLI-2012 [24], Zapletal [44], Stanojevic [16], Polgar [12], Quanjer [13], Hankinson [25], Knudson [45], Rosenthal [46] y Wang [47].

población). La distribución de las frecuencias acumuladas en mujeres blancas sanas (figura 18) es diferente de la de las mujeres negras. Esto ilustra una vez más sobre la gran utilidad de los z-scores, ya que pueden ser interpretados independientemente del grupo étnico.

Interpretación de resultados de las pruebas

Las pruebas de función pulmonar muestran un resultado instantáneo. El resultado no sólo refleja la presencia o ausencia de enfermedad respiratoria, sino que sus resultados se ven también influidos por la hora del día, variaciones diarias y estacionales, etc. (fig. 19 [44]). Al interpretar los resultados, es necesario tener en cuenta que existe variabilidad inter-sujetos. Por lo tanto, hay una mayor incertidumbre respecto a la “anormalidad” de un hallazgo cerca del

límite inferior de la normalidad, que si el resultado se encontrase claramente dentro del rango de normalidad o del rango patológico (fig. 20 y 21).

La forma en que los resultados espirométricos son usualmente presentados ayuda poco en la interpretación y mistifica al evaluador inexperto: valores observados de FEV1, FVC, FEV1/FVC, junto con índices adicionales, tales como valores pre y post broncodilatadores, valores predichos, límites inferiores de la normalidad, porcentajes del predicho. Esto representa una matriz de datos impenetrable que confunde a la mayoría de los receptores de esta información, sean clínicos, técnicos o pacientes. Por el contrario, pictogramas en los que se muestran los z-scores en relación al rango de normalidad permiten la interpretación de los resultados en un parpadeo (fig. 20 y 21).

Comparación de valores predichos

Los pediatras en Holanda confían casi exclusivamente en los valores de referencia de Zapletal [45]. Estos están ba

Tabla 2 - Tasa de prevalencia de obstrucción de la vía aérea según GLI-2012 y otras ecuaciones de predicción.

Autor	FEV ₁ /FVC < LIN	
	Niños n = 2492	Niñas n = 2072
Hankinson	17.8%	14.3%
Knudson	21.0%	10.5%
Quanjer GLI-2012	15.0%	14.0%
Wang	21.6%	16.8%
Zapletal	23.1%	10.9%

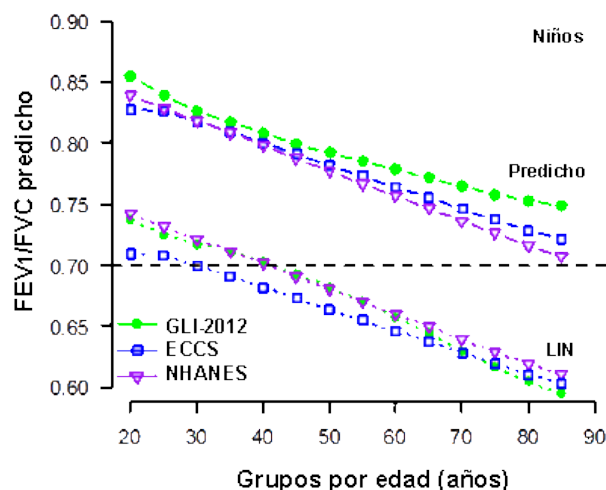
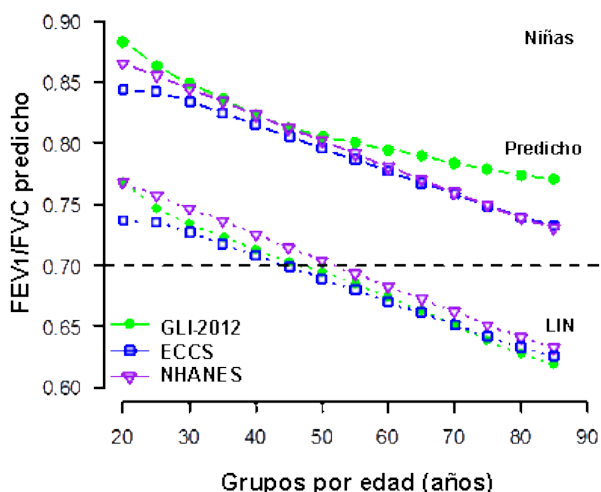


Fig. 23 - Comparación del cociente FEV1/FVC en adultos predicho por Quanjer GLI-2012 [24], Hankinson [25] y ECCS/ERS [10].

