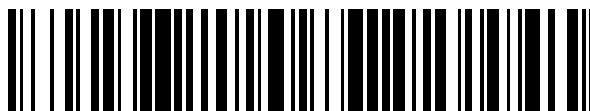


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 685 395**

51 Int. Cl.:

A61B 5/00 (2006.01)

G06F 19/00 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.07.2013 PCT/EP2013/064954**

87 Fecha y número de publicación internacional: **16.01.2014 WO14009569**

96 Fecha de presentación y número de la solicitud europea: **15.07.2013 E 13737244 (7)**

97 Fecha y número de publicación de la concesión europea: **20.06.2018 EP 2872031**

54 Título: **Método para proporcionar pruebas diagnósticas no invasivas confiables**

30 Prioridad:

13.07.2012 EP 12176372

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

08.10.2018

73 Titular/es:

**UNIVERSITÉ D'ANGERS (50.0%)
40 rue de Rennes
49000 Angers, FR y
CENTRE HOSPITALIER UNIVERSITAIRE
D'ANGERS (50.0%)**

72 Inventor/es:

**CALÈS, PAUL;
HUNAUULT, GILLES y
BOURSIER, JÉRÔME**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 685 395 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Método para proporcionar pruebas diagnósticas no invasivas confiables

5 Campo de la invención

La presente invención se refiere a un método para mejorar la confiabilidad de las pruebas diagnósticas no invasivas. Más específicamente, la presente invención se refiere a un método para analizar la confiabilidad de un Índice Inicial obtenido mediante la implementación de una prueba no invasiva para evaluar la presencia y/o la gravedad de una enfermedad, y para mejorar la exactitud del diagnóstico de dicha prueba.

El método de la invención puede aplicarse para mejorar la confiabilidad de cualquier método no invasivo para evaluar el presente diagnóstico y/o la gravedad de una enfermedad, y se aplica especialmente a enfermedades hepáticas.

15 Antecedentes de la invención

El diagnóstico de enfermedades hepáticas puede realizarse a través del análisis de la fibrosis hepática. La fibrosis hepática se refiere a la acumulación en el hígado de tejido cicatricial fibroso en respuesta a la lesión de los hepatocitos debido a varias etiologías, tales como por ejemplo la infección con un virus (tal como los virus de la hepatitis HCV y HBV), alto consumo de alcohol, toxinas o fármacos. La evolución de la lesión fibrótica puede conducir a la cirrosis, una afección en la que la capacidad de funcionamiento del hígado se afecta. Existen tratamientos para la fibrosis hepática, que pueden enlentecer o detener la progresión de la fibrosis, e incluso revertir los daños hepáticos existentes. Por el contrario, usualmente se considera que la cirrosis no es reversible.

La biopsia hepática es la herramienta implementada históricamente para diagnosticar enfermedades hepáticas en pacientes. Varias clasificaciones, basadas en las biopsias hepáticas, se usan para calificar fibrosis y cirrosis, tales como, por ejemplo, Metavir e Ishak (donde se califica la cirrosis). Por ejemplo, con el uso de la clasificación de puntuación de Metavir para la fibrosis, se distinguen cinco clases (denominadas estadios **F** de Metavir): F0 (sin fibrosis, sin cicatrización), F1 (fibrosis portal, cicatrización mínima), F2 (pocos septos, se ha producido cicatrización y se extiende fuera de las áreas en el hígado que contienen vasos sanguíneos), F3 (muchos septos, la fibrosis en puente se dispersa y se conecta con otras áreas que contienen fibrosis) y finalmente F4 (cirrosis o cicatrización avanzada del hígado). En esta solicitud de patente, cualquier mención de F0, F1, F2, F3 y F4 se hace con referencia a los estadios de Metavir.

Sin embargo, dado que la biopsia hepática es invasiva y costosa, el diagnóstico no invasivo de la fibrosis hepática ha ganado considerable atención en los últimos 10 años como una alternativa a la biopsia hepática. La primera generación de pruebas simples de fibrosis en sangre combinaba marcadores indirectos en sangre en una relación simple, como APRI (Wai y otros, Hepatology 2003) o FIB-4 (Valley-Pichard y otros, Hepatology 2007). La segunda generación de pruebas calculadas combinaba marcadores de fibrosis indirectos y/o directos mediante regresión logística, lo que conduce a una puntuación, como Fibrotest™ (Imbert-Bismut y otros, Lancet 2001), la puntuación ELF (Rosenberg y otros, Gastroenterology 2004), FibroMeter™ (Cales y otros, Hepatology 2005), Fibrospect™ (Patel y otros, J Hepatol 2004), y Hepascore (Adams y otros, Clin Chem 2005). Por ejemplo, el documento WO2005/116901 describe un método no invasivo para evaluar la presencia de una enfermedad hepática y su gravedad, mediante la medición de los niveles de variables específicas, que incluyen variables biológicas y variables clínicas, y la combinación de dichas variables en funciones matemáticas para proporcionar una puntuación, llamada frecuentemente "puntuación de fibrosis".

Sin embargo, estas pruebas diagnósticas no invasivas no son 100 % exactas. De hecho, pueden producirse resultados falsos positivos o falsos negativos, lo que conduce a la clasificación errónea del paciente. Los errores pueden atribuirse principalmente a la referencia (biopsia hepática) o a la construcción de la prueba (como se observa en los datos académicos). Por otra parte, otras fuentes de errores pueden surgir de la medición de los marcadores o de los datos físicos que fundamentan la prueba, del profesional, o del propio paciente.

Por lo tanto, existe la necesidad de un método para limitar la ocurrencia de clasificaciones erróneas de los pacientes, y mejorar la exactitud de las pruebas no invasivas. Un ejemplo de ineficacia de la suposición del estado de la técnica de la confiabilidad de Fibroscan en base al AUROC se muestra en el Ejemplo 6: Las AUROC de LSE en biopsias no confiables no fueron significativamente diferentes a las de biopsias confiables.

El documento WO2010/013235 describe un método para diagnosticar una enfermedad hepática que comprende calcular una puntuación confiable que incluye los datos derivados de una prueba de aliento estándar y otros parámetros, tales como, por ejemplo, ruido fisiológico. La determinación del ruido fisiológico puede incluir el uso de un sistema de decisión experto. Sin embargo, el método del documento WO2010/013235 es específico para una prueba diagnóstica que comprende realizar una prueba de aliento estándar, y no puede adaptarse a otros métodos de diagnóstico no invasivos.

El documento WO 01/86304 A2 se refiere a un método para diagnosticar la fibrosis hepática a través del uso de un algoritmo matemático para aplicarlo a mediciones de biomarcadores. En particular, describe el uso del análisis de función discriminante para identificar cuáles son los marcadores discriminantes a seleccionar para diseñar una prueba de fibrosis no invasiva.

BMC INFECTIOUS DISEASES, vol. 10, 251, 24 de agosto de 2010, páginas 1-8, XP021074532, ISSN 1471-2334 describe el uso de una red neural artificial de tres capas, desarrollada con un algoritmo de aprendizaje Bayesiano, para diseñar un modelo no invasivo para predecir el riesgo de fibrosis hepática en pacientes que padecen de hepatitis B crónica.

En LIVER INTERNATIONAL, vol. 28, núm. 10, 20 de mayo de 2008, páginas 1352-1362, XP055046946, ISSN 1478-3223, los Inventores publicaron un artículo titulado "Evaluating and increasing the reliable diagnosis rate of blood tests for liver fibrosis in chronic hepatitis C". En el documento se hace énfasis en que la confiabilidad es un término que tiene un significado diferente en bioestadística. En este documento del estado de la técnica, los términos "confiable" o "confiabilidad" se usaron para definir intervalos de diagnóstico confiable (RDI). El RDI se trata de un diagnóstico más preciso en términos de estadios de fibrosis. Por lo tanto, en lugar de un diagnóstico amplio de estadios F2/F3/F4 (el paciente se diagnostica como que tiene una fibrosis, que es significativa (F2) o avanzada (F3) o cirrótica (F4)), se puede obtener un RDI con diagnóstico F3/F4 (el paciente se diagnostica como que tiene una fibrosis avanzada (F3) o cirrótica (F4)); en este documento del estado de la técnica, la confiabilidad se refiere a la precisión del diagnóstico.

En la presente invención, la confiabilidad no se relaciona con el RDI. Esta invención es un método y un sistema experto para mejorar la confiabilidad de una prueba, y no puede usarse para definir un RDI.

Se hace énfasis en que, en esta invención, a diferencia del documento del estado de la técnica, no se define ningún RDI; las clases de confiabilidad representan los grupos de pacientes con diferentes niveles de exactitud definidos por predictores de exactitud independientes. Por ejemplo, en pacientes con insuficiencia renal una prueba diagnóstica tiene una exactitud significativamente menor que la misma prueba en pacientes con función renal normal.

Además, se hace énfasis en que el índice de dispersión descrito en esta invención es un nuevo índice del valor de dispersión de un resultado, especialmente una puntuación, comparable con una desviación estándar.

Por lo tanto, existe la necesidad de un método para mejorar la confiabilidad de las pruebas diagnósticas.

Definiciones

En la presente invención los siguientes términos tienen los siguientes significados:

- "Enfermedad hepática" se refiere a una insuficiencia hepática aguda, una enfermedad hepática crónica, una infección por hepatitis viral especialmente una infección causada por el virus de la hepatitis B, C o D, una hepatotoxicidad, un cáncer hepático, una esteatosis, una enfermedad de hígado graso no alcohólico (NAFLD), una esteatohepatitis no alcohólica (NASH), una enfermedad autoinmunitaria, una enfermedad hepática metabólica o una enfermedad con participación secundaria del hígado. De acuerdo con una modalidad, la hepatotoxicidad es hepatotoxicidad inducida por alcohol y/o hepatotoxicidad inducida por fármaco (es decir cualquier hepatotoxicidad inducida por un compuesto xenobiótico como alcohol o fármaco).

De acuerdo con una modalidad, la enfermedad autoinmunitaria se selecciona del grupo que consiste en hepatitis autoinmunitaria (AIH), cirrosis biliar primaria (PBC) y colangitis esclerosante primaria (PSC).

De acuerdo con otra modalidad, la enfermedad hepática metabólica se selecciona del grupo que consiste en NAFLD, hemocromatosis, enfermedad de Wilson y deficiencia de alfa 1 antitripsina.

- "Sujeto" se refiere a un animal. En una modalidad, el animal es un mamífero, tal como, por ejemplo, una rata o una mascota, tal como, por ejemplo, un gato o un perro. De acuerdo con una modalidad preferida, el animal es un ser humano. De acuerdo con una modalidad, el animal, que incluye un ser humano, está en riesgo de padecer o padece una enfermedad hepática como se definió anteriormente en la presente descripción.

- "Prueba diagnóstica no invasiva" se refiere a una prueba para diagnosticar, o para evaluar la presencia y/o la gravedad de una enfermedad que puede proporcionar un dato, un índice, o una puntuación. Tal prueba puede usar la medición de biomarcadores, marcadores clínicos, datos físicos (tales como los obtenidos por la técnica Fibroscan, por ejemplo), o puntuaciones. En una modalidad, un dato puede ser un biomarcador, un marcador clínico, un dato físico, un índice o una puntuación.

Los ejemplos de biomarcadores incluyen, pero no se limitan a, glicemia, colesterol total, colesterol HDL (HDL), colesterol LDL (LDL), AST (aspartato aminotransferasa), ALT (alanina aminotransferasa), AST/ALT, AST.ALT, ferritina, plaquetas (PLT), AST/PLT, tiempo de protrombina (PT) o índice de protrombina (PI), ácido hialurónico (HA o hialuronato), hemoglobina, triglicéridos, alfa-2 macroglobulina (A2M), gamma-glutamyl transpeptidasa (GGT), urea, bilirrubina, apolipoproteína A1 (ApoA1), propéptido N-terminal de procolágeno tipo III (P3NP), gammaglobulinas (GBL), sodio (Na), albúmina (ALB), ferritina (Fer), Glucosa (Glu), fosfatasas alcalinas (ALP), YKL-40 (glicoproteína de cartílago humano 39), inhibidor tisular de metaloproteínasa de matriz 1 (TIMP-1), TGF, citoqueratina 18 y metaloproteínasa de matriz 2 (MMP-2) a 9 (MMP-9), relaciones y combinaciones matemáticas de estos.

Los ejemplos de marcadores clínicos incluyen, pero no se limitan a, peso, índice de masa corporal, edad, sexo, perímetro de cadera, perímetro abdominal o altura y la relación de estos, tal como por ejemplo perímetro de cadera/perímetro abdominal.

Los ejemplos de métodos físicos para evaluar la enfermedad hepática incluyen, pero no se limitan a, datos de imágenes médicas y mediciones clínicas, tales como, por ejemplo, medición del bazo, especialmente la longitud del bazo. De acuerdo con una modalidad, el método físico se selecciona del grupo que comprende ultrasonografía, especialmente ultrasonografía Doppler y ultrasonografía por elastometría y ultrasonografía por velocimetría (las pruebas preferidas que usan dichos datos son Fibroscan™, ARFI, VTE, imágenes supersónicas), MRI (Imágenes de Resonancia Magnética), y MNR (Resonancia Magnética Nuclear) como se usa en espectroscopía, especialmente elastometría o velocimetría por MNR. Preferentemente, los datos son datos de Evaluación de la Rigidez Hepática (LSE) o evaluación de la rigidez esplénica. De acuerdo con una modalidad preferida de la invención, los datos de los métodos físicos se emiten desde una prueba Fibroscan™. De acuerdo con una modalidad preferida de la invención, las mediciones o los datos emitidos desde la prueba Fibroscan™ son uno del índice implicado en el método de la invención.

Los ejemplos de pruebas incluyen, pero no se limitan a ELF, FibroSpect™, APRI, FIB-4, Hepascore, Fibrotest™, FibroMeter™, CirrhoMeter™, CombiMeter™, InflaMeter™.

- ELF es una prueba de sangre basada en ácido hialurónico, P3P, TIMP-1 y la edad.
- FibroSpect™ es una prueba de sangre basada en ácido hialurónico, TIMP-1 y A2M.
- APRI es una prueba de sangre basada en plaqueta y AST.
- FIB-4 es una prueba de sangre basada en plaqueta, ASAT, ALT y la edad.
- HEPAScore es una prueba de sangre basada en ácido hialurónico, bilirrubina, alfa2-macroglobulina, GGT, la edad y el sexo.
- FIBROTEST™ es una prueba de sangre basada en alfa2-macroglobulina, haptoglobina, apolipoproteína A1, bilirrubina total, GGT, la edad y el sexo.
- En conjunto FIBROMETER™ y CIRRHOMETER™ forman una familia de pruebas de sangre, cuyo contenido depende de la causa de la enfermedad hepática crónica y el objetivo de diagnóstico, y esta familia de pruebas de sangre se llama familia FM y se detalla en el Ejemplo 1.
- COMBIMETER™ es una familia de pruebas basadas en la combinación matemática de variables de la familia FM (como se detalla en el Ejemplo 1) o del resultado de una prueba de la familia FM con el resultado de la prueba FIBROSCAN™. En una modalidad, dicha combinación matemática es una regresión logística binaria.

En una modalidad, CombiMeter es una puntuación basada en la combinación matemática de datos físicos de elastometría del hígado o el bazo tales como el índice de dispersión de Fibroscan™ tal como IQR o IQR/mediana o mediana, preferentemente de la mediana de Fibroscan™ con al menos 3, preferentemente al menos 4, 5, 6, 7 o más y con mayor preferencia de 7 o 8 o 9 biomarcadores y/o datos clínicos seleccionados de la lista que comprende glicemia, colesterol total, colesterol HDL (HDL), colesterol LDL (LDL), AST (aspartato aminotransferasa), ALT (alanina aminotransferasa), AST/ALT, AST.ALT, ferritina, plaquetas (PLT), AST/PLT, tiempo de protrombina (PT) o índice de protrombina (PI), ácido hialurónico (HA o hialuronato), hemoglobina, triglicéridos, alfa-2 macroglobulina (A2M), gamma-glutamyl transpeptidasa (GGT), urea, bilirrubina, apolipoproteína A1 (ApoA1), propéptido N-terminal de procolágeno tipo III (P3NP), gammaglobulinas (GBL), sodio (Na), albúmina (ALB), ferritina (Fer), Glucosa (Glu), fosfatasas alcalinas (ALP), YKL-40 (glicoproteína de cartílago humano 39), inhibidor tisular de metaloproteínasa de matriz 1 (TIMP-1), TGF, citoqueratina 18 y metaloproteínasa de matriz 2 (MMP-2) a 9 (MMP-9), diabetes, peso, índice de masa corporal, edad, sexo, perímetro de cadera, perímetro abdominal o altura y las relaciones de estos y combinaciones matemáticas de estos.

En una modalidad, CombiMeter es una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con al menos 3, preferentemente al menos 4, 5, 6, 7 o más y con mayor preferencia de 7 o 8 o 9 biomarcadores y/o datos clínicos seleccionados de la lista que comprende alfa-2-macroglobulina (A2M), ácido hialurónico (HA), índice de protrombina (PI), plaquetas (PLT), AST, Urea, GGT, Bilirrubina (Bili), ALT, Ferritina, Glucosa, edad, sexo y peso.

En una modalidad, la prueba CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con plaquetas, PI, AST, HA, A2M, el sexo y la edad. En una modalidad preferida CombiMeter™ es una puntuación basada en la combinación matemática de la mediana de Fibroscan™ (rigidez hepática) con una prueba FibroMeter con el uso de los marcadores que se comprueban más abajo:

55

60

65

Causa	Variables													Nº	Ref		
	Edad	Sexo	peso	A2M	HA	PI	PLT	AST	Urea	GGT	Bili	ALT	Fer			Glu	FS
Virus																	
FM V 2G	x	x		x	x	x	x	x								8	2
CM V 2G	x	x		x	x	x	x	x								8	3
FM V 3G ^b	x	x		x	x		x	x								8	4
CM V 3G ^b	x	x		x	x		x	x								8	4

En una modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con plaquetas, PI, AST, HA, A2M, urea, ALT, el sexo y la edad.

5 En una modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con plaquetas, PI, AST, HA, A2M, urea, el sexo y la edad.

En una modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con plaquetas, PI, AST, HA, A2M, ALT, el sexo y la edad.