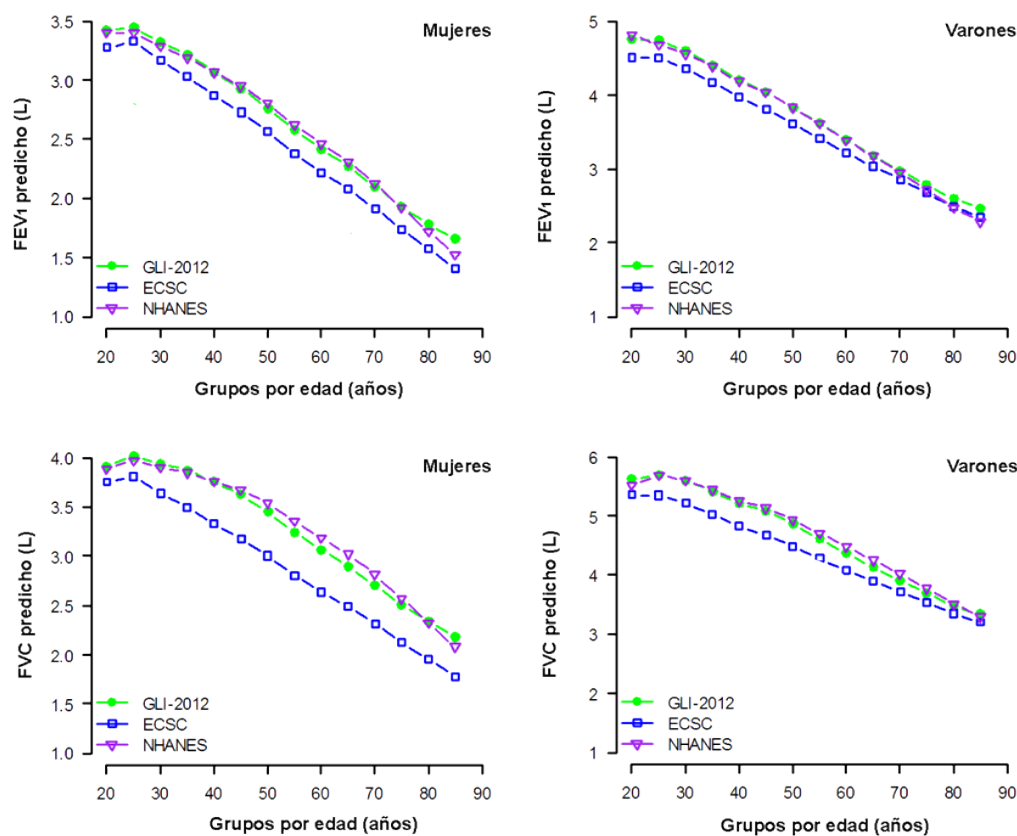


Fig. 24 - Comparación de los valores predichos de FEV₁ y FVC en adultos sanos, según Quanjer GLI-2012 [24], ECCS/ERS [10] y NHANES [25].

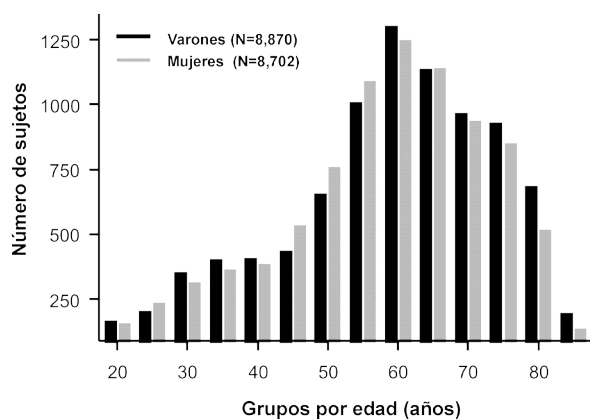


Fig. 25 - Distribución de los pacientes por edad (Australia, Polonia).

ados en un número muy limitado de sujetos (111 niños y niñas), y las ecuaciones de regresión toman en cuenta sólo la estatura, no la edad (rango 6 a 17 años). En otros países, los valores predichos de Polgar [12], Knudson [46], Quanjer [13], Rosenthal [47], Wang [48] y Hankinson [25] son frecuentemente utilizados. A diferencia de los de Zapletal, Polgar, Wang, Rosenthal y Knudson, los valores predichos por Stanojevic [16] provienen de una población de niños sanos (fig. 22).

En adultos (fig. 23) el cociente FEV₁/FVC predicho por ECCS/ERS [10] y NHANES [25] difieren de los de GLI-2012 [24]. Esto se debe principalmente al hecho de que las ecuaciones de GLI-2012 toman en cuenta que el cociente está inversamente relacionado a la talla de pie, en tanto las

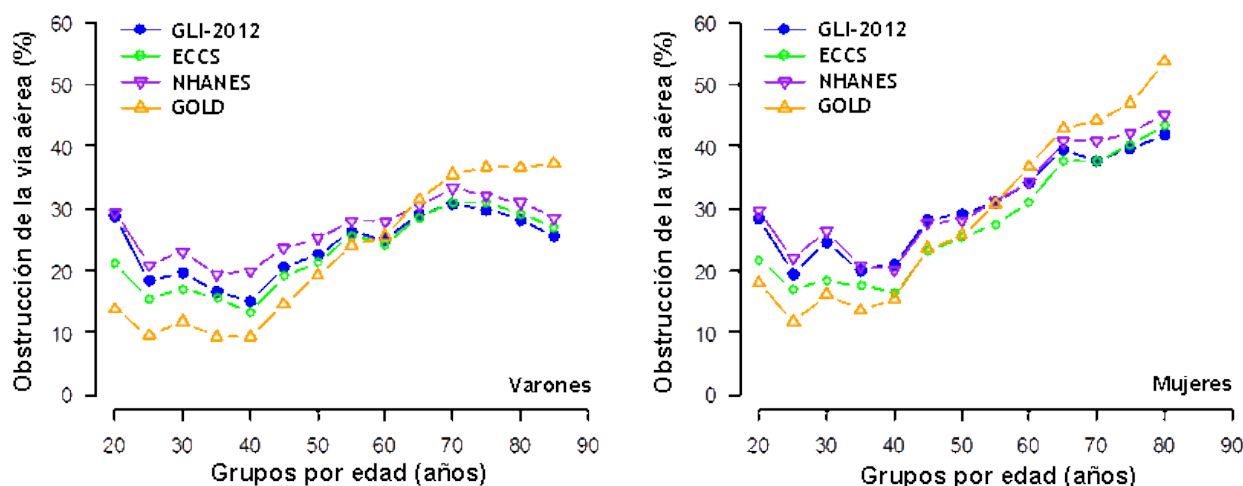
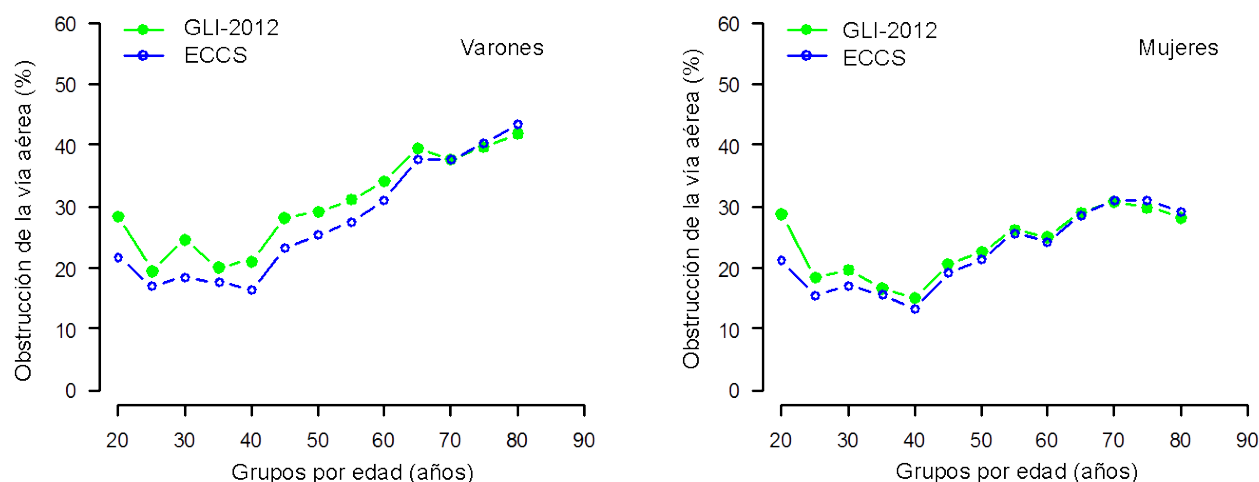
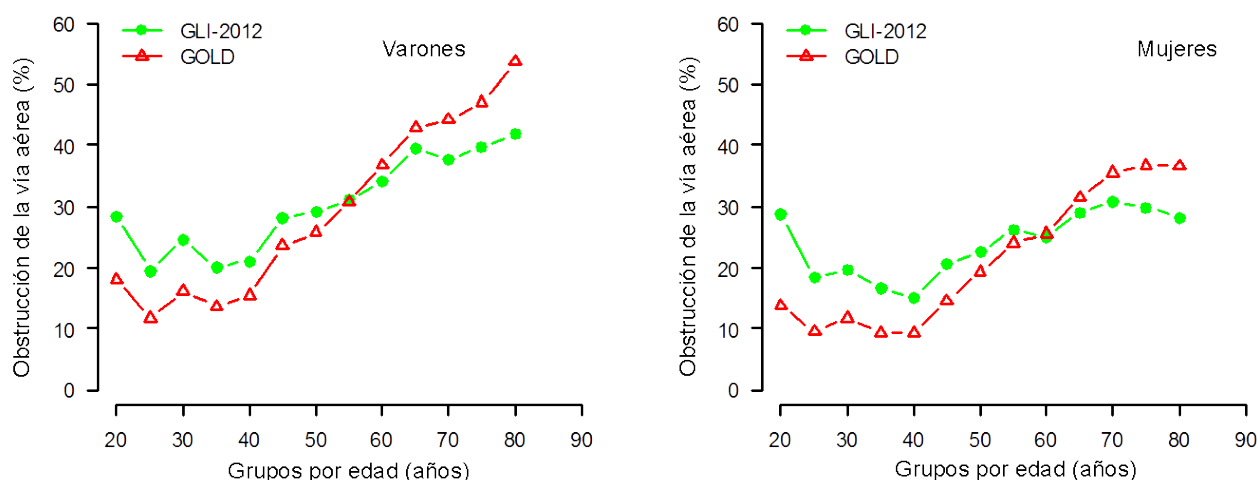


Fig. 26 - Porcentaje de pacientes con obstrucción de la vía aérea (FEV₁/FVC < LIN) basado en los valores predichos de Quanjer GLI-2012, Hankinson (NHANES), ECCS/ERS y en las guías GOLD.

Fig. 27 - Porcentaje de pacientes con obstrucción de la vía aérea ($FEV_1/FVC < LIN$) basado en los valores predichos de Quanjer GLI-2012 y ECCS/ERS.Fig. 28 - Porcentaje de pacientes con obstrucción de la vía aérea ($FEV_1/FVC < LIN$) basado en los valores predichos de GLI-2012 [23], o en los estadios GOLD 2-4.

otras dos ecuaciones consideran sólo la edad. Los valores predichos de FEV_1 y FVC predichos por NHANES concuerdan con los de GLI-2012, en tanto los de ECCS/ERS son definitivamente más bajos (fig. 24). Consecuentemente, los valores predichos por ECCS/ERS, ampliamente utilizados en Europa, debieran ser abandonados.