10 En una modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con PI, AST, A2M, diabetes y edad.

15 En una modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con PI, AST / ALT, A2M, plaquetas, y diabetes.

20 En una modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con PI, HA, y A2M. En otra modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con una puntuación de CirrhoMeter™, preferentemente una puntuación de CirrhoMeter^{2G}.

En otra modalidad, CombiMeter™ da lugar a una puntuación basada en la combinación matemática de la mediana de Fibroscan™ con una puntuación de FibroMeter™, preferentemente una puntuación de FibroMeter^{2G}.

25 • INFLAMETER™ es una prueba acompañante que refleja la actividad necroinflamatoria que incluye ALT, A2M, PI, y plaquetas.

De acuerdo con una modalidad preferida de la invención, al menos una prueba de la familia FM está implicada en el método de la invención.

30 • "Índice" se refiere a cualquier valor numérico obtenido mediante una prueba para diagnosticar, o para evaluar la presencia y/o la gravedad de una enfermedad, que implica la combinación matemática de al menos dos datos, preferentemente de al menos dos biomarcadores, marcadores clínicos y/o datos resultantes del método físico. En una modalidad de la invención, la combinación matemática es una combinación lineal de varios marcadores (x,y,...) como $a + bx + cy$ (a, b, c... son los coeficientes), preferentemente una regresión logística binaria o una regresión lineal múltiple. En una modalidad, un índice es un valor numérico, preferentemente un valor numérico no ligado. En una modalidad, una puntuación es un valor numérico ligado. En una modalidad, una puntuación es el resultado de la transformación de un índice no ligado en un valor ligado mediante una función matemática tal como, por ejemplo, una función logit. Preferentemente, la puntuación está en el intervalo de 0 a 1. En una modalidad, los datos combinados matemáticamente en un índice, preferentemente en una puntuación, son independientes, es decir cada uno proporciona una información que es diferente y no vinculada a la información proporcionada por los otros. Como el experto en la técnica conoce, la implementación de ELF, FibroSpect™, Hepascore, Fibrotest™, FibroMeter™, InflaMeter™, CirrhoMeter™, y CombiMeter™ da lugar a una puntuación con valores ligados mientras que APRI y FIB-4 dan lugar a un índice. De acuerdo con la presente invención, un índice o una puntuación puede ser la medición de al menos un biomarcador, al menos un marcador clínico, al menos un dato resultante de un método físico para evaluar la enfermedad hepática y/o cualquier combinación matemática de estos.

45 • "Dato Explicado" corresponde a un dato, un índice o una puntuación que se ha obtenido mediante la implementación de un método no invasivo para determinar una enfermedad. El Dato Explicado puede considerarse un dato primario. Cuando el Dato Explicado es una puntuación o un índice, este se refiere como "Índice Inicial".

50 • "Dato Explicativo" corresponde a un dato, un índice o una puntuación que es un dato, índice o puntuación, de referencia que se usa para evaluar la confiabilidad de un Dato Explicado. Un Dato Explicativo puede ser el nivel de un dato en una población de referencia. Un Dato Explicativo puede ser un dato intrínseco o extrínseco tal como por ejemplo el dato esperado en la población de referencia o un predictor de confiabilidad.

55 • "Índice Final o Dato Final": corresponde al índice o dato resultante del procesamiento del Índice Inicial o el Dato Explicado respectivamente mediante el método de la invención. En una modalidad, el Índice Final o Dato Final se refiere al último valor del resultado de la prueba diagnóstica proporcionado por el sistema experto al profesional. En una modalidad, si el sistema experto considera que el Índice Inicial o el Dato Explicado es confiable, el Índice Final o el Dato Final es igual al Índice Inicial o al Dato Explicado, respectivamente. En otra modalidad, si el sistema experto considera que el Índice Inicial o el Dato Explicado es no confiable, el Índice Final o el Dato Final es diferente al Índice Inicial o al Dato Explicado respectivamente (a menos que la diferencia se considere no significativa). El Índice Final o el Dato Final puede considerarse un dato procesado.

60 • "Dato Intrínseco": se refiere a los datos combinados matemáticamente en una prueba determinada.

• "Dato Extrínseco": se refiere a un dato que no se usa en una prueba determinada para calcular el índice correspondiente.

• "Exactitud": se refiere a la proporción de pacientes clasificados correctamente mediante una prueba diagnóstica.

• "Analizar la confiabilidad de una prueba": significa analizar la confiabilidad de una prueba diagnóstica, en donde la confiabilidad es la probabilidad de que el resultado de una prueba diagnóstica sea exacto para un sujeto individual o un

subgrupo de sujetos. Por lo tanto, de acuerdo con la invención, la confiabilidad de la prueba se mide para un sujeto individual o para un subgrupo de sujetos, mientras que la exactitud se mide para la población completa. En esta solicitud de patente, no confiable, no fiable y sin confiabilidad tienen el mismo significado, es decir indican la probabilidad de que un resultado de una prueba diagnóstica se considere inexacto (exactitud < 50 %). En una modalidad de la invención, analizar la confiabilidad de una prueba corresponde a evaluar la probabilidad de error de diagnóstico relacionada con dicha prueba, es decir comprobar la probabilidad de que los datos resultantes de dicha prueba puedan ser erróneos. Mientras más confiable es una prueba, menor es la probabilidad de error de clasificación. En consecuencia, la confiabilidad también puede corresponder a la exactitud del diagnóstico en un subconjunto de pacientes, en donde la exactitud del diagnóstico se refiere al porcentaje de pacientes con un diagnóstico correcto en la población completa.

• "Comprobar la confianza de un dato" significa identificar datos anómalos, datos inconsistentes y/o datos no homogéneos. En esta solicitud de patente, no seguro e inseguro tienen el mismo significado cuando se aplican a un dato, y se refieren a un dato anómalo, inconsistente o no homogéneo.

• "Alerta de Evento": corresponde a una alerta que se emite si se identifica un dato no seguro o una prueba no confiable. Una alerta de evento activa una respuesta. En una modalidad, una alerta de evento también puede activar el envío de un mensaje o señal de alerta al profesional o a un médico experto.

• "Sistema Experto" se refiere a un sistema informático. En una modalidad, un sistema experto incluye reglas expertas y análisis automático.

Descripción detallada

Por lo tanto, la presente invención se refiere a un método para mejorar la confiabilidad de una prueba diagnóstica no invasiva de la presencia y/o la gravedad de una enfermedad hepática, dicho método comprende:

a. recopilar un índice de diagnóstico, llamado Índice Inicial, que implica la combinación matemática de al menos dos datos, en donde el Índice Inicial se selecciona de pruebas de sangre que comprenden ELF, FibroSpect, APRI, FIB-4, Hepascore, Fibrotest, pruebas de la familia FibroMeter, y pruebas derivadas de la familia FibroMeter donde la urea se ha eliminado de los marcadores;

b. analizar la confiabilidad de cada uno de los datos combinados para recopilar el Índice Inicial mediante la identificación de si al menos un dato del Índice Inicial recopilado en la etapa a) es un dato anómalo, un dato inconsistente, o un dato no homogéneo, o es responsable de una disminución en el Índice de Dispersión mayor que la observada con otros datos:

i. con la comparación de cada dato con el dato esperado en una población de referencia, o

ii. en vista del(de los) predictor(es) de confiabilidad intrínseco(s) o extrínseco(s), o

iii. mediante el cálculo del Índice de Dispersión del Índice Inicial recopilado en la etapa a), un Índice de Dispersión alto significa datos homogéneos del Índice Inicial, y la realización de una serie de cálculos del Índice de Dispersión del Índice Inicial que comprende n datos donde se eliminan 1 a (n-2) datos del Índice Inicial, lo que da lugar a la clasificación de los datos que más reducen el Índice de Dispersión;

c. si un dato es anómalo, inconsistente o un dato no homogéneo, o un dato que reduce un Índice de Dispersión, generar una Alerta de Evento;

d. si se genera una Alerta de Evento, calcular nuevos índices, donde el dato anómalo, inconsistente y/o no homogéneo, o un dato responsable del Índice de Dispersión más bajo, se suprime para obtener de este modo un Índice Alternativo, o donde el dato anómalo, inconsistente y/o no homogéneo se sustituye por su valor medio para obtener de este modo un Índice Estimado o, si al menos dos datos son anómalos, inconsistentes o no homogéneos, el más discordante se suprime y el(los) otro(s) se sustituye(n) por su valor medio para obtener de este modo un Índice Mixto;

e. reemplazar el Índice Inicial que comprende un dato anómalo, inconsistente y/o no homogéneo o un dato que afecta el Índice de Dispersión, con un Índice Alternativo, un Índice Estimado, o un Índice Mixto.

Una modalidad preferida de la invención se refiere al método como se describió anteriormente en la presente descripción, en donde:

• en la etapa a) se recopila un Índice Inicial, dicho Índice Inicial es una prueba de la familia FibroMeter, o una prueba derivada de la familia FibroMeter donde la urea se ha eliminado de los marcadores;

• en la etapa b) el Índice de Dispersión del Índice Inicial recopilado en la etapa a) se calcula, para identificar de este modo cuál es el dato del Índice Inicial recopilado en la etapa a) que más disminuye el Índice de Dispersión, mediante la realización de una serie de cálculos del Índice de Dispersión del Índice Inicial que comprende n datos donde se eliminan 1 a (n-2) datos del Índice Inicial;

• en la etapa e) el Índice Inicial que comprende un dato que más reduce el Índice de Dispersión se reemplaza con un Índice Alternativo donde el dato que reduce el Índice de Dispersión se elimina.

En una modalidad preferida del método como se describió anteriormente en la presente descripción, la etapa c) se realiza después de calcular si el Índice de Dispersión del Índice Inicial corresponde a los de sus deciles intermedios, preferentemente del 2^{do} al 9^{no} decil, de la población de referencia.

En una modalidad preferida del método de la invención, cuando un dato es un dato anómalo, la Alerta de Evento es una Alerta de Dato Anómalo; cuando un dato es un dato inconsistente, la Alerta de Evento es una Alerta de Dato Inconsistente; cuando un dato es un dato no homogéneo, la Alerta de Evento es una Alerta de Dato no Homogéneo.

- 5 En una modalidad preferida del método de la invención, las Alertas de Evento emitidas en la etapa c) se analizan preliminarmente, antes de la etapa d), y en donde dicho tratamiento comprende:
- la supresión de una o más Alertas de Evento cuando el dato que ha activado la Alerta de Evento se encuentra confiable, y/o
 - cuando se emiten varias Alertas de Evento, la priorización de dichas Alertas de Evento para identificar una Alerta de Evento principal de acuerdo con su impacto sobre el Índice Inicial.
- 10

En una modalidad preferida del método de la invención, los predictores de confiabilidad se seleccionan del grupo que comprende urea, ALT, AST, puntuación de CirrhoMeter2G, puntuación de FibroMeter2G, clases de Fibroscan, clases de FibroMeter, clases de CirrhoMeter, mediana de Fibroscan, IQR, IQR/M, plaquetas, A2M, relación urea/ Fibroscan e Índice de Dispersión del Índice Inicial.

15

En una modalidad preferida del método de la invención, en la etapa b) la confiabilidad de cada dato se analiza mediante la identificación de si al menos un dato del Índice Inicial recopilado en la etapa a) es un dato anómalo, un dato inconsistente, o un dato no homogéneo, o es responsable de una disminución en el Índice de Dispersión mayor con la observada con otros datos con comparación de cada dato con el dato esperado en una población de referencia, o en vista del(de los) predictor(es) de confiabilidad intrínseco(s) o extrínseco(s).

20

En una modalidad preferida, el método de la invención comprende en la etapa e) reemplazar el Índice Inicial con el índice más confiable seleccionado del grupo que consiste en un Índice Alternativo, un Índice Estimado, o un Índice Mixto.

25

La presente invención también se refiere a un sistema experto que implementa el método de la invención como se describió anteriormente en la presente descripción.

La presente invención se refiere además a un software que implementa el método de la invención como se describió anteriormente en la presente descripción.

30

En una modalidad de la invención, etapa c., es decir proporcionar la información de si el Dato Explicado es confiable o no, y si no es confiable, proporcionar un Índice Final confiable incluye:

- 35 a. generar una Alerta de Evento siempre que un dato sea no seguro y/o siempre que un Dato Explicado sea no confiable de acuerdo con su Dato Explicativo o con predictores de confiabilidad; y cuando se emite una Alerta de Evento, realizar un análisis preliminar de la Alerta de Evento,
- b. activar una respuesta a la presencia o la ausencia de alerta de evento, que puede ser:
- 40 i. la emisión del Dato Explicado como Dato Final si no se produce o se mantiene una Alerta de Evento después del tratamiento previo o
- ii. el cálculo y la emisión de un Dato Final más confiable que el Dato Explicado, junto con un comentario.

En una modalidad, el método de la invención incluye:

- 45 • identificar si un dato es un dato anómalo, y emitir una Alerta de Evento que es una Alerta de Dato Anómalo, y/o
- identificar si un dato es un dato inconsistente y emitir una Alerta de Evento que es una Alerta de Dato Inconsistente, y/o
- identificar si un dato es un dato no homogéneo y emitir una Alerta de Evento que es una Alerta de Dato no Homogéneo.
- 50

En una modalidad, las Alertas de Evento emitidas se tratan, y el tratamiento comprende:

- la supresión de una o más Alertas de Evento, y/o
 - cuando se emiten varias Alertas de Evento, la priorización de dichas Alertas de Evento para identificar una Alerta de Evento principal.
- 55

En una modalidad de la invención, activar la respuesta comprende calcular nuevos índices, preferentemente nuevas puntuaciones después de emitir y tratar previamente la(s) Alerta(s) de Evento.

60 En una modalidad, recopilar un Dato Explicado significa recopilar un índice de diagnóstico, llamado Índice Inicial, preferentemente una puntuación, con mayor preferencia una puntuación seleccionada de pruebas de sangre que comprenden ELF, FibroSpect™, APRI, FIB-4, Hepascore, Fibrotest™, o una puntuación de la familia FibroMeter, tal como por ejemplo FibroMeter™ y CirrhoMeter™ o CombiMeter™; o una prueba derivada de la familia FibroMeter, donde la urea se ha eliminado de los marcadores; u otra prueba diagnóstica; dicho índice implica la combinación matemática de al

menos dos datos, preferentemente de al menos dos biomarcadores, marcadores clínicos y/o datos resultantes del método físico; dicho índice se refiere como Índice Inicial.

5 En una modalidad donde el Dato Explicado es un Índice Inicial cuando se emite una Alerta de Evento, se calcula al menos uno de los índices nuevos siguientes:

- un Índice Alternativo, preferentemente una Puntuación Alternativa, en donde el dato que ha emitido la Alerta de Evento se suprime, y/o
- un Índice Estimado, preferentemente una Puntuación Estimada, en donde el dato que ha emitido la Alerta de Evento se sustituye por su valor central, tal como, por ejemplo, por su valor medio.

10 En otra modalidad donde el Dato Explicado es un Índice Inicial, cuando se emiten al menos dos Alertas de Evento, al menos uno de los índices siguientes, preferentemente puntuaciones, se calcula:

- 15 • un Índice Mixto, preferentemente una Puntuación Mixta, en donde el dato que ha emitido la Alerta de Evento Principal se suprime y el(los) dato(s) que ha(n) emitido la(s) Alerta(s) de Evento Secundaria(s) se sustituye(n) por su valor central, tal como, por ejemplo, por sus valores medios, y/o
- un Índice Estimado, preferentemente una Puntuación Estimada, en donde todos los datos que han emitido las Alertas de Evento se sustituyen por su valor central, tal como, por ejemplo, por sus valores medios.

20 En una modalidad de la invención donde el Dato Explicado es un Índice Inicial, activar una respuesta a dicha(s) Alerta(s) de Evento comprende además identificar el índice más confiable, preferentemente puntuación, entre el índice inicial y los nuevos índices.

25 En una modalidad de la invención, el Dato Explicativo es al menos un predictor de confiabilidad. De acuerdo con la invención, los predictores de confiabilidad se seleccionan del grupo de datos, llamados variables, incluidos en o derivados de pruebas no invasivas, como se indica más abajo, preferentemente FibroMeter™, InflaMeter™ y Fibroscan™, y con mayor preferencia urea, ALT, AST, puntuación de CirrhoMeter^{2G}, puntuación de FibroMeter^{2G}, clases de Fibroscan™, clases de FibroMeter™, clases de CirrhoMeter™, mediana de Fibroscan™, IQR, IQR/M, plaquetas, A2M, relación urea/Fibroscan™ e Índice de Dispersión del Índice Inicial.

30 En una modalidad, el Dato Explicado es un índice o puntuación ("Índice Inicial"), el Índice de Dispersión del Índice Inicial se calcula y la confiabilidad de cada dato incluido en el Índice Inicial se evalúa por su impacto sobre un Índice de Dispersión. Evaluar el impacto de un dato sobre un Índice de Dispersión significa realizar una serie de cálculos del Índice de Dispersión del Índice Inicial que comprende n datos donde se eliminan 1 a (n-2) datos del Índice Inicial, lo que da lugar a la identificación y clasificación de los datos que más reducen el Índice de Dispersión; el Índice de Dispersión más bajo en la serie indica qué dato eliminar. Por lo tanto, el método de la invención incluye esta evaluación, y una vez que los datos se identifican, el cálculo de Índices Alternativos donde los datos que reducen el Índice de Dispersión se eliminan (respuesta activada).