Obstrucción de la vía aérea

La aplicación de valores predichos de varios autores al in-

dice FEV_1/FVC de un grupo de pacientes pediátricos del Children's Hospital de Pittsburgh (cortesía Dr. Weiner) muestra diferencias en la tasa de prevalencia de obstrucción de la vía aérea en niños y en niñas, aunque algo menor en estas últimas (tabla 2).

Datos de pacientes con una amplia variedad de diagnósticos provenientes de dos hospitales en Australia y uno en Polonia mostraron la siguiente tendencia (ver figura abajo). Existe concordancia en la tasa de obstrucción de la vía aérea

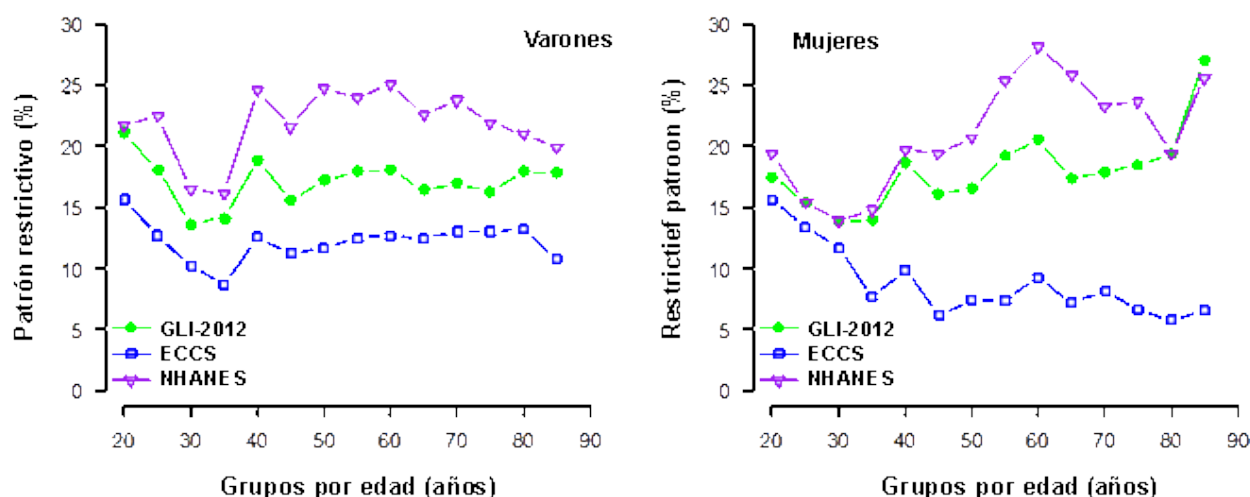
Fig. 29 - Porcentaje de pacientes con un "patrón restrictivo" espirométrico: VC demasiado baja pero cociente FEV_1/FVC normal o alto.

Tabla 3 - Redondeo de la edad, en este caso de 0,75 años, conduce a errores en los valores predichos para FEV₁ y FVC.

Edad (años) redondeada	Varones		Mujeres	
	FEV ₁ % error	FVC % error	FEV ₁ % error	FVC % error
3 vs 3.75	-2.8	-3.4	-2.9	-3.6
10 vs 10.75	-1.3	-1.4	-2.6	-2.7
15 vs 15.75	-3.4	-2.9	-3.4	-2.9
50 vs 50.75	+0.4	+0.4	+0.6	+0.7
85 vs 85.75	+0.7	+0.5	+0.9	+1.0

entre los valores predichos por GLI-2012 [24] y NHANES [25], aunque esta última produce un incremento sistemático de la prevalencia de obstrucción de la vía aérea en mujeres. Las ecuaciones de predicción de ECCS/ERS [3] generan una prevalencia algo menor en varones hasta 60 años de edad y en mujeres jóvenes (ver figura). En general, las diferencias son relativamente pequeñas. En consecuencia, la adopción de las ecuaciones de Quanjer GLI-2012 no provocarán un cambio significativo en la tasa de prevalencia de obstrucción de la vía aérea.

Como fuera explicado previamente el estadio 1 no representa enfermedad respiratoria. Por lo tanto, el análisis se encuentra limitado a los estadios 2 a 4 de la clasificación GOLD. La tasa de prevalencia de los estadios 2 a 4 de GOLD tiene el mismo patrón anteriormente publicado para el estadio 1: subdiagnóstico de obstrucción de la vía aérea ($\approx 20\%$) hasta la edad de 55-60 años, y sobrediagnóstico ($\approx 20\%$) por encima de esa edad. Estos porcentajes concuerdan con los reportados en un estudio clínico previo [48]. Esto indica que un sesgo relacionado a la edad afecta al estadio 2 de GOLD. Esto se debe en parte al hecho de que el FEV₁ debería ser $<80\%$ del valor predicho. Concluimos previamente que no sólo FEV₁/FVC $<0,70$ sino que también FEV₁ $<80\%$ se asociaba a un fuerte sesgo relacionado a la edad (fig. 26, 27 y 28).

“Patrón restrictivo”

En 1991 un comité de la ATS sugirió que era posible evidenciar un defecto ventilatorio restrictivo, es decir, una condición en la cual la capacidad pulmonar total se encuentra reducida, sobre la base de una VC anormalmente baja en combinación con un cociente FEV₁/FVC normal o aumentado: “patrón restrictivo” [22]. Desde entonces el patrón restrictivo ha sido regularmente descripto en la literatura, sugiriendo que sea considerado un patrón clínicamente significativo. La prevalencia en una población de pacientes hospitalarios australianos y polacos variaba según la edad entre 5 y 20% (figura 29); la cantidad de observaciones por encima de la edad de 80 años era muy limitada, por lo que ese patrón, por encima de esa edad, debería ser olvidado. Las diferencias en la tasa de prevalencia de acuerdo a tres ecuaciones de predicción es considerable. El patrón general es que la adopción de las ecuaciones GLI-2012 conducen a un incremento en la tasa de prevalencia del patrón restrictivo, comparada con ECCS/ERS. Esto es preocupante, ya que puede llevar a un incremento en las solicitudes de medición

de capacidad pulmonar total, llevando a un incremento en el gasto médico. Es conocido que este patrón espirométrico tiene una baja sensibilidad para diagnosticar correctamente enfermedad pulmonar restrictiva: 50% o menos en una población de pacientes [50-52]. La restricción pulmonar es rara en la población general, de modo que es mejor si los médicos generalistas ignoran un patrón restrictivo. De hecho, en general es mejor ignorar este patrón, a menos que haya evidencia clínica compatible con restricción pulmonar (resección pulmonar, xifoesciosis severa, etc) y que documentar ese defecto sea clínicamente relevante. La idea general debería ser: “trate al paciente, no a los números”.

Determinación precisa de la talla y la edad

Talla

La talla debe ser medida, ya que el dato auto-reportado no es confiable. La diferencia entre la talla medida y la auto-reportada puede ser de hasta 6,9 cm y es generalmente más marcada en sujetos ancianos [53-58]. FEV₁ y FVC son función de la talla multiplicada por una constante (k), donde $k \approx 2,2$. En un niño de 110 cm de estatura, o un adulto de 180 cm, un error de 1 cm conduce a un error en las variables de función pulmonar predichas de 2% y 1,2% respectivamente. No sólo debería ser medida la talla de pie, sino que el tallímetro debería ser calibrado anualmente, mientras que el valor de la talla ingresado para la predicción de las variables espirométricas debería tener una precisión de 1 decimal [24, 59].

Edad

El efecto de la imprecisión de la edad sobre los valores predichos no puede ser estimado tan fácilmente, debido a la contribución de la variable del *spline* en la edad. Si la edad es sistemáticamente subestimada en 0,75 años debido al redondeo, el porcentaje de error es el mostrado en la tabla 3.

Los errores varían con la edad, siendo más manifiestos en la infancia. Por lo tanto, en el cálculo de las variables predichas, la edad debería ser ingresada con una precisión de 1 decimal [24, 59].

Validación

Los errores varían con la edad, siendo más manifiestos en la infancia. Por lo tanto, en el cálculo de las variables predichas, la edad debería ser ingresada con una precisión de 1 decimal [60-62].

Software

Dos tipos de software (gratuito) están disponibles para generar valores predichos de acuerdo a las ecuaciones de referencia Quanjer GLI-2012:

- 1 **Software para calcular valores predichos para un individuo**
Este software se encuentra disponible como un programa para sistema Windows, y en formato Excel

2 **Software para transformar grandes conjuntos de datos**, de modo que los valores predichos, el límite inferior de la normalidad y los z-scores sean adicionados a los datos. Este software gratuito está asimismo disponible como una aplicación para sistema Windows y como una planilla Excel.

Este *software* puede ser descargado desde

[software](#)

Adicionalmente, los fabricantes de espirómetros han implementado las ecuaciones GLI-2012 en su software, o se encuentran en el proceso de hacerlo. Se puede hallar información al respecto en

[fabricantes](#).

Flujos

Existen preguntas recurrentes sobre por qué no fueron incluidos en GLI-2012 valores de referencia para flujos instantáneos, tales como el FEF₅₀. Estos flujos nunca demostraron agregar utilidad al FEV₁ y la VC. Estos flujos son a menudo considerados como índices sensibles de enfermedad de la “pequeña vía aérea”, un síndrome que podría estar ocurriendo sin afectar a la gran vía aérea intrapulmonar, de una manera que sería detectable mediante la espirometría; esta visión fue refutada ya en 1991 [22]. El coeficiente de variación de los flujos instantáneos es tan grande que parcialmente explica su poco satisfactorio desempeño en la toma de decisiones clínicas. Asimismo, los flujos pre y post broncodilatador no pueden ser comparados si ocurre un cambio en la FVC, ya sea por diferente técnica o espontáneo, y los valores predichos para los flujos son inválidos si el FVC se ve afectado por una enfermedad. Es por esta razón que el uso de flujos instantáneos para propósitos diagnósticos no es recomendado en los reportes de estandarización, y no desempeñan un rol en los algoritmos diagnósticos [10, 14, 22, 61-62]

En pacientes pediátricos, los flujos instantáneos son aún frecuentemente utilizados. Por esta razón, y ante una solicitud especial, el grupo GLI agregó valores predichos para FEF_{75%} y FEF_{25-75%}.

Factor de transferencia o capacidad de difusión

El grupo GLI ha iniciado la derivación de valores predichos para factor de transferencia/capacidad de difusión. El grupo, bajo el liderazgo de Brian Graham y Graham Hall, recibió estado de “task force” de la ATS.