35 En una modalidad, el método de la invención, para proporcionar una prueba diagnóstica no invasiva confiable de la presencia y/o la gravedad de una enfermedad, comprende:

- 45 a. recopilar un índice de diagnóstico, llamado Índice Inicial, preferentemente una puntuación, con mayor preferencia una puntuación seleccionada de pruebas de sangre que comprenden ELF, FibroSpect™, APRI, FIB-4, Hepascore, Fibrotest™, o una puntuación de la familia FibroMeter, tal como por ejemplo FibroMeter™ y CirrhoMeter™ o CombiMeter™; o una prueba derivada de la familia FibroMeter, donde la urea se ha eliminado de los marcadores; u otra prueba diagnóstica; dicho índice implica la combinación matemática de al menos dos datos, preferentemente de al menos dos biomarcadores, marcadores clínicos y/o datos resultantes del método físico; dicho índice se refiere como Índice Inicial,
- 50 b. analizar la confiabilidad de cada dato mediante la identificación de si al menos un dato del índice recopilado en la etapa a) es un dato anómalo, un dato inconsistente; y/o un dato no homogéneo, o es el que disminuye el Índice de Dispersión más que lo observado con otros datos:

- 55 i. con la comparación de cada dato con el dato esperado en la población de referencia, o
- ii. con la predicción del índice de la etapa a) a predictor(es) de confiabilidad intrínseco(s) o extrínseco(s), o
- iii. mediante el cálculo del Índice de Dispersión del Índice Inicial y después la realización de una serie de cálculos del Índice de Dispersión de los Índices Alternativos correspondientes al Índice Inicial que comprende n datos donde se eliminan 1 a (n-2) datos, lo que da lugar a la identificación y clasificación de los datos que más reducen el Índice de Dispersión,
- 60 iv. mediante la ubicación del Índice en una clase de confiabilidad de la clasificación de confiabilidad en base a los datos obtenidos en una población de referencia,

c. si un dato es un dato anómalo, inconsistente y/o no homogéneo, o un dato que reduce un Índice de Dispersión, generar una Alerta de Evento,

65

d. si se genera una Alerta de Evento, si no se ha hecho ya, calcular nuevos índices, preferentemente nuevas puntuaciones, dichos nuevos índices son:

- 5 i. el mismo índice que el Índice Inicial, pero donde el dato anómalo, inconsistente y/o no homogéneo se suprime (Índice Alternativo) o se sustituye por su valor central, preferentemente su valor medio (Índice Estimado) o, si al menos dos datos son anómalos, inconsistentes o no homogéneos, el más discordante se suprime y el(los) otro(s) se sustituye(n) por su valor central, preferentemente su valor medio (Índice Mixto), o
- 10 ii. el índice correspondiente al Índice Alternativo donde el dato responsable del Índice de Dispersión más bajo se elimina,
- e. reemplazar el Índice Inicial que comprende un dato anómalo, inconsistente y/o no homogéneo o un dato que disminuye el Índice de Dispersión, con un Índice Alternativo, un Índice Estimado, o un Índice Mixto.

En una modalidad, el método de la invención comprende:

- 15 • recopilar un índice de la familia FibroMeter, tal como por ejemplo FibroMeter™ y CirrhoMeter™ o CombiMeter™; o una prueba derivada de la familia FibroMeter, donde la urea se ha eliminado de los marcadores; este índice se refiere como Índice Inicial,
- 20 • calcular el Índice de Dispersión del Índice Inicial recopilado en la etapa a); identificar cuál es el dato que afecta el Índice de Dispersión, mediante la realización de una serie de cálculos del Índice de Dispersión del Índice Inicial que comprende n datos donde se eliminan 1 a (n-2) datos del Índice Inicial,
- reemplazar el Índice Inicial que comprende un Índice Alternativo, donde el(los) dato(s) que reduce(n) el DI se elimina(n).

25 En una modalidad la etapa c) se realiza después de calcular si el Índice de Dispersión del Índice Inicial corresponde a los de sus deciles intermedios, preferentemente los deciles 2-9, 2-8, 2-7, 3-9, 3-8, 3-7, 4-9, 4-8, 4-7, de la población de referencia (deciles - 10 %- de acuerdo con los cuantiles de pacientes del Índice de Dispersión del Índice Inicial).

30 En una modalidad de la invención, el resultado de la prueba es una puntuación o un índice, y la confiabilidad de la prueba se analiza en vista del dato intrínseco de dicha puntuación o índice. En una segunda modalidad de la invención, el resultado de la prueba es una puntuación o un índice y la confiabilidad de la prueba se analiza en vista del dato extrínseco (con o sin dato intrínseco).

[Primera Modalidad - ES2G]

- 35 En una modalidad, la presente invención se refiere a un método para mejorar la confiabilidad de las pruebas diagnósticas no invasivas que comprende:
- a1. - recopilar la medición de al menos dos datos seleccionados de al menos un biomarcador, al menos un marcador clínico, al menos un dato resultante de un método físico para evaluar la enfermedad y/o al menos una puntuación o índice, y
- 40 a2. - calcular un Índice Inicial (dato explicado) con las mediciones o índice o puntuación recopilados en la etapa (a), preferentemente por medio de una prueba de la familia FibroMeter™, y
- b1. - comprobar la confianza de cada dato usado en la prueba, y/o
- b2. - analizar la confiabilidad del índice inicial, por ejemplo, en vista del dato intrínseco y/o en vista de un predictor de confiabilidad,
- 45 c1. - generar una alerta de evento siempre que un dato sea no seguro y/o siempre que el Índice Inicial sea no confiable,
- d1. - activar una respuesta a dicha alerta de evento, que incluye proporcionar un Índice Final.

Recopilación de los datos (etapa a) y cálculo del Índice Inicial (etapa b)

50 En una modalidad, el método comienza con la recopilación a partir de un sujeto de la medición de al menos dos datos seleccionados de al menos un biomarcador, al menos un marcador clínico, al menos un dato resultante de un método físico para evaluar la enfermedad y/o una puntuación o un índice.

55 En esta modalidad, ventajosamente, se miden preferentemente al menos 3 datos, con mayor preferencia al menos 4 datos, y aun con mayor preferencia al menos 5 datos. En una modalidad de la invención, se recopilan 2 datos, preferentemente 3, 4, 5, 6, o 7 datos o más.

60 Los al menos dos datos se combinan matemáticamente para obtener un índice, preferentemente una puntuación, con mayor preferencia una puntuación seleccionada de ELF, FibroSpect™, APRI, FIB-4, Hepascore, Fibrotest™, o una prueba de la familia FibroMeter, tal como por ejemplo FibroMeter™ y CirrhoMeter™ o CombiMeter™.

Dado que este índice se calcula antes de comprobar la confianza de cada dato, en la presente descripción se refiere como Índice Inicial.

65 En esta modalidad, la confiabilidad de la prueba puede analizarse en vista del dato intrínseco del índice.

Identificación de datos anómalos, datos inconsistentes y/o datos no homogéneos.

Dato anómalo

5 Como se usa en la presente descripción, el término anómalo que antecede al término dato ("dato anómalo") se refiere a un valor numérico que no entra dentro del intervalo de los valores numéricos usuales de este dato medidos en una población de referencia; generalmente un dato anómalo es el resultado de un error de medición o de un registro erróneo.

10 En una modalidad, un dato anómalo es menor, preferentemente dos veces menor, con mayor preferencia tres veces menor y aun con mayor preferencia 5 o 10 veces menor que el dato esperado más pequeño medido en la población de referencia.

15 En otra modalidad, un dato anómalo es mayor, preferentemente dos veces mayor, con mayor preferencia tres veces mayor, y aun con mayor preferencia 5 o 10 veces mayor que el dato esperado más alto medido en la población de referencia.

20 En una modalidad, la población de referencia es una población de al menos 500 sujetos, preferentemente al menos 1000 sujetos. En una modalidad, la población de referencia comprende sujetos sanos. En otra modalidad, la población de referencia comprende sujetos no sanos, en donde dichos sujetos no sanos tienen preferentemente una enfermedad hepática. En otra modalidad, la población de referencia comprende tanto sujetos sanos como no sanos.

El Ejemplo 3 muestra un ejemplo de datos que se consideran anómalos para algunos biomarcadores y datos clínicos.

Dato inconsistente

25 Como se usa en la presente descripción, un "dato inconsistente" se usa con referencia a un índice, y es un dato que es externo al intervalo de consistencia para dicho dato, medido de la siguiente manera:

30 • fijar arbitrariamente los valores para todos los datos combinados matemáticamente dentro de dicho índice, dicha fijación arbitraria de valores es de manera tal que el resultado del índice es un valor alto o bajo para este índice. Como un ejemplo, si el índice es una puntuación en el intervalo de 0 a 1, los valores de cada dato implicado en la puntuación se fijan arbitrariamente para conducir a una puntuación en el intervalo de 0,6 a 0,9, preferentemente de 0,7 o 0,8 (valor alto), o para conducir a una puntuación en el intervalo de 0,1 a 0,4, preferentemente de 0,2 o 0,3 (valor bajo),

35 • variar arbitrariamente el valor de uno de los datos, mientras que los otros son constantes; y determinar la influencia de dicha variación sobre el índice,

• fijar dos "umbrales de influencia" para dicho dato, es decir un umbral alto y uno bajo, correspondientes a un valor del dato que se considera inductor de error, es decir que afecta el índice en términos de confiabilidad, es decir colocar el índice en riesgo de proporcionar un resultado falso positivo o un falso negativo.

40 De acuerdo con la invención, un intervalo de consistencia de un dato es dependiente del índice. Por ejemplo, el intervalo de consistencia medido para el biomarcador TIMP-1 dentro de una puntuación de ELF puede ser diferente del intervalo de consistencia de TIMP-1 dentro de una puntuación de FibroSpect™.

45 En una modalidad, el índice es una puntuación y los umbrales de influencia son los valores del dato que conducen a una variación de la puntuación de más de $\pm 0,05$, $\pm 0,1$, $\pm 0,2$, $\pm 0,3$ o $\pm 0,35$.

50 En los métodos de diagnóstico, cuando se mide una puntuación, una puntuación de 0 es usualmente indicativa de la ausencia de la enfermedad (o 0 % de probabilidad o ausencia del objetivo de diagnóstico), mientras que una puntuación de 1 representa la forma más grave de la enfermedad (o 100 % de probabilidad o presencia del objetivo de diagnóstico).

55 Por lo tanto, en una modalidad, un dato que conduce a una disminución o a un aumento de la puntuación de más de 0,05, 0,1, 0,2, 0,3 o 0,35 cuando los otros datos que se usan para el cálculo de esta puntuación se fijan arbitrariamente para conducir a un valor bajo, tal como, por ejemplo, 0,2, o un valor alto, tal como, por ejemplo, 0,8, puede conducir a un falso negativo o falso positivo.

60 Una inconsistencia de falso positivo es un valor de dato que da lugar a una desviación positiva (aumento) clínicamente significativa del resultado de la prueba proporcionado por los otros datos. Por ejemplo, en un índice que incluye 8 marcadores, con un resultado de puntuación en 0,6 proporcionado por 7 marcadores, el 8vo marcador será la causa de un resultado positivo si su entrada conduce a un resultado de puntuación $\geq 0,75$.

65 Una inconsistencia de falso negativo es un valor de dato que da lugar a una desviación negativa (disminución) clínicamente significativa del resultado de la prueba proporcionado por los otros datos. Por ejemplo, en un índice que incluye 8 marcadores, con un resultado de puntuación en 0,6 proporcionado por 7 marcadores, el 8vo marcador será la causa de un resultado negativo si su entrada proporciona un resultado de puntuación $\leq 0,45$.

En ambos casos, el cálculo de la desviación significativa para un marcador / dato se proporciona mediante simulaciones. Las simulaciones se realizan con diferentes escenarios:

- Dos valores de puntuación: a un valor de puntuación bajo (0,1 a 0,3) y uno alto (0,6 a 0,8),
- Dos grados de significación para la desviación: significación baja y alta, por ejemplo, de 0,15 a < 0,30 y $\geq 0,30$, respectivamente, para una puntuación de 0 a 1.

En una modalidad, el intervalo de inconsistencia de falso negativo puede dividirse en varios intervalos, en donde el grado de la inconsistencia aumenta. En una modalidad, el intervalo de inconsistencia de falso negativo puede dividirse en dos intervalos:

- una zona verde, correspondiente a los valores del dato que conducen a una desviación negativa (disminución) de la puntuación en el intervalo de $0,15 \leq 0,30$, cuando los valores de otros datos que se usan para el cálculo de la puntuación se fijan arbitrariamente (preferentemente imitando un escenario clínicamente verosímil) para conducir a una puntuación baja, tal como, por ejemplo, una puntuación de 0,3 y a una puntuación alta, tal como, por ejemplo, una puntuación de 0,7, y
- una zona azul, correspondiente a los valores del dato que conducen a una desviación negativa (disminución) de la puntuación $\geq 0,3$, cuando los valores de otros datos que se usan para el cálculo de la puntuación se fijan arbitrariamente para conducir a una puntuación baja, tal como, por ejemplo, una puntuación de 0,3 y a una puntuación alta, tal como, por ejemplo, una puntuación de 0,7.

En una modalidad de la invención, un dato de la zona verde se considerará menos inconsistente que un dato de la zona azul.

En una modalidad, el intervalo de inconsistencia de falso positivo puede dividirse en varios intervalos, en donde el grado de la inconsistencia aumenta. En una modalidad, el intervalo de inconsistencia de falso positivo puede dividirse en dos intervalos:

- una zona amarilla, correspondiente a valores del dato que conducen a una desviación positiva (aumento) de la puntuación en el intervalo de $0,15 \leq 0,30$, cuando los valores de otros datos que se usan para el cálculo de la puntuación se fijan arbitrariamente para conducir a una puntuación baja, tal como, por ejemplo, una puntuación de 0,3 y a una puntuación alta, tal como, por ejemplo, una puntuación de 0,7, y
- una zona roja, correspondiente a valores del dato que conducen a una desviación positiva (aumento) de la puntuación $\geq 0,3$, cuando los valores de otros datos que se usan para el cálculo de la puntuación se fijan arbitrariamente para conducir a una puntuación baja, tal como, por ejemplo, una puntuación de 0,3 y a una puntuación alta, tal como, por ejemplo, una puntuación de 0,7.

En una modalidad de la invención, un dato de la zona amarilla se considerará menos inconsistente que un dato de la zona roja.

El Ejemplo 3 muestra un ejemplo de intervalos de consistencia, intervalos de inconsistencia de falso positivo e intervalos de inconsistencia de falso negativo para algunos biomarcadores y datos clínicos de FibroMeter™.

Dato no homogéneo

Como se usa en la presente descripción, un "dato no homogéneo" es un dato que es no homogéneo es decir no concordante o no consistente con los otros datos medidos en una prueba para calcular un índice o una puntuación.

En una modalidad, un dato se considera no homogéneo con otros cuando la eliminación de dicho dato conduce a un 10 % de variación del índice (o puntuación) real, con mayor preferencia a un 20 % de variación, preferentemente a un 30 % de variación del índice (o puntuación). Si no es así, el dato se considera homogéneo.

En otra modalidad, un dato se considera no homogéneo, cuando está por debajo del umbral del índice de dispersión de dicha prueba, que es conocido o calculado por el experto. Un dato se considera homogéneo, cuando está por encima del umbral del índice de dispersión de dicha prueba.

Índice de Dispersión

En una modalidad de la invención, la confiabilidad de una prueba se comprueba mediante la medición del índice de dispersión (DI), de los datos o las pruebas que se usan en el método (la medición de biomarcador(es), marcador(es) clínico(s), datos resultantes de un método físico para evaluar la enfermedad y/o cualquier combinación matemática de estos).

El índice de dispersión es un índice que representa la dispersión de dichos datos, que, en una modalidad, se combinan matemáticamente para obtener un índice, por ejemplo, el Índice Inicial. En una modalidad, cuando el índice comprende la combinación de n datos, se miden los índices alternativos de diagnóstico con el uso de una combinación matemática

de $n-1$ de estos n datos para determinar el índice de dispersión. Un índice de dispersión de 1 (100 %) corresponde a una homogeneidad perfecta de los datos, y un índice de dispersión de 0 corresponde a la dispersión total de los datos.

5 Un valor de índice se considera no confiable cuando se alcanza un umbral de dispersión. El umbral de dispersión del DI se calcula gracias al gráfico de DI contra la diferencia absoluta δ_i entre el índice inicial y final como se definió anteriormente (como se muestra en el Ejemplo 2 y en la Figura 1).

Por ejemplo, con FibroMeter²⁰ el umbral de confiabilidad del DI es 0,81.

10 Por lo tanto, esta invención demuestra que:

- si el DI está por encima del umbral de dispersión, (es decir cuando los datos son homogéneos), el Índice Inicial se afecta poco, lo que significa que el Índice Inicial es confiable,
- si el DI está por debajo del umbral de dispersión, el índice inicial no es confiable y se requieren Índices Alternativos o Estimados.

En una modalidad, dicho índice de dispersión puede medirse de la siguiente manera:

20 X es una muestra de los n valores x_i del índice, en donde los valores de x_i están en el intervalo de a hasta b ; por lo tanto

$$RI = s(X) / s_m(a, b)$$

en donde $s(X)$ es la desviación estándar empírica de X y en donde $s_m(a, b)$ es la desviación estándar empírica máxima en $[a, b]$ para una muestra que comprende n valores si n es impar, o $n+1$ valores si n es par.

25 En una modalidad, cuando se usa el índice de dispersión, el método de la invención comprende las siguientes etapas:

- (a) el índice de dispersión se calcula como se describió anteriormente en la presente descripción, mediante el uso de todos los n datos combinados matemáticamente en el índice,
- 30 (b) se determina el umbral del índice de dispersión, por debajo del cual la tasa de exactitud de la prueba diagnóstica disminuye significativamente,
- (c) cuando se alcanza este umbral de dispersión, los n índices alternativos (índices correspondientes a la combinación matemática de $n-1$ datos como se describió anteriormente en la presente descripción) se comparan,
- (d) se determina el índice alternativo responsable de la mayor variación en el índice de dispersión, es decir aquel con el valor más disperso en comparación con el Índice Inicial (o con el índice anterior para la etapa f), y el dato se excluye,
- 35 (e) se calcula un nuevo índice de dispersión sin el dato excluido de la etapa (d),
- (f) si todavía se alcanza el umbral de dispersión, se repiten las etapas (b) a (e) hasta que no se alcance el umbral del índice de dispersión o hasta que el número de datos restantes sea menor que 4.

40 En una modalidad, para un índice resultante de la combinación de n datos, el número máximo de datos excluidos es menor que o igual a $n/3$, preferentemente menor que o igual a $n/4$.

El índice inicial o alternativo que conduce a un índice de dispersión que no alcanza el umbral de dispersión se considera el Índice Final.

45 Por lo tanto, dicho Índice Final puede usarse finalmente para evaluar la presencia y/o la gravedad de una enfermedad hepática en el sujeto. En una modalidad, dicha evaluación incluye la clasificación del sujeto en una clase de una clasificación de estadio de fibrosis, es decir una clase de una clasificación en donde cada clase se asocia con un estadio de fibrosis, tal como, por ejemplo, uno o más estadios F de Metavir.