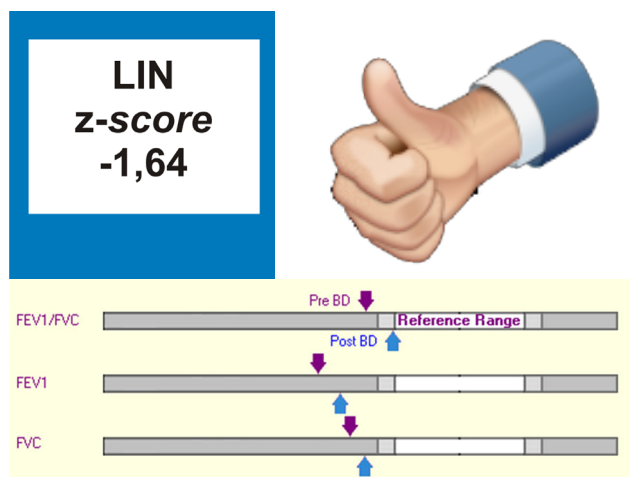
El factor de transferencia del pulmón es a menudo llamado capacidad de difusión del pulmón. Sin embargo, el pulmón no difunde. Además, esta medición no representa una capacidad ya que, por ejemplo durante el ejercicio, la transferencia de O₂ o CO a través del pulmón es mucho mayor que durante el reposo. Por lo tanto, factor de transferencia es un mejor nombre.

Volúmenes pulmonares

En esta etapa no hay planes para generar ecuaciones de regresión para volúmenes pulmonares (RV, TLC, FRC). Esto es en parte debido a que hay demasiadas técnicas diferentes para medirlos y porque existen pocos datos sobre sujetos sanos. Adicionalmente, muchos sostienen que la medición de volúmenes pulmonares es de limitada utilidad en la práctica clínica.

Conclusiones

- 1 El estudio desarrollado por la Global Lung Initiative está basado en una muy amplia muestra representativa de la población.
- 2 Sus recomendaciones han sido respaldadas por 6 grandes sociedades respiratorias internacionales: ERS, ATS, Australian and New Zealand Society of Respiratory Science, Asian Pacific Society for Respiriography, Thoracic Society of Australia and New Zealand y el American College of Chest Physicians.
- 3 GLI-2012 provee ecuaciones de regresión para el rango de edad de 3 a 95 años, y para un cierto número de grupos étnicos.
- 4 La dependencia del LIN respecto a la edad fue tenida en cuenta.
- 5 Los z-scores ofrecen la oportunidad de interpretar los resultados de las pruebas independientemente de la edad, talla, sexo y grupo étnico.



- 6 La adopción de las ecuaciones Quanjer GLI-2012 conducirá a cambios menores en la tasa de prevalencia de obstrucción de la vía aérea en poblaciones clínicas.
- 7 El uso de porcentaje del valor predicho conduce a un sesgo dependiente de la edad inaceptable y debe ser reemplazado por el uso de *z-scores*.
- 8 La doctrina GOLD no respeta el LIN clínicamente válido y conduce a considerable sub y sobrediagnóstico de obstrucción de la vía aérea.
- 9 La adopción de las ecuaciones Quanjer GLI-2012 conducirá a incrementar la tasa de prevalencia del "patrón restrictivo" respecto a ECCS [64-66] : la recomendación clínica en estos casos es: "trate al paciente, no los datos".

Agradecimientos

- Figuras 6, 7 y 20: las figuras fueron modificadas y reproducidas con permiso de la European Respiratory Society. *Eur Respir J* 2012 40:1324-1343; doi:10.1183/09031936.00080312. Epub 2012 Jun 27.
- Figura 11: modificadas y reproducidas con permiso de la European Respiratory Society. *Eur Respir J* Diciembre 2010 36:1391-1399; doi:10.1183/09031936.00164109. Epub Mar 2010.
- Figura 22: modificada y reproducida con permiso de la European Respiratory Society. *Eur Respir J* July 2012 40:190-197; Epub December 19, 2011, doi:10.1183/09031936.00161011.
- Figuras 25 y 29: las figuras fueron modificadas y reproducidas con permiso de la European Respiratory Society. *Eur Respir J* 2013; in press; doi: 10.1183/09031936.00195512.

Referencias

- 1 Hutchinson J. On the capacity of the lungs, and on the respiratory functions, with a view of establishing a precise and easy method of detecting disease by the spirometer. *Med Chir Trans (London)* 1846; 29: 137-252.
- 2 Tiffeneau R, Pinelli A. Air circulant et air captif dans l'exploration de la fonction ventilatrice pulmonaire. *Paris Méd* 1947; 37: 624-628.
- 3 Yernault JC. The birth and development of the forced expiratory manoeuvre: a tribute to Robert Tiffeneau (1910-1961). *Eur Respir J* 1997; 10: 2704-2710.
- 4 Jouasset D. Normalisation des épreuves fonctionnelles respiratoires dans les pays de la Communauté Européenne du Charbon et de l'Acier. *Poumon Coeur* 1960; 16: 1145-1159.
- 5 Cara M, Hentz P (1971). Aide-mémoire of spiographic practice for examining ventilatory function, 2nd edn. (Industrial Health and Medicine series, vol 11) pp. 1-130.
- 6 Ferris BC: Epidemiology Standardization Project. *Am Rev Respir Dis* 1978; 118 (Suppl, part 2): 1-120.
- 7 American Thoracic Society. 1979. Standardization of spirometry. *Am Rev Respir Dis* 1979; 119: 831-838.
- 8 Quanjer PH, ed. Standardized lung function testing. Report Working Party Standardization of Lung Function Tests. European Community for Coal and Steel. *Bull Eur Physiopathol Respir* 1983; 19: Suppl. 5, 1-95.
- 9 American Thoracic Society. Standardization of spirometry: 1987 update. *Am Rev Respir Dis* 1987; 136: 1285-1298.
- 10 Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault J-C. Lung volume and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J* 1993; 6: Suppl. 16, 5-40. Erratum *Eur Respir J* 1995; 8: 1629.
- 11 American Thoracic Society. Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med* 1995; 152: 1107-1136.
- 12 Polgar, G, Promadhat V. Pulmonary function testing in children: techniques and standards. Philadelphia, WB Saunders C, 1971.
- 13 Quanjer PH, Borsboom GJ, Brunekreef B, Zach M, Forche G, Cotes JE, Sanchis J, Paoletti P. Spirometric reference values for white European children and adolescents: Polgar revisited. *Pediatr Pulmonol* 1995;19: 135-142.
- 14 Miller MR, Hankinson J, Brusasco V, et al. ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J* 2005; 26: 319-338.
- 15 <http://www.lungfunction.org>.
- 16 Stanojevic S, Wade A, Stocks J, et al. Reference ranges for spirometry across all ages. A new approach. *Am J Respir Crit Care Med* 2008; 177: 253-260.
- 17 Gramprey J. Les valeurs "dites normales" en spiropgraphie. *Rev méd Nancy* 1954; 79: 648-657.
- 18 Bates DV, Christie RV. (1964). Respiratory Function in Disease, p. 91. Saunders, Philadelphia and London.
- 19 Sobol BJ. Assessment of ventilatory abnormality in the asymptomatic subject: an exercise in futility. *Thorax* 1966; 2: 445-449.
- 20 Sobol BJ, Sobol PG. Editorial. Percent of predicted as the limit of normal in pulmonary function testing: a statistically valid approach. *Thorax* 1979; 34: 1-3.
- 21 Miller MR, Pincock AC. Predicted values: how should we use them? *Thorax* 1988; 43: 265-267.
- 22 ATS Statement. Lung function testing: selection of reference values and interpretative strategies. *Am Rev Resp Dis* 1991; 144: 1202-1218.
- 23 Miller MR, Quanjer PH, Swanney MP, Ruppel G, Enright PL. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011; 139: 52-59.
- 24 Quanjer PH, Stanojevic S, Cole TJ et al. and the ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3-95 years age range: the Global Lung Function 2012 equations. *Eur Respir J* 2012; 40: 1324-1343.
- 25 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general US population. *Am J Respir Crit Care Med* 1995; 152: 179-187.
- 26 Wang X, Dockery DW, Wypij D, Fay ME, Ferris BG Jr. Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993; 15: 75-88.
- 27 Falaschetti E, Laiho J, Primatesta P, Purdon S. Prediction equations for normal and low lung function from the Health Survey for England. *Eur Respir J* 2004; 23: 456-463.
- 28 Brändli O, Schindler Ch, Künzli N, Keller R, Perruchoud AP, and SA-PALDIA team. Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. *Thorax* 1996; 51: 277-283.
- 29 Pistelli F, Bottai M, Viegi G, et al. Smooth reference equations for slow vital capacity and flow-volume curve indexes. *Am J Respir Crit Care Med* 2000; 161: 899-905. Erratum in: *Am J Respir Crit Care Med* 2001; 164: 1740.
- 30 Pistelli F, Bottai M, Carrozzi L, et al. Reference equations for spirometry from a general population sample in central Italy. *Respir Med* 2007; 101: 814-825.
- 31 Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). *Appl Statist* 2005; 54: 507-554.
- 32 Quanjer PH, Stanojevic S, Stocks J et al., for and on behalf of the Global Lung Initiative. Changes in the FEV₁/FVC ratio during childhood and adolescence: an intercontinental study. *Eur Respir J* 2010; 36: 1391-1399.
- 33 West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science* 1997; 276: 122-126.
- 34 Quanjer PH, Enright PL, Miller MR et al. Open Letter. The need to change the method for defining mild airway obstruction. *Eur Respir J* 2011; 37: 720-722.
- 35 Ekberg-Aronsson M, Pehrsson K, Nilsson JA, Nilsson PM, Löfdahl CG. Mortality in GOLD stages of COPD and its dependence on symptoms of chronic bronchitis. *Respir Res* 2005; 6: 98.
- 36 Vaz Fragoso CA, Concato J, McAvay G, et al. Chronic obstructive pulmonary disease in older persons: a comparison of two spirometric definitions. *Respir Med* 2010; 104: 1189 - 1196.
- 37 Pedone C, Scarlata S, Sorino C, Forastiere F, Bellia V, Antonelli Incalzi R. Does mild COPD affect prognosis in the elderly? *BMC Pulm Med* 2010; 10: 35.
- 38 Mannino DM, Doherty DE, Buist AS. Global Initiative on Obstructive Lung Disease (GOLD) classification of lung disease and mortality: findings from the Atherosclerosis Risk in Communities (ARIC) study. *Respir Med* 2006; 100: 115-122.
- 39 Vaz Fragoso C, Gill T, McAvay G, et al. Use of lambda-mu-sigma-derived Z score for evaluating respiratory impairment in middle-aged persons. *Respir Care* 2011; 56: 1771-1777.