50 En una modalidad, el método de la invención comprende emitir un comentario de confiabilidad, que incluye uno de:

- indicar el número de Alertas de Evento,
- indicar el dato que ha emitido una Alerta de Evento o la Alerta de Evento principal,
- 55 • indicar cuál(es) de los datos se excluye(n),
- indicar el Índice Final.

Alerta de Evento

60 Tipos de Alerta de Evento

De acuerdo con una modalidad de la invención, una alerta de evento se emite si se identifica un dato no seguro.

65

En una modalidad, dicha alerta de evento puede corresponder a la identificación de un dato anómalo (Alerta de Dato Anómalo), de un dato inconsistente (Alerta de Dato Inconsistente) y/o de un dato no homogéneo (Alerta de Dato no Homogéneo).

5 En una modalidad, un dato puede generar tanto una Alerta de Dato Inconsistente como una Alerta de Dato no Homogéneo.

En una modalidad, de acuerdo con los intervalos de inconsistencia descritos anteriormente en la presente descripción, una Alerta de Dato de Inconsistencia puede ser:

- 10
- una Alerta de Dato de Inconsistencia de Falso Positivo (si el dato pertenece al intervalo de inconsistencia de falso positivo), o
 - una Alerta de Dato de Inconsistencia de Falso Negativo (si el dato pertenece al intervalo de inconsistencia de falso negativo).

15 En una modalidad, una Alerta de Dato de Inconsistencia de Falso Positivo puede ser:

- una Alerta Roja (si el dato pertenece a la zona roja de inconsistencia de falso positivo), o
- una Alerta Amarilla (si el dato pertenece a la zona amarilla de inconsistencia de falso positivo).

20 En una modalidad, una Alerta de Dato de Inconsistencia de Falso Negativo puede ser:

- una Alerta Azul (si el dato pertenece a la zona roja de inconsistencia de falso negativo), o
- una Alerta Verde (si el dato pertenece a la zona amarilla de inconsistencia de falso negativo).

25 **Análisis Preliminar de Alertas de Evento**

En una modalidad, un análisis preliminar de Alertas de Evento puede llevarse a cabo antes de activar una respuesta a dicha alerta de evento.

30 En una modalidad, dicho análisis preliminar puede corresponder a la supresión de una Alerta de Evento (es decir el dato que ha emitido la Alerta de Evento finalmente se considera confiable) o a la priorización de dichas alertas, para definir de este modo una "Alerta de Evento principal" y "Alerta(s) de Evento secundaria(s)".

Supresión de Alertas de Evento

35 En una modalidad, cada dato se considera de manera independiente: cuando un dato genera una Alerta de Dato Inconsistente pero no una Alerta de Dato no Homogéneo, es decir cuando el dato es inconsistente pero homogéneo con los otros datos combinados matemáticamente dentro del índice, la Alerta de Dato Inconsistente se suprime posteriormente (es decir no conduce a una variación de la puntuación). Esta modalidad se repite para cada dato del índice.

40 En otra modalidad de la invención, varios datos se consideran de manera simultánea: cuando al menos dos, preferentemente al menos tres, con mayor preferencia al menos cuatro Alertas de Inconsistencia de Dato del mismo tipo, es decir 2 o 3 o 4 o más Alertas de Dato de Inconsistencia de Falso Positivo o Falso Negativo se activan de manera simultánea, dichas Alertas de Dato Inconsistente pueden suprimirse posteriormente (es decir no conducen a una variación de la puntuación). En una modalidad de la invención, cuando al menos dos, preferentemente al menos tres, con mayor preferencia al menos cuatro Alertas de Dato de Inconsistencia Negativa se activan de manera simultánea, dichas Alertas de Dato Inconsistente no se suprimen.

50 El experto en la técnica, en vista de lo anterior, es capaz de redactar reglas adicionales para el tratamiento previo de los datos (tales como, por ejemplo, suprimir una Alerta de Evento individual cuando se encuentra implicado un biomarcador o un dato clínico específico (tal como, por ejemplo, edad, sexo o ASAT); nunca suprimir Alertas de Dato de Inconsistencia de Falso Negativo, y similares). Los ejemplos de tales Reglas cuando se lleva a cabo una prueba FibroMeter™ se muestran en el Ejemplo 4.

55 **Priorización:**

En una modalidad, si se emiten varias Alertas de Evento, puede llevarse a cabo un análisis preliminar que comprende la priorización de dichas alertas, para definir de este modo una "Alerta de Evento principal" y "Alerta(s) de Evento secundaria(s)".

60 Los ejemplos de Reglas de análisis preliminar cuando se emiten varias Alertas de Dato Inconsistente se muestran más abajo:

- 65
- una Alerta de Dato de Inconsistencia de Falso Positivo puede considerarse más importante (es decir "principal") que una Alerta de Dato de Inconsistencia de Falso Positivo (que por lo tanto será "secundaria"),
 - una Alerta Azul puede ser más importante que una Alerta Verde,

- una Alerta Roja puede ser más importante que una Alerta Amarilla,
- si se emiten dos Alertas de Evento similares, tales como, por ejemplo, dos Alertas Azules, la Alerta principal puede ser la emitida por el dato más alejado de su valor central como la media o con el mayor impacto sobre el valor de puntuación,
- etc...

Otros ejemplos de reglas de análisis preliminar de Alertas de Evento cuando se lleva a cabo la prueba FibroMeter™ se muestran en el Ejemplo 4.

10 Nuevos índices - Tratamiento de los datos

En una modalidad de la invención, nuevos índices, preferentemente nuevas puntuaciones, se calculan después de emitir y tratar previamente las Alertas de Evento, si algunas Alertas de Evento subsisten después de dicho análisis preliminar.

15 En una modalidad, se emite una Alerta de Evento individual, y al menos uno de los índices siguientes, preferentemente puntuaciones, se calcula:

- un Índice Alternativo, preferentemente una Puntuación Alternativa, en donde el dato que ha emitido la Alerta de Evento se suprime, y/o
- un Índice Estimado, preferentemente una Puntuación Estimada, en donde el dato que ha emitido la Alerta de Evento se sustituye por su valor central, tal como, por ejemplo, por su valor medio.

En una modalidad, se emiten al menos dos Alerta de Evento, y al menos uno de los índices siguientes, preferentemente puntuaciones, se calcula:

- un Índice Mixto, preferentemente una Puntuación Mixta, en donde el dato que ha emitido la Alerta de Evento Principal se suprime y el(los) dato(s) que ha(n) emitido la(s) Alerta(s) de Evento Secundaria(s) se sustituye(n) por su(s) valor(es) central(es) como la media, y/o
- un Índice Estimado, preferentemente una Puntuación Estimada, en donde todos los datos que han emitido las Alertas de Evento se sustituyen por sus valores centrales, tales como, por ejemplo, por sus valores medios.

En una modalidad de la invención, el valor medio de un dato corresponde al valor medio (preferentemente la media aritmética) para dicho dato medido en una población de referencia.

35 Respuesta activada

En una modalidad de la invención, la respuesta puede ser la emisión del Índice Inicial como Índice Final si no queda ninguna Alerta de Evento después del tratamiento previo o el cálculo de nuevos índices, preferentemente nuevas puntuaciones y la emisión de un Índice Final diferente al Índice Inicial (Índice Alternativo o Índice Estimado), preferentemente junto con un comentario.

Selección del índice más confiable

En una modalidad, se emite una Alerta de Evento individual y la selección del índice más adecuado se lleva a cabo mediante la comparación del valor del Índice Inicial (II) con el(los) valor(es) del Índice Alternativo (AI) y/o del Índice Estimado (EI).

En una modalidad, si $|II - AI|$ es superior a δ_2 , el AI se considera el índice más confiable, en donde δ_2 es la diferencia clínicamente significativa. En una modalidad, δ_2 está en el intervalo de 0,05 a 0,3, preferentemente de 0,1 a 0,2, con mayor preferencia es aproximadamente 0,15.

En una modalidad, si $|II - AI|$ es inferior o igual a δ_2 , en donde δ_2 es como se describió anteriormente en la presente descripción, se distinguen 6 situaciones:

- si $II < Ct \leq AI$, AI se considera el índice más confiable; en donde Ct representa el umbral crítico, y corresponde al umbral de cirrosis para el índice, es decir el valor numérico de dicho índice que conduce a la clasificación del sujeto en el estadio de cirrosis o F4 de Metavir,
- si $AI \leq Ct < II$, AI se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción,
- si $Ct \leq II \leq AI$, AI se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción,
- si $Ct \leq AI \leq II$, AI se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción,

- si $AI \leq II \leq Ct$, II se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción, y
- si $II \leq AI \leq Ct$, II se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción.

5

En una modalidad de la invención, la prueba es una FibroMeter^{1G}, y el Ct es de aproximadamente 0,88.

10

En una modalidad, se emiten al menos dos Alertas de Evento y la selección del índice más adecuado se lleva a cabo mediante la comparación del valor del índice inicial (II) con el(los) valor(es) del Índice Mixto (MI) y/o del Índice Estimado (EI).

15

En una modalidad, si $|II - MI|$ es superior a δ_2 , el MI se considera el índice más confiable, en donde δ_2 es la diferencia clínicamente significativa. En una modalidad, está en el intervalo de 0,05 a 0,3, preferentemente de 0,1 a 0,2, con mayor preferencia es aproximadamente 0,15.

20

En una modalidad, si $|II - MI|$ es inferior o igual a δ_2 , en donde δ_2 es como se describió anteriormente en la presente descripción, se distinguen 6 situaciones:

25

- si $II < Ct \leq MI$, MI se considera el índice más confiable; en donde Ct representa el umbral crítico, y corresponde al umbral de cirrosis para el índice, es decir el valor numérico de dicho índice que conduce a la clasificación del sujeto en el estadio de cirrosis o F4 de Metavir,

30

- si $MI \leq Ct < II$, MI se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción,

- si $Ct \leq II \leq MI$, MI se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción,

- si $Ct \leq MI \leq II$, MI se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción,

- si $MI \leq II \leq Ct$, II se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción, y

- si $II \leq MI \leq Ct$, II se considera el índice más confiable; en donde Ct es como se describió anteriormente en la presente descripción.

En una modalidad de la invención, la prueba es una FibroMeter^{1G}, y el Ct es de aproximadamente 0,88.

35

En una modalidad de la invención, el índice seleccionado se llama el Índice Final. Por lo tanto, dicho índice puede usarse finalmente para evaluar la presencia y/o la gravedad de una enfermedad hepática en el sujeto. En una modalidad, dicha evaluación incluye la clasificación del sujeto en una clase de una clasificación de estadio de fibrosis, es decir una clase de una clasificación en donde cada clase se asocia con un estadio de fibrosis, tal como, por ejemplo, uno o más estadios F de Metavir.

40

En una modalidad, el método de la invención comprende emitir un comentario. En una modalidad, dicho comentario comprende indicar el número de Alertas de Evento, y el dato que ha emitido una Alerta de Evento o la Alerta de Evento principal. En una modalidad, dicho comentario comprende indicar cuál de los índices se ha considerado el más confiable. En una modalidad, dicho comentario comprende un mensaje de advertencia, en donde mientras más alertas existan, con más cuidado se deberá considerar el resultado.

45

[Segunda modalidad - ES3G]

50

La presente invención se refiere a un método que comprende:

a'1 - recopilar al menos uno, preferentemente al menos dos índices o puntuaciones o datos físicos resultantes de las pruebas no invasivas implementadas en un sujeto,

b'1 - analizar la confiabilidad de cada índice o puntuación,

c'1 -generar una alerta de evento si una prueba es no confiable,

55

d'1 - activar una respuesta a dicha alerta de evento.

Datos

60

En esta modalidad, en la etapa a'1, se miden preferentemente al menos 2, 3, 4, 5 o más índices o puntuaciones o datos físicos. En una modalidad de la invención, dichos índices o puntuaciones o datos físicos son preferentemente índices, con mayor preferencia puntuaciones y/o datos físicos.

En una modalidad, el método de la invención comprende al menos una puntuación resultante de una prueba de la familia FibroMeter y al menos un dato físico resultante de la prueba FibroscanTM. De acuerdo con una modalidad específica, los

resultados recopilados de FibroMeter™ (puntuación), y Fibroscan™ (datos físicos) se obtuvieron del mismo sujeto dentro de seis meses, preferentemente dentro de dos meses.

5 Análisis de la confiabilidad

En una modalidad, analizar la confiabilidad de una prueba significa considerar un dato (puntuación o dato físico, en nuestra modalidad) como el Dato Explicado mientras que los otros son Datos Explicativos.

10 En esta modalidad, la confiabilidad de la prueba se analiza en vista de los datos intrínsecos y/o extrínsecos del índice.

Predictores de confiabilidad

15 En una modalidad, el término puntuación significa la puntuación proporcionada por una función logit incluida en la regresión logística binaria donde la función logit: $1 / 1 - e^{-R}$ con $R = c + ax + \text{por...}$ Estas puntuaciones proporcionan predictores de confiabilidad independientes. Por consiguiente, estos determinan las clases de confiabilidad con descriptores de confiabilidad cualitativos de resultado de la prueba diagnóstica no confiable a muy confiable. La mayoría de los resultados se basan en el siguiente objetivo de diagnóstico: diagnóstico correcto por clasificación detallada de la fibrosis a menos que se indique de otra manera (tal como, por ejemplo, objetivo binario y/o presencia o ausencia).

20 Predictores de confiabilidad sin segmentación (para la población completa)

Los ejemplos de predictores de confiabilidad son: el Índice de Dispersión de la puntuación, AST, ALT, una puntuación que combina Índice de Dispersión y AST.

25 Predictores de confiabilidad con segmentación (para un subconjunto de pacientes)

30 Para subconjuntos específicos de pacientes, los ejemplos de predictores de confiabilidad de una prueba (tal como por ejemplo FIBROSCAN™, FIBROMETER™, CIRRHOMETER™, COMBIMETER™) son: la mediana de FIBROSCAN™; clasificaciones de FIBROSCAN™, IQR/M, IQR, clasificación de FIBROMETER™, clasificación de CIRRHOMETER™, urea, ALT, AST, puntuación de CirrhoMeter^{2G}, puntuación de FibroMeter^{2G}; en un predictor de confiabilidad puede definirse un umbral para valores predictivos negativos; y otro umbral para valores predictivos positivos

En una modalidad, el análisis de confiabilidad comprende dos etapas:

- 35 • la primera etapa corresponde al cálculo de probabilidades clasificadas que determinan las clases de confiabilidad. Por ejemplo, cuando la mayoría de los resultados de los sujetos en un subgrupo de sujetos se clasifican erróneamente, el resultado se considera no confiable. En una modalidad, cuando la probabilidad o la frecuencia de resultados exactos es $\geq 90\%$, se dice que los resultados son muy confiables.
- 40 • La segunda etapa incluye la determinación de factores predictivos de las clases de confiabilidad. Esto es obligatorio para identificar las clases de confiabilidad. Esta etapa es la diferencia principal entre exactitud y confiabilidad. Por ejemplo, una prueba diagnóstica que tiene una exactitud de diagnóstico de 90% significa que se obtiene una clasificación correcta en 90% de los pacientes mediante la prueba. La prueba es inexacta o completamente (100%) no confiable en 10% de los pacientes de la población completa. La evaluación de la confiabilidad puede determinar que en un subgrupo de sujetos (por ejemplo 2% de los pacientes), la tasa de exactitud puede ser 30% (es decir mayor que la tasa de inexactitud en la población completa), que es un resultado no confiable. Este resultado no confiable se diagnostica mediante uno o algunos predictores de confiabilidad significativos tales como peso corporal alto y/o un marcador discordante en comparación con otros marcadores (por ejemplo, el primero es altamente anómalo mientras que todos los otros marcadores son normales o ligeramente anómalos).

50 En una modalidad, una prueba se considera suficientemente confiable para el diagnóstico cuando su exactitud es mayor que 50% , preferentemente mayor que $60, 70, 75, 80, 85, 90$ o 95% . En consecuencia, la probabilidad de error del diagnóstico que se considera mayor que la tolerada para el diagnóstico es de $50\%, 40\%, 30\%, 20\%, 15\%, 10\%$ o 5% . En otras palabras, una prueba se considera suficientemente confiable cuando el porcentaje de sujetos con diagnóstico erróneo es menor que 50% , preferentemente menor que $40, 30, 20, 15, 10$ o 5% .

55 De acuerdo con una modalidad, el método de la invención comprende:

- 60 a. recopilar un índice de diagnóstico, llamado Índice Inicial, preferentemente una puntuación, con mayor preferencia una puntuación seleccionada de pruebas de sangre que comprenden ELF, FibroSpect™, APRI, FIB-4, Hepascore, Fibrotest™, o una puntuación de la familia FibroMeter, tal como por ejemplo FibroMeter™ y CirrhoMeter™ o CombiMeter™; o una prueba derivada de la familia FibroMeter, donde la urea se ha eliminado de los marcadores; u otra prueba diagnóstica; dicho índice implica la combinación matemática de al menos dos datos, preferentemente de al menos dos biomarcadores, marcadores clínicos y/o datos resultantes del método físico; dicho índice se refiere como Índice Inicial,
- 65 b. analizar la confiabilidad de cada dato mediante la ubicación del Índice en una clase de confiabilidad de la clasificación de confiabilidad en base a los datos obtenidos en una población de referencia;

De acuerdo con esta modalidad, la confiabilidad del Dato Explicado se determina mediante la ubicación de dicho dato en una clase de confiabilidad de una clasificación de confiabilidad. En una modalidad, dicha clasificación de confiabilidad proporciona, para cada conjunto de datos (es decir para cada combinación de dato explicado y dato explicativo), la confiabilidad del dato explicado. En una modalidad, dicha clasificación de confiabilidad se estableció en base a los datos obtenidos en una población de referencia. En una modalidad, la confiabilidad de un Índice Inicial se determina mediante la ubicación de dicho índice en una clase de confiabilidad de una clasificación de confiabilidad que incluye una tabla de dos entradas de Dato Explicado y Dato Explicativo (predictores), establecida en base a una población de referencia, en donde la lectura de la posición del Índice en la tabla proporciona la confiabilidad del Índice con consideración al Dato Explicativo. En una modalidad, las diferentes pruebas que se realizan son preferentemente predictores independientes de pacientes bien clasificados. En una modalidad, los predictores independientes de pacientes bien clasificados mediante el dato explicado se han determinado por análisis multivariante, tal como, por ejemplo, por regresión logística binaria.

En una modalidad, la confiabilidad de una prueba se determina en base a la comparación del resultado del paciente (dato explicado) con las clases de confiabilidad, en donde las clases de confiabilidad son parte de una clasificación basada en el nivel de confiabilidad de los resultados de la prueba (dato explicado) determinado en una base de datos de referencia.

En una modalidad, en esta base de datos de referencia, los predictores independientes de pacientes bien clasificados se han determinado por análisis multivariante (por ejemplo regresión logística binaria o RLB) en donde la variable dependiente es los pacientes bien clasificados mediante la prueba diagnóstica y las variables independientes son todas las variables compuestas incluidas en la prueba (datos intrínsecos) y las variables disponibles proporcionadas por datos demográficos y variables compuestas incluidas en otras pruebas disponibles (datos extrínsecos). El RLB proporciona una puntuación que incluye varias variables o una clasificación segmentada. Ambos dan lugar a una clasificación de confiabilidad:

- La puntuación se representa gráficamente contra la tasa de pacientes bien clasificados y las clases se determinan mediante el umbral elegido por el experto en este gráfico (ver los Ejemplos en la Figura 3).
- La clasificación segmentada se determina mediante la segmentación de las variables de acuerdo con la relación entre las variables independientes. El umbral puede determinarse mediante un gráfico que representa ambas variables o mediante índices clásicos como el índice de Youden. Esta clasificación se obtiene preferentemente cuando existe una interacción significativa entre dos variables independientes.
- Puede producirse un modelo mixto que incluye las clases de confiabilidad determinadas mediante RLB y las clases de confiabilidad proporcionadas por variable(s) individual(es) (tal(es) como, por ejemplo, variables inconsistentes como se definió anteriormente en la presente descripción).

Finalmente, el resultado del paciente (dato explicado) se compara con estas clases de confiabilidad y el nivel de confiabilidad subsiguiente se fija para este paciente con esta prueba.

En una modalidad, el dato explicado es un resultado de una prueba diagnóstica o un índice, preferentemente una puntuación, y los predictores probados para definir si son predictores independientes son:

- los datos combinados matemáticamente para calcular el índice (dato intrínseco), y
- el dato explicativo como se describió anteriormente en la presente descripción (dato extrínseco), y/o
- el dato extrínseco disponible con la prueba diagnóstica.

Un ejemplo de una variable intrínseca de una prueba es el índice de dispersión (DI), como se describió anteriormente en la presente descripción.

Los ejemplos de predictores de confiabilidad incluyen, pero no se limitan a las variables incluidas en o derivadas de las pruebas no invasivas, como se indica más abajo, preferentemente FibroMeter, InflaMeter y Fibroscan, y con mayor preferencia urea, ALT, AST, Índice de Dispersión, puntuación de CirrhoMeter^{2G}, puntuación de FibroMeter^{2G}, clases de FibroscanTM, clases de FibroMeterTM, clases de CirrhoMeterTM, mediana de FibroscanTM, IQR, IQR/M, plaquetas, A2M y t=relación urea/ Fibroscan.

Los ejemplos de predictores de confiabilidad para FibroMeter²⁰ y para CombiMeterTM se muestran en el Ejemplo 6.

Alerta de Evento

En una modalidad, una Alerta de Evento se emite cuando el dato explicado es no confiable.

Respuesta activada

En una modalidad, si se emite una Alerta de Evento del dato explicado, los datos explicativos se consideran, cada uno a su vez, como el dato explicado, y su confiabilidad se comprueba tal y como se describió anteriormente en la presente descripción.

En una modalidad ilustrativa, una prueba FibroMeter™ y una prueba Fibroscan™ se llevan a cabo considerando el resultado de FibroMeter™ como el dato explicado en una primera etapa. Si se emite una Alerta de Evento de FibroMeter™, el resultado de Fibroscan™ se considerará en una segunda etapa como el dato explicado, con el resultado de FibroMeter™ como dato explicativo, y se comprobará la confiabilidad de Fibroscan™.

5 En una modalidad de la invención, si solamente una prueba es confiable, su resultado se considera el Dato Final. En otra modalidad de la invención, si varias pruebas son confiables, se considera que la más confiable proporciona el Dato Final. En otra modalidad de la invención, si varias pruebas son confiables, se considera que su combinación proporciona el Dato Final. Por ejemplo, en una modalidad, si Fibroscan™ y FibroMeter™ o CirrhoMeter™ son confiables, se considera que la prueba CombiMeter™ correspondiente proporciona el Dato Final.

10 En una modalidad, por lo tanto, el Dato Final puede usarse finalmente para evaluar la presencia y/o la gravedad de una enfermedad hepática en el sujeto. En una modalidad, dicha evaluación incluye la clasificación del sujeto en una clase de una clasificación de estadio de fibrosis, es decir una clase de una clasificación en donde cada clase se asocia con un estadio de fibrosis, tal como, por ejemplo, uno o más estadios F de Metavir.

15 En una modalidad, el método de la invención comprende emitir un comentario. En una modalidad, dicho comentario comprende indicar cuál de los datos se ha considerado el más confiable. En una modalidad, dicho comentario comprende un mensaje de advertencia, en donde mientras existan más datos que se consideren no confiables, con más cuidado se deberá considerar el resultado.

20 Un ejemplo de un método de la invención que comprende llevar a cabo una prueba FibroMeter™ y una prueba Fibroscan™ se muestra en el Ejemplo 5, y el establecimiento de una clasificación de confiabilidad se muestra en el Ejemplo 7.

25 En una modalidad, el método de la invención proporciona una mejora de la precisión del diagnóstico a través de valores predictivos. En una modalidad, se fijan umbrales de 100 % de valores predictivos para el objetivo de diagnóstico principal con varias pruebas diagnósticas obtenidas anteriormente en una base de datos de referencia. En una modalidad, si el resultado del paciente es igual o superior al umbral correspondiente al 100 % del valor predictivo positivo, el objetivo de diagnóstico se ratifica. En otra modalidad, si el resultado del paciente es igual o superior al umbral correspondiente al 100 % del valor predictivo negativo, el objetivo de diagnóstico se excluye.

30 La presente invención también se refiere a un sistema experto que implementa el método de la invención.

35 La presente invención también se refiere a un software para implementar el método de la invención.

En una modalidad, el sistema experto es un software, en donde:

- a. el software recibe datos de entrada (niveles de biomarcadores, datos físicos, datos clínicos, puntuaciones, etc.),
- 40 b. el software se implementa para un sujeto individual, en base a los datos de entrada (niveles de biomarcadores, datos físicos, datos clínicos, puntuaciones etc...), varias pruebas que usan varias estrategias,
- b. el software comprueba la confianza de los datos de entrada de cada prueba,
- c. el software analiza la confiabilidad de cada prueba realizada en la etapa a,
- d. el software informa una prueba final correspondiente a la prueba más confiable para este individuo, y opcionalmente un comentario sobre la confiabilidad de la prueba final.

45 En otra modalidad, el sistema experto es un software, en donde:

- a. el software recibe la prueba de entrada (datos físicos, puntuaciones etc.), llamada Dato Explicado,
- 50 b. el software analiza la confiabilidad del Dato Explicado con el uso del Dato Explicativo,
- d. el software informa una prueba final correspondiente a la prueba más confiable para este individuo, opcionalmente junto con un comentario sobre la confiabilidad de la prueba final.

Breve descripción de las figuras

55 La Figura 1 es un gráfico que muestra que el índice de dispersión de la prueba FibroMeter™ inicial se correlaciona bien (línea discontinua oblicua) con la diferencia entre FibroMeter™ inicial (sin ES -sistema experto-) y final (con ES) modificada de acuerdo con el método de la invención. Las líneas horizontales indican las medias. Población de 825 pacientes con hepatitis C crónica.

60 La Figura 2 es un gráfico que muestra que el índice de dispersión es significativamente diferente entre los valores de las pruebas FibroMeter no modificada (No) o modificada (Si) (en 9,2 % de los casos en la caja a la derecha) mediante el método de la invención. La caja representa: mediana, intercuartiles y extremos. Población de 825 pacientes con hepatitis C crónica.

La Figura 3 es una combinación de gráficos que muestra el porcentaje de pacientes clasificados correctamente con el uso de FibroMeter^{2G} (FM2G) de acuerdo con la puntuación de confiabilidad para FibroMeter^{2G} en las clases de Fibroscan™ < 3 (A) o > 2 (B).

5 La Figura 4 es una combinación de gráficos que muestra la relevancia de las clases de confiabilidad. (4A) muestra la relación entre la puntuación de FibroMeter (líneas continuas) y la exactitud (líneas discontinuas) en el eje Y como una función de 10 clases de confiabilidad para FibroMeter en el eje X. Las clases de confiabilidad en el eje X se clasifican de acuerdo con el aumento de la puntuación de FibroMeter. (4B) muestra la exactitud de la puntuación de FibroMeter como una función de 10 clases de confiabilidad para FibroMeter. En la figura, las 5 primeras clases son las de clases de FS < 3 ($p < 0,001$) y las 5 últimas son las de clases de FS ≥ 3 ($p < 0,001$).

10 La Figura 5 es un gráfico que muestra la tasa de pacientes bien clasificados mediante la clasificación de la evaluación de la rigidez hepática (LSE) derivada de los puntos de corte de Castera y otros, como una función de las 3 clases de la clasificación e IQR/M, que muestra que IQR/M es un buen predictor de confiabilidad.

15 La Figura 6 es una combinación de gráficos que muestra la tasa de pacientes bien clasificados mediante LSE para el diagnóstico de (A) fibrosis significativa (punto de corte del diagnóstico: 7,1 kPa), o (B) cirrosis (punto de corte del diagnóstico: 12,5 kPa) como una función de la mediana de LSE e IQR/M.

20 EJEMPLOS

La presente invención se ilustra además por los siguientes ejemplos.

25 Ejemplo 1: Composición de la familia FM

30

35

40

45

50

55

60

65

Causa	Variables										N ^a	Ref					
	Edad	Sexo	peso	A2M	HA	PI	PLT	AST	Urea	GGT			Bili	ALT	Fer	Glu	FS
Virus																	
FM V 1G	x			x	x	x	x	x	x								7 1
FM V 2G	x	x		x	x	x	x	x	x								8 2
CM V 2G	x	x		x	x	x	x	x	x								8 3
FM V 3G ^b	x	x		x	x		x	x	x								8 4
CM V 3G ^b	x	x		x	x		x	x	x								8 4
Alcohol																	
FM A 1G	x			x	x	x											4 2
FM A 2G				x	x	x											3 -
NAFLD(esteatosis)																	
FM S	x		x											x	x	x	7 5

FM: FibroMeter, CM: CirrhoMeter,
A2M: alfa-2 macroglobulina, HA: ácido hialurónico, PI: índice de protrombina, PLT: plaquetas, Bili: bilirrubina, Fer: ferritina, Glu: glucosa, FS: Fibroscan
^a Número de variables
^b HA puede reemplazarse con GGT

En una modalidad, el Dato Explicado es la prueba derivada de la familia FibroMeter, donde la urea se ha eliminado de los marcadores

Referencias

- 5 1. Calès P, Oberti F, Michalak S, Hubert-Fouchard I, Rousselet MC, Konate A, Gallois Y, y otros. A novel panel of blood markers to assess the degree of liver fibrosis. *Hepatology* 2005;42:1373-1381.
2. Leroy V, Hilleret MN, Sturm N, Trocme C, Renversez JC, Faure P, Morel F, y otros. Prospective comparison of six non-invasive scores for the diagnosis of liver fibrosis in chronic hepatitis C. *J Hepatol* 2007; 46: 775-782.
- 10 3. Boursier J, Bacq Y, Halfon P, Leroy V, de Ledinghen V, de Muret A, Bourlière M, y otros. Improved diagnostic accuracy of blood tests for severe fibrosis and cirrhosis in chronic hepatitis C. *Eur J Gastroenterol Hepatol* 2009;21:28-38.
4. Calès P, Boursier J, Bertrais S, Oberti F, Gallois Y, Hubert IF, Dib N, Zarski JP, Rousselet MC; grupos multicéntricos (SNIFF 14 & 17, ANRS HC EP 23 Fibrostar). Optimization and robustness of blood tests for liver fibrosis and cirrhosis. *Clin Biochem.* 2010;43:1315-1322.
- 15 5. Calès P, Lainé F, Boursier J, Deugnier Y, Moal V, Oberti F, Hunault G, y otros. Comparison of blood tests for liver fibrosis specific or not to NAFLD. *J Hepatol* 2009; 50:165-173.

Ejemplo 2: Índice de Dispersión de FibroMeter modificado mediante el método de la invención

20 La Figura 1 y la Figura 2 muestran que el método de la invención conduce a una mejora de la determinación de la confiabilidad de FibroMeter™, como se indica mediante el análisis del índice de dispersión de FibroMeter™ antes (es decir Puntuación inicial) y después (es decir Puntuación final) de llevar a cabo el método de la invención. Los pacientes por debajo de los umbrales de dispersión (eje Y) tienen una disminución de la confiabilidad. Note que en esta población de 825 pacientes la mayoría de los símbolos de los pacientes se superponen en la esquina superior derecha.

25 La Figura 1 muestra que cuando el índice de dispersión (DI) es alto, el Sistema Experto de la invención no proporciona un Índice Alternativo o Estimado, y el Índice Final es el Índice Inicial. En otras palabras, el DI es un buen predictor de la confiabilidad de un método no invasivo para evaluar la presencia y/o la gravedad de una enfermedad hepática. La Figura 1 muestra la eficacia del método de la invención.

30 Ejemplo 3: Datos anómalos, intervalos de consistencia e inconsistencia para FibroMeter™ (Tabla 1)

35

40

45

50

55

60

65

Tabla 1

		PLT (G/L)	AST (IU/L)	Urea (mmol/L)	HA (µg/L)	PI (%)	A2M (mg/dl)	Edad (años)
Dato anómalo		5-20	500-20000	0-1	2000-3000	5-20	500-1000	99-120
Intervalo de inconsistencia falso positivo	Rojo	20-110	110-500	1-1.2	73-2000	20-79	402-500	82-99
	Amarilla	110-170	68-110	1.2-2.9	51-73	79-88	311-402	66-82
Intervalo de consistencia		170-308	6-68	2.9-6.7	24-51	88-109	105-311	29-66
Intervalo de inconsistencia falso positivo	Verde	308-390	N/A	6.7-8.8	11-24	109-121	100-105	8-29
	Azul	390-800	N/A	8.8-15	N/A	121-150	N/A	6-8
Dato anómalo		800-2000	0-6	15-70	0-11	150	10-100	0-6

A2M: alfa-2 macroglobulina, HA: ácido hialurónico, PI: índice de protrombina, PLT: plaquetas, AST: aspartato aminotransferasa

Ejemplo 4 Ejemplos de reglas de análisis preliminar de Alertas de Evento emitidas en donde se lleva a cabo una prueba FibroMeter™

5 Aquí se presentan algunas reglas de análisis preliminar de las Alertas de Evento de Dato Inconsistente que se emiten cuando se lleva a cabo una prueba FibroMeter™. En este ejemplo, se emiten 4 tipos de Alertas de Dato Inconsistente: Roja, Amarilla, Verde y Azul, de acuerdo con la Tabla del Ejemplo 1.

Supresión de Alertas de Evento

- 10
- si AST es la única Alerta Roja y si no hay Alerta Amarilla, la Alerta Roja se suprime.
 - si hay una Alerta Roja y una Alerta Amarilla, ambas pueden suprimirse, excepto cuando una de estas Alertas de Evento se emite por el dato EDAD. En esta situación, solamente se suprime la Alerta de Evento que se emite por EDAD.
 - si hay más de una Alerta Roja o dos Amarillas, estas alertas se suprimen.
 - si se emite una Alerta de Evento por el dato EDAD, dicha Alerta de Evento se suprime.

15 Priorización de Alertas de Evento

20 Si varias Alertas de Dato Inconsistente subsisten después de la supresión de las Alertas de Evento como se describió anteriormente en la presente descripción, la priorización de dichas Alertas de Evento (es decir la selección de la alerta principal) se lleva a cabo de la siguiente manera (Tabla 2 - no se representan todas las situaciones):

Tabla 2

25

Número de Alertas de Evento	Tipo y número de Alertas de Evento				Alerta Principal
	Rojo	Amarilla	Azul	verde	
30 2	1		1		Rojo
			1	1	Azul
	1			1	Rojo
	1	1			Rojo
35 3	1	1	1		Rojo
	1		1	1	Rojo
	1	1		1	Rojo
40 4	1	1	1	1	Rojo

Si se emiten varias Alertas de Evento del mismo tipo (tales como, por ejemplo, dos Alertas Rojas o dos Alertas Verdes), la principal será la emitida por el dato más alejado de su valor medio.

45 Ejemplo 5: Ejemplos de clasificaciones de confiabilidad

50 En las tablas más abajo, la "confiabilidad" corresponde al porcentaje de pacientes clasificados correctamente en una clase determinada. El "% de pacientes" corresponde al porcentaje de pacientes de la población de referencia clasificados en esta clase determinada. Cualquier referencia a una clase "F" se hace con referencia a los estadios F de Metavir estimados. La población de referencia comprende aproximadamente 600 pacientes.

Confiabilidad de Fibroscan (Tabla 3)

55 El dato explicado es la clasificación basada en Fibroscan, mientras que los datos explicativos son la puntuación de CirrhoMeter^{2G} (CM^{2G}), la puntuación de FibroMeter^{2G} (FM^{2G}), AST, y ALT.

Tabla 3

60

65

Dato Explicado (clasificación de Fibroscan)	Dato Explicativo				Confiabilidad	% de pacientes
	CM ^{2G}	FM ^{2G}	AST	ALT		
≤ F2	<0,13		<26		81	5.9
			≥26		92.1	60.1
	>0,13			<70	87.5	3.6
				≥70	68.6	5.2
≥ F3		<0,40			20	2.2
		0,40≤FM ^{2G} <0,71			69.2	3.9
		≥0,71			95.3	19.1

En conclusión, como se muestra en la Tabla 3, más de 92 % de los pacientes clasificados como ≤ F2 después de llevar a cabo una prueba Fibroscan, y con una puntuación de CirrhoMeter^{2G} inferior o igual a 0,13 y cuantificación de ASAT superior o igual a 26 IU/L se encuentran bien clasificados. Esta situación involucra a aproximadamente 60 % de los pacientes.