- 40 Bridevaux P-O, Gerbase MW, Probst-Hensch NM, Schindler C, Gaspoz JM, Rochat T. Long-term decline in lung function, utilisation of care and quality of life in modified GOLD stage 1 COPD. *Thorax* 2008; 63: 768 - 774.
- 41 Akkermans R, Biermans M, Robberts B *et al.* COPD prognosis in relation to diagnostic criteria for airflow obstruction in smokers. *Eur Respir J* 2013 published ahead of print as doi: 10.1183/09031936.00158212.
- 42 Mannino DM, Buist AS, Vollmer WM. Chronic obstructive pulmonary disease in the older adult: what defines abnormal lung function? *Thorax* 2007; 62: 37-241
- 43 Vaz Fragoso CA, Concato J, McAvay G, *et al.* The ratio of FEV₁ to FVC as a basis for establishing chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010; 181: 446 - 451.
- 44 Borsboom GJJM, van Pelt W, van Houwelingen HC, van Vianen BG, Schouten JP, Quanjer PH. Diurnal variation in lung function in subgroups from two Dutch populations. Consequences for longitudinal analysis. *Am J Respir Crit Care Med* 1999; 159: 1163-1171.
- 45 Zapletal A, Paul T, Samanek N. Die Bedeutung heutiger Methoden der Lungenfunktionsdiagnostik zur Feststellung einer Obstruktion der Atemwege bei Kindern und Jugendlichen. *Z Erkrank Atm-Org* 1977; 149: 343-371.
- 46 Knudson RJ, Lebowitz MD, Holberg CJ, *et al.* Changes in the normal maximal expiratory flow-volume curve with growth and aging. *Am Rev Respir Dis* 1983; 127: 725-734.
- 47 Rosenthal M, Bain SH, Cramer D, *et al.* Lung function in white children aged 4-19 years: I - Spirometry. *Thorax* 1993; 48: 794-802.
- 48 Wang X, Dockery DW, Wypij D, *et al.* Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993; 15: 75-88.
- 49 Miller MR, Quanjer PH, Swanney MP, Ruppel G, Enright PL. Interpreting lung function data using 80% predicted and fixed thresholds misclassifies more than 20% of patients. *Chest* 2011; 139: 52-59.
- 50 Aaron SD, Dales RE, Cardinal P. How accurate is spirometry at predicting restrictive pulmonary impairment? *Chest* 1999; 115: 869-873.
- 51 Glady CA, Aaron SD, Lunau ML, *et al.* A spirometry-based algorithm to direct lung function testing in the pulmonary function laboratory. *Chest* 2003; 123: 1939-1946.
- 52 Swanney MP, Beckert LE, Frampton CM, *et al.* Validity of the American Thoracic Society and other spirometric algorithms using FVC and Forced Expiratory Volume at 6 s for predicting a reduced total lung capacity. *Chest* 2004; 126: 1861-1866.
- 53 Parker JM, Dillard TA, Phillips YY. Impact of using stated instead of measured height upon screening spirometry. *Am J Respir Crit Care Med* 1994; 150(6 Pt 1):1705-1708.
- 54 Brenner ND, McManus T, Galuska DA, Lowry R, Wechsler H. Reliability and validity of self-reported height and weight among high school students. *J Adolesc Health* 2003; 32: 281-287.
- 55 Braziuniene I, Wilson TA, Lane AH. Accuracy of self-reported height measurements in parents and its effect on mid-parental target height calculation. *BMC Endocrine Disorders* 2007; 7: 2.
- 56 Jansen W, van de Looij-Jansen P. M, Ferreira I, de Wilde EJ, Brug J. Differences in measured and self-reported height and weight in Dutch adolescents. *Ann Nutr Metab* 2006; 50: 339-346.
- 57 Lim LLY, Seubsman S-A, Sleight A. Validity of self-reported weight, height, and body mass index among university students in Thailand: Implications for population studies of obesity in developing countries. *Population Health Metrics* 2009; 7: 15.
- 58 Wada K, Tamakoshi K, Tsunekawa T *et al.* Validity of self-reported height and weight in a Japanese workplace population. *Intern J Obesity* 2005; 29: 1093-1099.
- 59 Quanjer PH, Hall GL, Stanojevic S, Cole TJ, Stocks J, on behalf of the Global Lungs Initiative. Age- and height-based prediction bias in spirometry reference equations. *Eur Respir J* 2012; 40: 190-197.
- 60 Lum S, Bonner R, Kirkby J, Sonnappa S, Stocks J. S33 Validation of the GLI-2012 multi-ethnic spirometry reference equations in London school children. *Thorax* 2012; 67: A18 (http://thorax.bmj.com/content/67/Suppl_2/A18.2).
- 61 Hall GL, Thompson BR, Stanojevic S, *et al.* The Global Lung Initiative 2012 reference values reflect contemporary Australasian spirometry. *Respirology* 2012; 17: 1150-1151.
- 62 Bonner R, Lum S, Stocks J, Kirkby J, Wade A, Sonnappa S. Applicability of the Global Lung Function spirometry equations in contemporary multi-ethnic children. Research letter accepted for publication in *Am J Respir Crit Care Med*.
- 63 Pellegrino R, Viegi G, Brusasco V, *et al.* ATS/ERS Task Force. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948-968.
- 64 Quanjer PH, Brazzale DJ, Boros PW, Pretto JJ. Implications of adopting the Global Lungs 2012 all-age reference equations for spirometry. *Eur Respir J* 2013; i42: 1046-1054.
- 65 Swanney MP, Miller MR. Adopting universal lung function reference equations. *Eur Respir J* 2013; 42: 901-903.
- 66 Quanjer PH, Weiner DJ. Interpretative consequences of adopting the global lungs 2012 reference equations for spirometry for children and adolescents. *Pediatr Pulmonol* 2014; 49: 118-125.