Por el contrario, aproximadamente 80 % de los pacientes clasificados en ≥ F3 después de una prueba Fibroscan y con una puntuación de FibroMeter^{2G} inferior a 0,40 tienen un diagnóstico erróneo. En consecuencia, Fibroscan no es una prueba confiable para estos pacientes, que representan aproximadamente 2 % de los pacientes.

Confiabilidad de FibroMeter²⁰ (Tabla 4)

El dato explicado es la clasificación basada en FibroMeter^{2G}(clases de FM^{2G}), mientras que los datos explicativos son la clasificación de Fibroscan (clases de FS), la puntuación de FibroMeter^{2G} (FM^{2G}), la relación urea/puntuación resultante de una prueba Fibroscan (urea/FS), IQR (Intervalo entre Cuartiles de Fibroscan) y plaquetas.

Tabla 4

Dato Explicado (clases de FM ^{2G})	Dato Explicativo					Confiabilidad	% de pacientes
	clases de FS	IQR	Plaquetas	FM ^{2G}	urea/FS		
≤ F2	<2	≤1,0				91,7	23.4
		> 1,0				84,8	9.9
						76,9	3.9
	2		>285			95	18.0
	>2				≤0,43	23,5	2.5
					72	3.7	
≥ F3					<1,015	92,1	34.3
					≥1,015	78,6	4.2

En conclusión, como se muestra en la Tabla 4, más de 91 % de los pacientes clasificados como ≤ F2 después de llevar a cabo una prueba FibroMeter^{2G}, y clasificados como F<2 después de una prueba Fibroscan con un IQR inferior o igual a 1,0 se encuentran bien clasificados. Por lo tanto, FibroMeter^{2G} es una prueba confiable para estos pacientes, que representan aproximadamente 23 % de la población. La misma conclusión puede hacerse para pacientes clasificados como ≥ F3 después de llevar a cabo una prueba FibroMeter^{2G}, y con una relación urea/FS inferior a 1,015.

Por el contrario, aproximadamente 78 % de los pacientes clasificados en ≥ F3 después de una prueba FibroMeter^{2G} y con una puntuación de FibroMeter^{2G} inferior a 0,43 tienen un diagnóstico erróneo. En consecuencia, FibroMeter^{2G} no es una prueba confiable para estos pacientes, que representan aproximadamente 2,5 % de la población.

Ejemplo 6: Ejemplos de predictores de confiabilidad

Predictores de confiabilidad para FibroMeter²⁰ (Tabla 5)

5 Los ejemplos de predictores de confiabilidad para FibroMeter²⁰ se muestran en la Tabla 5 más abajo. Población de 597 pacientes con hepatitis C crónica. FS: Fibroscan, FM: FibroMeter.

Tabla 5

Segmentación	Predictores de confiabilidad	Notas
Todos los pacientes	clases de FS, ALT, IQR/M	Interacción significativa entre clases de FS y ALT que justifica la próxima etapa:
1. clases de FS < 3	clases de FM, clases de FS, AST, IQR	Interacción significativa con AST que justifica la próxima etapa:
AST > 185		100 % de pacientes con clasificación errónea
AST < 185	clases de FM, clases de FS, IQR	Estos 3 datos se incluyen en una Puntuación #1
2. clases de FS ≥ 3		
	clases de FM, ALT, A2M	Estos 3 datos se incluyen en una Puntuación #2
ALT > 430		100 % de pacientes con clasificación errónea

El método de la invención para determinar la confiabilidad de FibroMeter²⁰ para un paciente incluye, en este ejemplo, dos etapas:

- 1/ aplicar un umbral de variables (para AST o ALT) que determina 100 % de pacientes con clasificación errónea; y
- 2/ después aplicar la puntuación multivariante en los otros pacientes.

La Figura 3 y la Figura 4 muestran la relevancia de las clases de confiabilidad.

Para los pacientes clasificados en clases de fibrosis de Fibroscan (FS) < 3, la eficacia de la puntuación predictiva que incluye 3 predictores (puntuación #1) se muestra en la Figura 3A donde la puntuación distingue 4 clases de confiabilidad (puntuaciones de confiabilidad) en el eje X (p < 0,001 por ANOVA).

Para los pacientes clasificados en clases de FS > 2, la eficacia de la puntuación predictiva que incluye 3 predictores (puntuación #2) se muestra en la Figura 3B donde la puntuación distingue 4 clases de confiabilidad (puntuaciones de confiabilidad) en el eje X (p < 0,001 por ANOVA).

En la Figura 4, las 5 primeras clases son las de clases de FS < 3 (p < 0,001) y las 5 últimas son las de clases de FS ≥ 3 (p < 0,001). La Figura 4A muestra la relación entre la puntuación de FibroMeter (líneas continuas) y la exactitud (líneas discontinuas) en el eje Y como una función de 10 clases de confiabilidad para FibroMeter en el eje X. Las clases de confiabilidad en el eje X se clasifican de acuerdo con el aumento de la puntuación de FibroMeter. La Figura 4B muestra la exactitud de la puntuación de FibroMeter como una función de 10 clases de confiabilidad para FibroMeter.

Predictores de confiabilidad para CombiMeter™ (Tabla 6)

Los ejemplos de predictores de confiabilidad para CombiMeter™ se muestran en la Tabla 6 más abajo. Estos incluyen clases de Fibroscan, mediana de Fibroscan, Urea, puntuación de FibroMeter^{2G} (FM2G), clases de CirrhoMeter^{2G} (clases de CM2G) y AST.

Tabla 6

Predictores de confiabilidad		Clases de confiabilidad	Exactitud
1 ^{er} lugar	2 ^{do} lugar		
Clases de Fibroscan			
≤ 2	Mediana de Fibroscan		
	$\geq 10,5$	1a	Confiabilidad subóptima
	$< 10,5$	1b	Muy confiable
= 3	Urea		
	> 8	2a	No confiable
	≤ 8	2b	Confiabilidad subóptima
= 3.5	FM^{2G}		
	$< 0,3$	3a	Bastante confiable
	$\geq 0,3$	3b	Muy confiable
= 4	clases de CM2G ≤ 2	4a	No confiable
	AST > 400	4b	No confiable
	Otros	4c	Confiable

Ejemplo 7: Clasificación de exactitud de la LSE

Pacientes y métodos

Pacientes

Dos poblaciones con biopsia hepática y LSE se incluyeron en el presente estudio. LSE representa la evaluación de la rigidez hepática (LSE) y corresponde a todas las mediciones registradas durante un examen con el dispositivo de Fibroscan. La primera población estaba compuesta por pacientes con enfermedad hepática crónica reclutados en 3 centros de Francia entre 2004 y 2009 (Angers: n=383; Bordeaux: n=309; y Grenoble: n=142). Los pacientes incluidos en los centros de Angers y Bordeaux tenían varias causas de enfermedades hepáticas crónicas mientras que los de Grenoble tenían CHC. La segunda población fue la del estudio multicéntrico ANRS/HC/EP23 de Fibrostar promovido por la Agencia Nacional Francesa de Investigación de SIDA y Hepatitis (Zarski y otros, J Hepatol 2012; 56:55-62). Los pacientes incluidos en ambas poblaciones se identificaron y en última instancia se agruparon como una sola observación para el análisis estadístico. Todos los pacientes proporcionaron su consentimiento informado escrito. El protocolo del estudio cumplía con las directrices éticas de la Declaración de Helsinki vigente y recibió la aprobación de los comités de ética locales.

Evaluación histológica

La fibrosis hepática se evaluó de acuerdo con la estadificación de fibrosis de Metavir (F_M). La fibrosis significativa se definió como F_M de Metavir ≥ 2 , la fibrosis grave como F_M de Metavir ≥ 3 , y la cirrosis como F_M de Metavir 4. En la primera población, las evaluaciones histológicas se realizaron en cada centro a ciegas por patólogos titulares especializados en hepatología. En el estudio de Fibrostar, las lesiones histológicas se evaluaron centralmente por dos expertos titulares con una lectura consensuada en casos de discordancia. La estadificación de la fibrosis se consideró confiable cuando la longitud del espécimen hepático era ≥ 15 mm y/o el número de conductos portaes ≥ 8 .

Evaluación de la rigidez hepática

Condiciones del examen - la LSE mediante Fibroscan (Echosens, París, Francia) se realizó con la sonda M (en donde la sonda M es una sonda estándar mientras que existen sondas específicas para pacientes obesos o niños) y por un observador experimentado (>50 exámenes antes del estudio), a ciegas para los datos de los pacientes. Un intervalo de tiempo de ≤ 3 meses entre la biopsia hepática y la LSE se consideró aceptable para los propósitos del estudio. Las condiciones del examen fueron las recomendadas por el fabricante, con el objetivo de obtener al menos 10 mediciones válidas. Una medición de LSE corresponde a una medición individual de rigidez hepática, es decir, después de presionar el botón de la sonda de Fibroscan. La validez de la medición se determinó mediante un proceso interno definido por el fabricante del Fibroscan. La validez se evalúa para cada medición de LSE: la rigidez hepática se muestra en la pantalla cuando la medición se considera "válida", y no se muestra si se considera "no válida". Los resultados se expresaron como la mediana y el IQR (kilopascales) de todas las mediciones válidas: La mediana de LSE (kPa) corresponde a la mediana de todas las mediciones válidas realizadas durante la LSE; y el intervalo entre cuartiles (IQR) de la LSE (kPa) corresponde al intervalo alrededor de la mediana de LSE que contiene 50 % de las mediciones válidas entre el 25^{to} y 75^{to} percentil.

De acuerdo con la definición usual, la LSE se consideró confiable cuando incluía ≥ 10 mediciones válidas con una tasa de éxito $\geq 60\%$ e $IQR/M \leq 0,30$, en donde la tasa de éxito (%) se calcula como la relación del número de mediciones válidas respecto al número total de mediciones realizadas durante la LSE.

5

Interpretación del resultado de la LSE - la mediana de LSE se interpretó de acuerdo con los puntos de corte del diagnóstico publicados en estudios anteriores. Dado que la hepatitis C crónica (CHC) era la causa principal de enfermedad hepática en la población del estudio (68 %), se probaron los puntos de corte publicados por Castera y otros (Castera y otros, Gastroenterology, 2005;128:343-50): $\geq 7,1$ kPa para $F_{M\geq 2}$ y $\geq 12,5$ kPa para F_{M4} , los de Ziol y otros (Hepatology 2005;41:48-54): $\geq 8,8$ kPa para $F_{M\geq 2}$ y $\geq 14,6$ kPa para F_{M4} , y los calculados específicamente para CHC en el metaanálisis de Stebbing y otros (Journal of Clinical Gastroenterology 2010; 44:214-9): $\geq 8,5$ kPa para $F_{M\geq 2}$ y $\geq 16,2$ kPa para F_{M4} . Dado que existían varias causas de enfermedad hepática crónica en la población del estudio, también se probó el punto de corte publicado en el metaanálisis de Friedrich-Rust y otros (Gastroenterology 2008; 134:960-74): $\geq 7,7$ kPa para $F_{M\geq 2}$ y $\geq 13,1$ kPa para F_{M4} . Mediante el uso de los puntos de corte del diagnóstico, la mediana de LSE se categorizó en estadios F_{FS} estimados de acuerdo con el(los) estadio(s) F de Metavir más probable(s). Este enfoque proporcionó la siguiente clasificación de LSE: resultado de LSE <punto de corte para $F_{M\geq 2}$: $F_{FS} 0/1$; \geq punto de corte para $F_{M\geq 2}$ y <punto de corte para F_{M4} : $F_{FS} 2/3$; \geq punto de corte para F_{M4} : $F_{FS} 4$.

10

15

Análisis estadístico

20

Debido a que la distribución era sesgada para la mayoría de las variables cuantitativas, estas se expresaron como mediana con el 1^{er} y el 3^{er} cuartil entre paréntesis. La exactitud del diagnóstico se expresó principalmente como AUROC (para diagnósticos binarios de fibrosis significativa, fibrosis grave, o cirrosis) o la tasa de pacientes bien clasificados mediante la clasificación de LSE. Las AUROC se compararon de acuerdo con Delong y otros para grupos pareados (Biometrics 1988;44:837-45), y Hanley y otros para grupos no pareados (Radiology 1982;143:29-36).

25

30

Para identificar los factores que influyen sobre la exactitud de la LSE, se determinaron las variables asociadas independientemente con el siguiente objetivo de diagnóstico: fibrosis significativa, fibrosis grave, o cirrosis mediante regresión logística binaria directa gradual. De hecho, por definición, cada variable seleccionada mediante un análisis multivariante es un predictor independiente del objetivo de diagnóstico estudiado. En otras palabras, cuando se selecciona con la mediana de LSE, un predictor independiente influye sobre el resultado (objetivo de diagnóstico) para cada nivel fijado de rigidez hepática. Por consiguiente, el análisis multivariante permitió la identificación del predictor que influye sobre la exactitud de la LSE con respecto a la estadificación de la fibrosis. En el presente estudio, la variable dependiente -mediana de LSE- se probó con las siguientes variables independientes: edad, sexo, índice de masa corporal, causa de la enfermedad hepática crónica (CHC frente a otras), mediciones válidas de LSE ≥ 10 , tasa de éxito de la LSE, IQR/M (en donde IQR/M corresponde a la relación IQR de LSE/ mediana de LSE), y longitud de la biopsia como una variable de confusión putativa. Los análisis estadísticos se realizaron con el uso del software SPSS versión 18.0 (IBM, Armonk, NY, Estados Unidos) y SAS 9.1 (SAS Institute Inc., Cary, NC, Estados Unidos).

35

40

Resultados

Pacientes

45

Las características principales de los 1165 pacientes incluidos en el estudio se presentan en la Tabla 7. La causa de la enfermedad hepática crónica era CHC en 68,5 % de los pacientes, mono infección por hepatitis B: 5,7 %, alcohol: 12,4 %, enfermedad de hígado graso no alcohólico: 3,3 %, y otras: 10.1 El estado de sobrepeso (índice de masa corporal $\geq 25,0$ kg/m²) estaba presente en 44,0 % de los pacientes. Las biopsias hepáticas se consideraron confiables en 92,0 % de los casos. La prevalencia para fibrosis significativa, fibrosis grave, y cirrosis fue 63,3 %, 38,9 %, y 21,0 % respectivamente.

50

Tabla 7

55

60

65

		Causa de la enfermedad hepática			
		Todos	CHC	Otro	P ^a
5	Pacientes (n)	1165	798	367	-
	Edad (años)	<u>51,1 (43,9-60,5)</u>	<u>50,1 (43,9-59,7)</u>	<u>54,2 (43,9-63,3)</u>	0.084
	Varón (%)	65.2	62.9	70.0	0.018
10	Índice de masa corporal (kg/m ²)	<u>24,5 (22,2-27,6)</u>	<u>24,2 (22,1-26,7)</u>	<u>25,1 (22,5-29,4)</u>	<10 ⁻³
	Índice de masa corporal ≥25 kg/m ² (%)	44.0	40.1	50.9	10 ⁻³
	Estadio F _M de Metavir (%):				<10 ⁻³
15	- 0	5.6	3.5	10.3	
	- 1	31.0	37.1	17.6	
	- 2	24.5	27.4	17.9	
20	- 3	17.9	17.5	18.7	
	- 4	21.0	14.5	35.5	
	Longitud de la biopsia (mm)	<u>25 (18-30)</u>	<u>24 (18-30)</u>	<u>25 (17-32)</u>	0.093
25	Biopsia confiable (%)	92.0	93.8	88.0	10 ⁻³
	Mediana de LSE (kPa)	<u>8,1 (5,8-14,0)</u>	<u>7,8 (5,6-11,1)</u>	<u>11,0 (6,6-25,1)</u>	<10 ⁻³
	Mediciones válidas (n)	<u>10 (10-10)</u>	<u>10 (10-10)</u>	<u>10 (10-10)</u>	0.227
30	Mediciones válidas de LSE ≥10 (%)	92.8	93.3	91.6	0.291
	Tasa de éxito de la LSE (%)	<u>100 (83-100)</u>	<u>100 (83-100)</u>	<u>91 (77-100)</u>	10 ⁻³
	Tasa de éxito de la LSE ≥60 % (%)	89.8	91.9	85.1	<10 ⁻³
35	IQR/M	<u>0,17 (0,12-0,25)</u>	<u>0,17 (0,12-0,24)</u>	<u>0,18 (0,11-0,25)</u>	0.211
	IQR/M ≤0,30 (%)	85.5	86.1	84.3	0.416
	LSE confiable (%) ^b	75.7	77.6	71.6	0.027
40	CHC: monoinfección por hepatitis C crónica, IQR/M: intervalo entre cuartiles de LSE/mediana de LSE; ^a Entre CHC y otras causas de enfermedad hepática; ^b De acuerdo con la definición usual de confiabilidad de la LSE (mediciones válidas ≥10 y tasa de éxito ≥60 % e IQR/M ≤0,30)				

45 Exactitud de la LSE

Las AUROC (±SD) de LSE para el diagnóstico de fibrosis significativa, fibrosis grave, y cirrosis fueron respectivamente 0,822±0,012, 0,872±0,010, y 0,910±0,011 (Tabla 8).

50 Tabla 8

55

60

65

Causa de la enfermedad hepática	Objetivo de diagnóstico	Evaluación de la rigidez hepática				
		Todos	Confiable ^a	No confiable	p ^b	
5	Todos	F _M ≥2	0,822 ± 0,012	0,835 ± 0,014	0,794 ± 0,026	0.165
		F _M ≥3	0,872 ± 0,010	0,881 ± 0,012	0,856 ± 0,023	0.344
		F _M 4	0,910 ± 0,011	0,913 ± 0,012	0,906 ± 0,022	0.780
10	CHC	F _M ≥2	0,787 ± 0,016	0,805 ± 0,018	0,733 ± 0,037	0.080
		F _M ≥3	0,843 ± 0,015	0,856 ± 0,016	0,811 ± 0,035	0.242
		F _M 4	0,897 ± 0,016	0,900 ± 0,018	0,918 ± 0,038	0.669
15	Otro	F _M ≥2	0,883 ± 0,019 ^c	0,888 ± 0,024 ^d	0,889 ± 0,032 ^c	0.980
		F _M ≥3	0,905 ± 0,016 ^d	0,913 ± 0,018 ^e	0,888 ± 0,034	0.516
		F _M 4	0,908 ± 0,016	0,920 ± 0,018	0,862 ± 0,037	0.159

CHC: mono infección por hepatitis C crónica; ^a De acuerdo con la definición usual de confiabilidad de la LSE (LSE con mediciones válidas ≥10 y tasa de éxito ≥60 % e intervalo entre cuartiles de LSE/mediana de LSE ≤0,30); ^b Entre LSE confiable y no confiable; ^cp≤10⁻³ vs pacientes con CHC; ^dp≤0,010 vs pacientes con CHC; ^ep≤0,05 vs paciente con CHC

25

Las AUROC de LSE en biopsias no confiables no fueron significativamente diferentes a las de las biopsias confiables (no se muestran los detalles). Las tasas de pacientes bien clasificados de acuerdo con los diversos puntos de corte del diagnóstico probados se presentan en la **Tabla 9**.

30 Tabla 9

Objetivo de diagnóstico	Punto de corte del diagnóstico	Causa de la enfermedad hepática					
			Referencia	Punto de corte	Todos	CHC	Otro
35	F _M ≥2	Castera y otros	≥7,1	74,8 ^b	71,2 ^b	82.7	<10 ⁻³
		Ziol y otros	≥8,8	70.1	66.0	79.1	<10 ⁻³
		Stebbing y otros	≥8,5	70.9	66.4	80,7 ^c	<10 ⁻³
		Friedrich-Rust y otros	≥7,7	73,2 ^d	69,3 ^d	81.8	<10 ⁻³
40	F _M 4	Castera y otros	≥12,5	84,8 ^d	86,2 ^d	81.8	0.059
		Ziol y otros	≥14,6	86.1	87.9	82.1	0.008
		Stebbing y otros	≥16,2	86.8	88.8	82.4	0.003
		Friedrich-Rust y otros	≥13,1	85,3 ^e	86,8 ^e	82.1	0.038
45	LSE	Castera y otros ^f	-	62,0 ^c	59,7 ^c	67.0	0.018
	clasificación	Ziol y otros ^g	-	58.8	56.2	64.5	0.008
		Stebbing y otros ^h	-	59.8	57.1	65.6	0.006
50		Friedrich-Rust y otros ⁱ	-	61,2 ^c	58.6	67.0	0.007

CHC: mono infección por hepatitis C crónica; ^a Entre CHC y otras; ^b p<0,05 vs otras; ^c p<0,05 vs Ziol; ^d p≤0,05 vs Ziol o Stebbing; ^e p≤0,05 vs Stebbing; ^f La clasificación de LSE se deriva del punto de corte para fibrosis significativa y el punto de corte para cirrosis de la siguiente manera: <7,1 kPa: F_{FS}0/1, ≥7,1 kPa y <12,5 kPa: F_{FS}2/3, ≥12,5 kPa: F_{FS}4; ^g clasificación de LSE: <8,8 kPa: F_{FS}0/1, ≥8,8 kPa y <14,6 kPa: F_{FS}2/3, ≥14,6 kPa: F_{FS}4; ^h clasificación de LSE: <8,5 kPa: F_{FS}0/1, ≥8,5 kPa y <16,2 kPa: F_{FS}2/3, ≥16,2 kPa: F_{FS}4; ⁱ clasificación de LSE: <7,7 kPa: F_{FS}0/1, ≥7,7 kPa y <13,1 kPa: F_{FS}2/3, ≥13,1 kPa: F_{FS}4

Los puntos de corte publicados por Castera y otros proporcionaron la mayor exactitud para fibrosis significativa y clasificación de LSE, y por lo tanto se usaron para los análisis estadísticos posteriores.

65

Definición usual de confiabilidad de la LSE

92,8 % de la LSE incluía al menos 10 mediciones válidas, 89,8 % alcanzó una tasa de éxito ≥ 60 %, y 85,5 % tenía una IQR/M $\leq 0,30$ (Tabla 7). Ninguna de estas condiciones condujo a un aumento significativo del AUROC de la LSE.

75,7 % de la LSE cumplió estos 3 criterios; por consiguiente, se consideraron confiables de acuerdo con la definición usual de confiabilidad de la LSE. Una medición de LSE individual puede ser errónea debido a varias condiciones (posición o inclinación de la sonda, movimiento respiratorio, etc.). Los criterios de confiabilidad de la LSE corresponden por lo tanto a las condiciones requeridas antes de considerar la mediana de LSE como la rigidez hepática "real". En este escenario, la evaluación de la rigidez hepática se considera usualmente confiable cuando cumple todos los criterios siguientes: mediciones válidas ≥ 10 , tasa de éxito ≥ 60 %, y relación IQR/mediana (IQR/M) $\leq 0,30$.

Las AUROC para fibrosis significativa, fibrosis grave, o cirrosis no fueron significativamente diferentes entre LSE confiable y no confiable (Tabla 8). Mediante el uso de los puntos de corte de Castera y otros ($\geq 7,1$ kPa para $F_{M \geq 2}$ y $\geq 12,5$ kPa para F_{M4}), la exactitud de la LSE no fue significativamente diferente entre LSE confiable y no confiable para el diagnóstico de fibrosis significativa (respectivamente: 75,5 % vs 72,1 %, $p=0,255$) o cirrosis (85,8 % vs 81,5 %, $p=0,082$). De manera similar, la tasa de pacientes bien clasificados mediante la clasificación de LSE ($F_{FS0/1}$, $F_{FS2/3}$, F_{FS4}) derivada de los puntos de corte de Castera no fue significativamente diferente LSE entre confiable y no confiable (respectivamente: 63,5 % vs 57,2 %, $p=0,064$).

Predictores independientes de estadificación de la fibrosis

Los predictores independientes de fibrosis significativa, fibrosis grave, o cirrosis se detallan en la Tabla 10.

Tabla 10

Objetivo de diagnóstico	Etapas	Variable	p	Relación de probabilidades (CI de 95 %)
$F_{M \geq 2}$	1 ^{ra}	Mediana de LSE	$<10^{-3}$	1,323 (1,262-1,387)
	2 ^{da}	Edad	$<10^{-3}$	1,023 (1,011-1,035)
	3 ^{ra}	IQR/M	0.002	0,197 (0,072-0,543)
$F_{M \geq 3}$	1 ^{ra}	Mediana de LSE	$<10^{-3}$	1,278 (1,234-1,324)
	2 ^{da}	IQR/M	10^{-3}	0,121 (0,034-0,433)
	3 ^{ra}	Edad	0.007	1,017 (1,005-1,030)
F_{M4}	1 ^{ra}	Mediana de LSE	$<10^{-3}$	1,201 (1,168-1,234)
	2 ^{da}	Longitud de la biopsia	0.002	0,965 (0,944-0,987)
	3 ^{ra}	IQR/M	0.005	0,070 (0,011-0,442)

En resumen, además de la mediana de LSE, IQR/M fue la única característica de LSE que se asoció independientemente con los tres objetivos de diagnóstico de fibrosis, sin influencia significativa del número de mediciones válidas de LSE, la tasa de éxito de la LSE, o la causa de la enfermedad hepática. No hubo colinealidad entre la mediana de LSE e IQR/M (coeficiente de correlación de Spearman=0,047, $p=0,109$). Los predictores independientes fueron los mismos cuando las variables se introdujeron como resultados dicotómicos (IQR/M $\leq 0,30$, tasa de éxito de la LSE ≥ 60 %, biopsia confiable vs no confiable) en el análisis multivariante (los datos detallados no se muestran).

Clasificación de exactitud de la LSE

En esta descripción se desarrolló una clasificación con el uso de los predictores independientes de exactitud anteriores.

IQR/M - la exactitud de la LSE disminuyó cuando IQR/M aumentó y se identificaron tres subgrupos de LSE: IQR/M $\leq 0,10$ (16,6 % de los pacientes); $0,10 < \text{IQR/M} \leq 0,30$ (69,0 %); IQR/M $> 0,30$ (14,5 %). La LSE con IQR/M $\leq 0,10$ tuvo una exactitud significativamente mayor que la LSE con IQR/M $> 0,10$ (Tabla 11).

Tabla 11

		AUROC			Exactitud del diagnóstico (%) ^a			
Objetivo de diagnóstico:		F _M ≥2	F _M ≥3	F _M 4	F _M ≥2	F _M 4	Clasificación de LSE	
5	≤0,1	0,886 ±	0,937 ±	0,970 ±	77,1	90,4	69,1	
		0,024	0,018	0,011				
	10	0,1 < y ≤0,3	0,822 ±	0,868 ±	0,895 ±	75,6	84,7	62,6
			0,015	0,013	0,015			
	15	>0,3	0,785 ±	0,842 ±	0,898 ±	69,1	80,6	53,9
			0,035	0,032	0,031			
Comparación (p):								
Tendencia lineal ^b		-	-	-	0,091	0,009	0,003	
≤0,1 vs 0,1 < y ≤0,3		0,024	0,002	<10 ⁻³	0,661	0,043	0,092	
≤0,1 vs >0,3		0,017	0,010	0,029	0,088	0,008	0,003	
20		0,1 < y ≤0,3 vs >0,3	0,331	0,451	0,931	0,081	0,196	0,039
^a Tasa de pacientes bien clasificados con el uso de 7,1 kPa como el punto de corte de LSE para el diagnóstico de fibrosis significativa (F _M ≥2), 12,5 kPa para el diagnóstico de cirrosis (F _M 4), o clasificación de LSE (F _{FS} 0/1, F _{FS} 2/3, F _{FS} 4) derivada de los 2 puntos de corte del diagnóstico anteriores (Castera y otros); ^b p para la tendencia lineal en los 3 subgrupos de IQR/M								

La LSE con 0,10 < IQR/M ≤0,30 tuvo mayor exactitud que la LSE con IQR/M >0,30, pero la diferencia no alcanzó significación estadística.

Mediana de LSE - Mediante el uso de 7,1 kPa como un punto de corte del diagnóstico (Castera y otros; Friedrich-Rust y otros), la tasa de pacientes bien clasificados para fibrosis significativa fue muy buena en medianas de LSE ≥7,1 kPa, pero solo pasable en medianas de LSE <7,1 kPa: 81,5 % vs 64,5 %, respectivamente (p<10⁻³). Mediante el uso de 12,5 kPa como punto de corte del diagnóstico (Castera y otros; Friedrich-Rust y otros), la tasa de pacientes bien clasificados para cirrosis fue excelente en medianas de LSE <12,5 kPa, pero solo pasable en medianas de LSE ≥12,5 kPa: 94,3 % vs 60,4 %, respectivamente (p<10⁻³). Por lo tanto, la LSE demostró un valor predictivo negativo excelente para cirrosis y un valor predictivo positivo muy bueno para fibrosis significativa. Por el contrario, tuvo un valor predictivo positivo insuficiente para cirrosis y un valor predictivo negativo insuficiente para fibrosis significativa. Finalmente, la tasa de pacientes bien clasificados mediante la clasificación de LSE derivada de los puntos de corte de Castera y otros no fue significativamente diferente entre sus 3 clases, F_{FS}0/1: 64,5 %, F_{FS}2/3: 60,4 %, y F_{FS}4: 60,4 % (p=0,379).

IQR/M y mediana de LSE - En pacientes con mediana de LSE <7,1 kPa, la exactitud del diagnóstico de la clasificación de LSE derivada de los puntos de corte de Castera y otros no fue significativamente diferente entre los tres subgrupos de IQR/M (p=0,458; Figura 5). Por el contrario, en pacientes con mediana de LSE ≥7,1 kPa, la exactitud del diagnóstico de la clasificación de LSE fue significativamente menor en la LSE con IQR/M >0,30 en comparación con la LSE con IQR/M ≤0,30 (43,8 % vs 64,1 %, p<10⁻³; Figura 5). Las tasas de pacientes bien clasificados para los diagnósticos binarios de fibrosis significativa o cirrosis como una función de IQR/M y mediana de LSE se detallan en la Figura 6. En resumen, en pacientes con mediana de LSE ≥7,1 kPa, la LSE con IQR/M >0,30 tuvo menor exactitud para fibrosis significativa que la LSE con IQR/M ≤0,30 (67,6 % vs 84,3 %, p<10⁻³). En pacientes con mediana de LSE ≥12,5 kPa, la LSE con IQR/M >0,30 tuvo menor exactitud para cirrosis que la LSE con IQR/M ≤0,30 (45,1 % vs 64,0 %, p=0,011). Propuesta de nuevos criterios de confiabilidad en la LSE

Los hallazgos anteriores condujeron al desarrollo de nuevos criterios para la interpretación de los resultados de la LSE (Tabla 12).

Tabla 12

		Mediana de LSE			Tasa de pacientes (%)
		<7,1	7,1 ≤ - <12,5	≥12,5	
5	<i>Diagnóstico de LSE^a:</i>	<i>F_{FS}0/1</i>	<i>F_{FS}2/3</i>	<i>F_{FS}4</i>	
10	IQR/M ≤0,10	<i>LSE muy confiable</i>			16.6
	0,10< y ≤0,30	<i>LSE confiable</i>			74,3 ^b
	>0,30		<i>LSE poco confiable</i>		9.1

^a Diagnóstico de LSE después de la categorización de la mediana de LSE en estadios F de Metavir (F_{FS}) estimados de acuerdo con los puntos de corte del diagnóstico de Castera y otros: 7,1 kPa para fibrosis significativa y 12,5 kPa para cirrosis (Castera y otros); ^b Con inclusión del subgrupo con IQR/M >0,30 y mediana de LSE <7,1

La exactitud de la LSE en el subgrupo de LSE con IQR/M ≤0,10 fue mayor que en la población completa (Tabla 13).

Tabla 13

		AUROC			Exactitud del diagnóstico (%) ^a		
Objetivo de diagnóstico:		F _M ≥2	F _M ≥3	F _M 4	F _M ≥2	F _M 4	Clasificación de LSE
25	LSE: <i>Todos^b</i>	0.822± 0.012	0.872± 0.010	0.910± 0.011	74.9	85.0	62.4
	Muy confiable	0,886 ± 0,024	0,937 ± 0,018	0,970 ± 0,011	77.1	90.4	69.1
	Confiable	0,823 ± 0,014	0,876 ± 0,012	0,904 ± 0,014	75.3	85.8	63.2
	Poco confiable	0,773 ± 0,045	0,745 ± 0,049	0,819 ± 0,052	67.6	69.5	43.8
<i>Comparación (p):</i>							
Tendencia lineal ^c		-	-	-	0.107	<10 ⁻³	<10 ⁻³
Muy confiable vs confiable		0.023	0.005	<10 ⁻³	0.603	0.090	0.125
Muy confiable vs poco confiable		0.027	<10 ⁻³	0.004	0.076	<10 ⁻³	<10 ⁻³
Confiable vs poco confiable		0.289	0.009	0.115	0.088	<10 ⁻³	<10 ⁻³

^a Tasa de pacientes bien clasificados con el uso de 7,1 kPa como el punto de corte de LSE para el diagnóstico de fibrosis significativa (F_M≥2), 12,5 kPa para el diagnóstico de cirrosis (F_M4), o clasificación de LSE (F_{FS}0/1, F_{FS}2/3, F_{FS}4) derivada de los 2 puntos de corte del diagnóstico anteriores (Castera y otros); ^b Este resultado, que ya se presentó en la tabla 2, se proporciona aquí para la comparación con subgrupos; ^c p para la tendencia lineal en los 3 subgrupos de LSE

Por lo tanto, las LSE en este subgrupo se consideraron "muy confiables". Las LSE con 0,10< IQR/M ≤0,30 o con IQR/M >0,30 y mediana de LSE <7,1 kPa proporcionaron una exactitud similar a la de la población completa y por lo tanto se consideraron "confiables". Finalmente, las LSE con IQR/M >0,30 y mediana de LSE ≥7,1 kPa proporcionaron una exactitud menor que la de la población completa y por lo tanto se consideraron "poco confiables". Los umbrales se determinaron como una función de los datos de la literatura y/o técnicas estadísticas clásicas para la determinación del punto de corte. La exactitud de las LSE en el subgrupo de LSE con IQR/M ≤0,10 fue mayor que en la población completa. Por lo tanto, las LSE en este subgrupo se consideraron "muy confiables". Las LSE con 0,10< IQR/M ≤0,30 o con IQR/M >0,30 y mediana de LSE <7,1 kPa proporcionaron una exactitud similar a la de la población completa y por lo tanto se consideraron "confiables". Finalmente, las LSE con IQR/M >0,30 y mediana de LSE ≥7,1 kPa proporcionaron una exactitud menor que la de la población completa y por lo tanto se consideraron "poco confiables".

De acuerdo con estos nuevos criterios, 16,6 % de las LSE se consideraron "muy confiables", 74,3 % "confiables", y 9,1 % "poco confiables". Lo que es más importante, las AUROC de LSE y las exactitudes de los diagnósticos fueron significativamente diferentes entre estos 3 subgrupos (Tabla 13). Finalmente, la tasa de LSE poco confiable de acuerdo con los nuevos criterios fue significativamente menor que la de LSE no confiable de acuerdo con la definición usual (9,1 % vs 24,3 %, p<10⁻³).

Análisis de sensibilidad

Los nuevos criterios se evaluaron en cuanto a la confiabilidad de la LSE como una función de varias características influyentes potenciales: causa de la enfermedad hepática (CHC vs. otras), índices de diagnóstico (AUROC, diagnóstico

5 binario de fibrosis significativa o cirrosis, clasificación de LSE), y puntos de corte del diagnóstico publicados por Ziol y *otros*, Stebbing y *otros*, y Friedrich-Rust y *otros*. En resumen, independientemente del factor influyente potencial, una disminución de la confiabilidad de la LSE, de acuerdo con los nuevos criterios, se asoció con una disminución de la exactitud de la LSE. El BMI (<25 vs ≥ 25 kg/m²) no influyó en la exactitud de la LSE en ninguna de las 3 categorías nuevas de confiabilidad de la LSE. Debido al bajo número de pacientes con hepatitis B, abuso de alcohol, o NAFLD, no fue posible realizar un análisis de sensibilidad para estas causas de la enfermedad hepática crónica.

Reivindicaciones

- 5 1. Un método para mejorar la confiabilidad de una prueba diagnóstica no invasiva de la presencia y/o la gravedad de una enfermedad hepática, dicho método comprende:
- 10 a. recopilar un índice de diagnóstico, llamado Índice Inicial, que implica la combinación matemática de al menos dos datos, en donde el Índice Inicial se selecciona de pruebas de sangre que comprenden ELF, FibroSpect, APRI, FIB-4, Hepascore, Fibrotest, pruebas de la familia FibroMeter, y pruebas derivadas de la familia FibroMeter donde la urea se ha eliminado de los marcadores;
- 15 b. analizar la confiabilidad de cada uno de los datos combinados para recopilar el Índice Inicial mediante la identificación de si al menos un dato del Índice Inicial recopilado en la etapa a) es un dato anómalo, un dato inconsistente, o un dato no homogéneo, o es responsable de una disminución en el Índice de Dispersión mayor que la observada con otros datos:
- 20 i. con la comparación de cada dato con el dato esperado en una población de referencia, o
 ii. en vista del(de los) predictor(es) de confiabilidad intrínseco(s) o extrínseco(s), o
 iii. mediante el cálculo del Índice de Dispersión del Índice Inicial recopilado en la etapa a), un Índice de Dispersión alto significa datos homogéneos del Índice Inicial, y la realización de una serie de cálculos del Índice de Dispersión del Índice Inicial que comprende n datos donde se eliminan 1 a (n-2) datos del Índice Inicial, lo que da lugar a la clasificación de los datos que más reducen el Índice de Dispersión;
- 25 c. si un dato es anómalo, inconsistente o un dato no homogéneo, o un dato que reduce un Índice de Dispersión, generar una Alerta de Evento;
 d. si se genera una Alerta de Evento, calcular nuevos índices, donde el dato anómalo, inconsistente y/o no homogéneo, o un dato responsable del Índice de Dispersión más bajo, se suprime para obtener de este modo un Índice Alternativo, o donde el dato anómalo, inconsistente y/o no homogéneo se sustituye por su valor medio para obtener de este modo un Índice Estimado o, si al menos dos datos son anómalos, inconsistentes o no homogéneos, el más discordante se suprime y el(los) otro(s) se sustituye(n) por su valor medio para obtener de este modo un Índice Mixto;
- 30 e. reemplazar el Índice Inicial que comprende un dato anómalo, inconsistente y/o no homogéneo o un dato que afecta el Índice de Dispersión, con un Índice Alternativo, un Índice Estimado, o un Índice Mixto.
- 35 2. El método de acuerdo con la reivindicación 1, en donde:
 - en la etapa a) se recopila un Índice Inicial, dicho Índice Inicial es una prueba de la familia FibroMeter, o una prueba derivada de la familia FibroMeter donde la urea se ha eliminado de los marcadores;
 - en la etapa b) el Índice de Dispersión del Índice Inicial recopilado en la etapa a) se calcula, para identificar de este modo cuál es el dato del Índice Inicial recopilado en la etapa a) que más disminuye el Índice de Dispersión, mediante la realización de una serie de cálculos del Índice de Dispersión del Índice Inicial que comprende n datos donde se eliminan 1 a (n-2) datos del Índice Inicial;
 - en la etapa e) el Índice Inicial que comprende un dato que más reduce el Índice de Dispersión se reemplaza con un Índice Alternativo donde el dato que reduce el Índice de Dispersión se elimina.
- 40 3. El método de acuerdo con la reivindicación 2, en donde la etapa c) se realiza después de calcular si el Índice de Dispersión del Índice Inicial corresponde a los de sus deciles intermedios, preferentemente del 2^{do} al 9^{no} decil, de la población de referencia.
- 45 4. El método de acuerdo con cualquiera de las reivindicaciones 1 a 3, en donde cuando un dato es un dato anómalo, la Alerta de Evento es una Alerta de Dato Anómalo; cuando un dato es un dato inconsistente, la Alerta de Evento es una Alerta de Dato Inconsistente; cuando un dato es un dato no homogéneo, la Alerta de Evento es una Alerta de Dato no Homogéneo.
- 50 5. El método de acuerdo con cualquiera de las reivindicaciones 1 a 4, en donde las Alertas de Evento emitidas en la etapa c) se analizan preliminarmente, antes de la etapa d), y en donde dicho tratamiento comprende:
- 55 - la supresión de una o más Alertas de Evento cuando el dato que ha activado la Alerta de Evento se encuentra confiable, y/o
 - cuando se emiten varias Alertas de Evento, la priorización de dichas Alertas de Evento para identificar una Alerta de Evento principal de acuerdo con su impacto sobre el Índice Inicial.
- 60 6. El método de acuerdo con la reivindicación 1, en donde los predictores de confiabilidad se seleccionan del grupo que comprende urea, ALT, AST, puntuación de CirrhoMeter^{2G}, puntuación de FibroMeter^{2G}, clases de Fibroscan, clases de FibroMeter, clases de CirrhoMeter, mediana de Fibroscan, IQR, IQR/M, plaquetas, A2M, relación urea/Fibroscan e Índice de Dispersión del Índice Inicial.
- 65 7. El método de acuerdo con la reivindicación 1, en donde en la etapa b) la confiabilidad de cada dato se analiza mediante la identificación de si al menos un dato del Índice Inicial recopilado en la etapa a) es un dato anómalo,

un dato inconsistente, o un dato no homogéneo, o es responsable de una disminución en el Índice de Dispersión mayor que la observada con otros datos con comparación de cada dato con el dato esperado en una población de referencia, o en vista del(de los) predictor(es) de confiabilidad intrínseco(s) o extrínseco(s).

- 5 8. Un método de acuerdo con cualquiera de las reivindicaciones 1 a 7, que comprende en la etapa e) reemplazar el Índice Inicial con el índice más fiable seleccionado del grupo que consiste en un Índice Alternativo, un Índice Estimado, o un Índice Mixto.
- 10 9. Un sistema experto que implementa el método de acuerdo con cualquiera de las reivindicaciones 1 a 8.
- 10 10. Un software que implementa el método de acuerdo con cualquiera de las reivindicaciones 1 a 8.

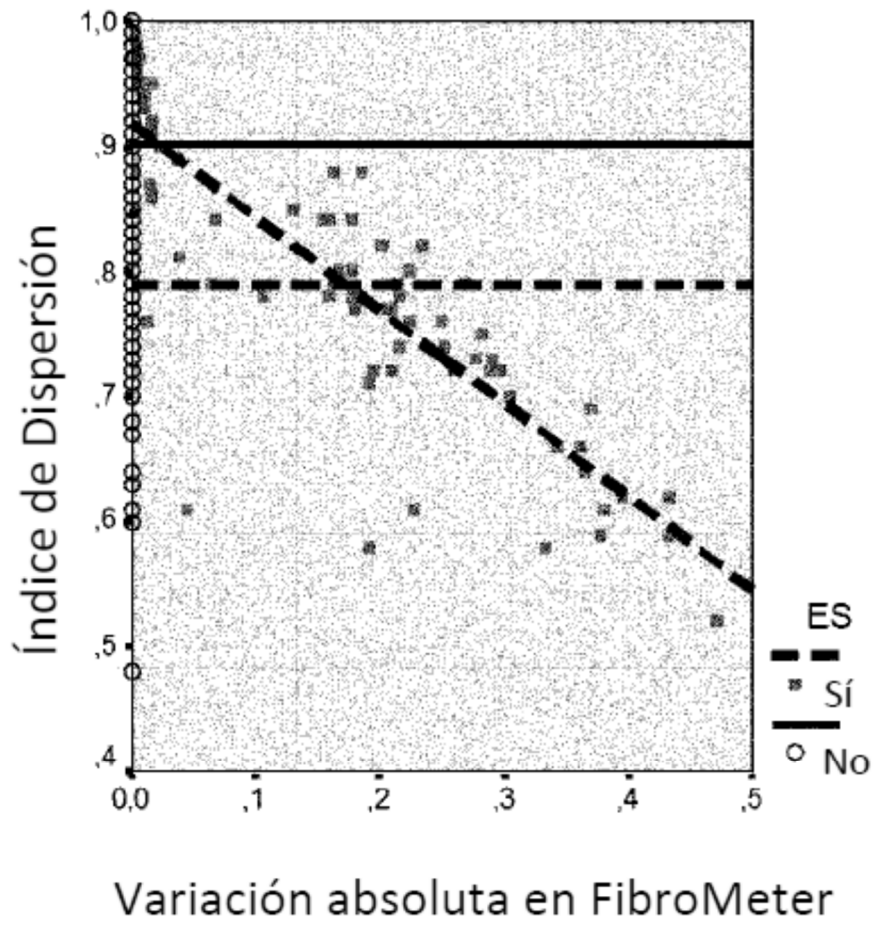


Figura 1

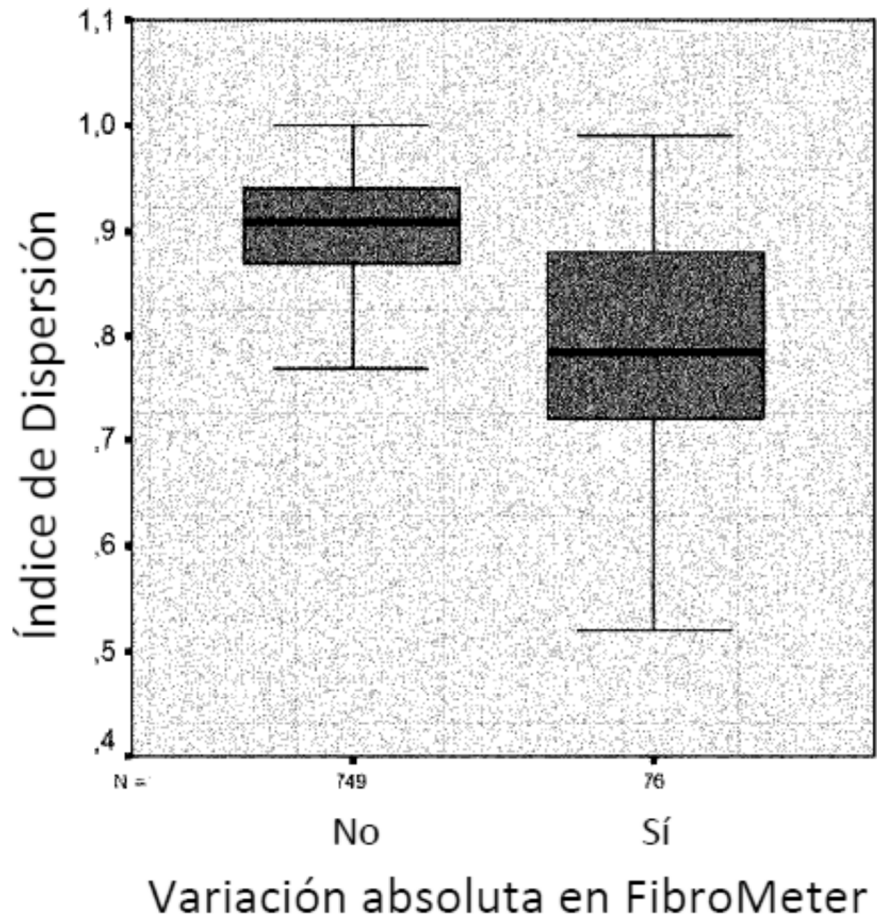


Figura 2

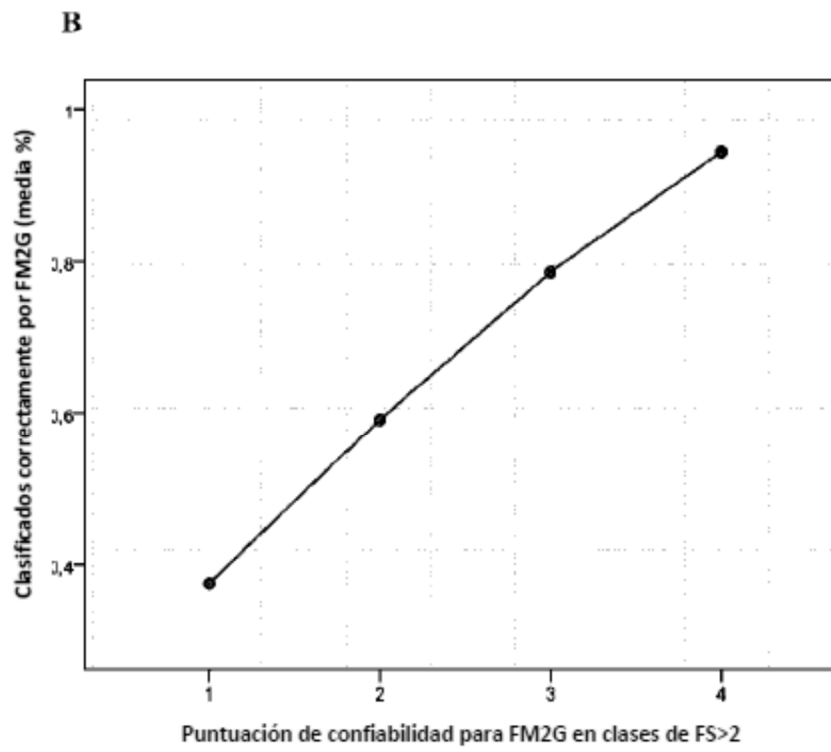
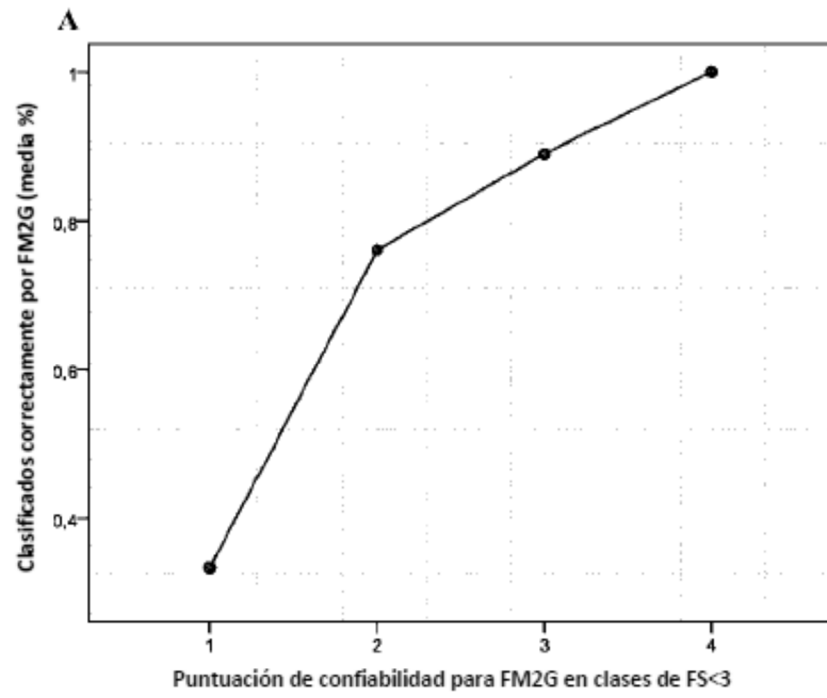


Figura 3

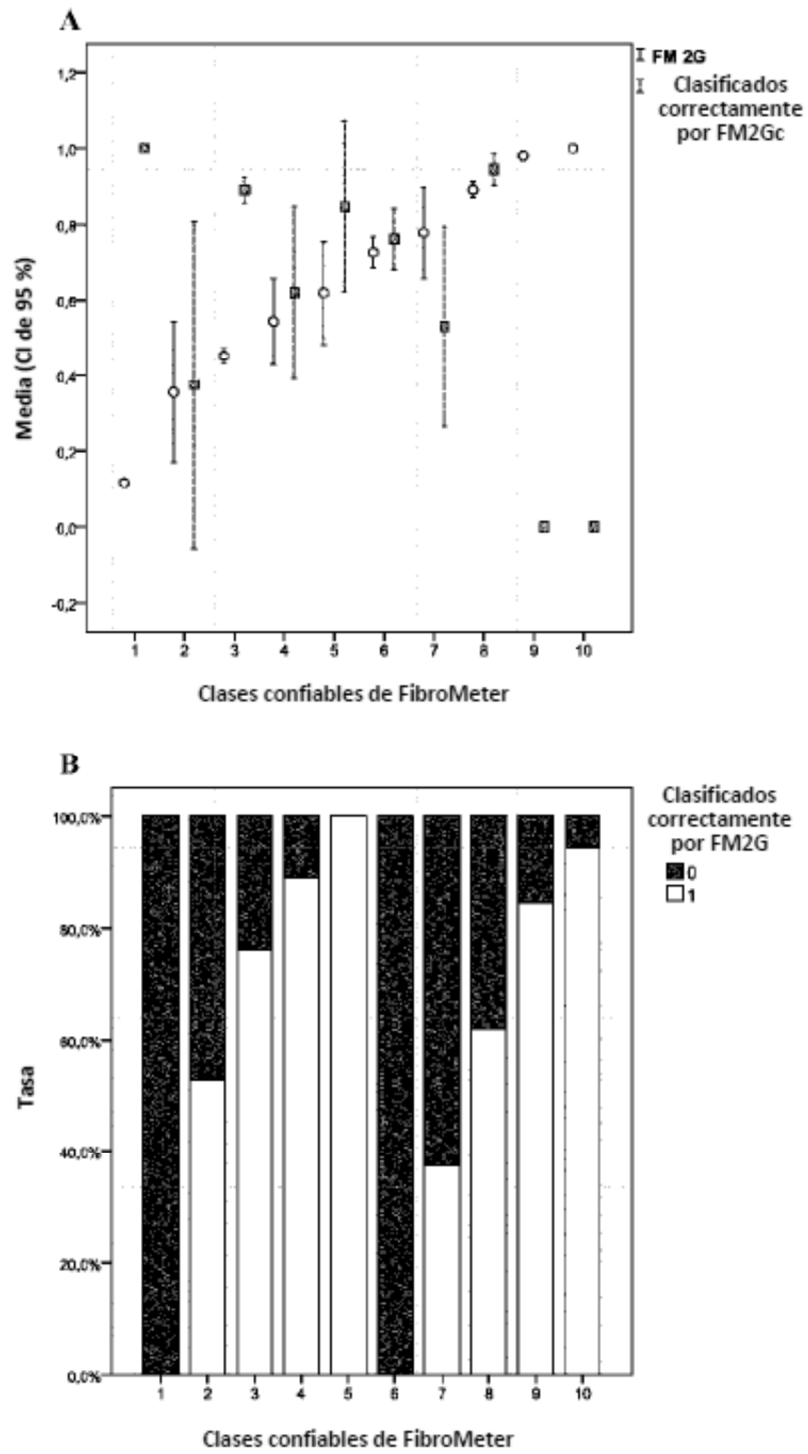


Figura 4

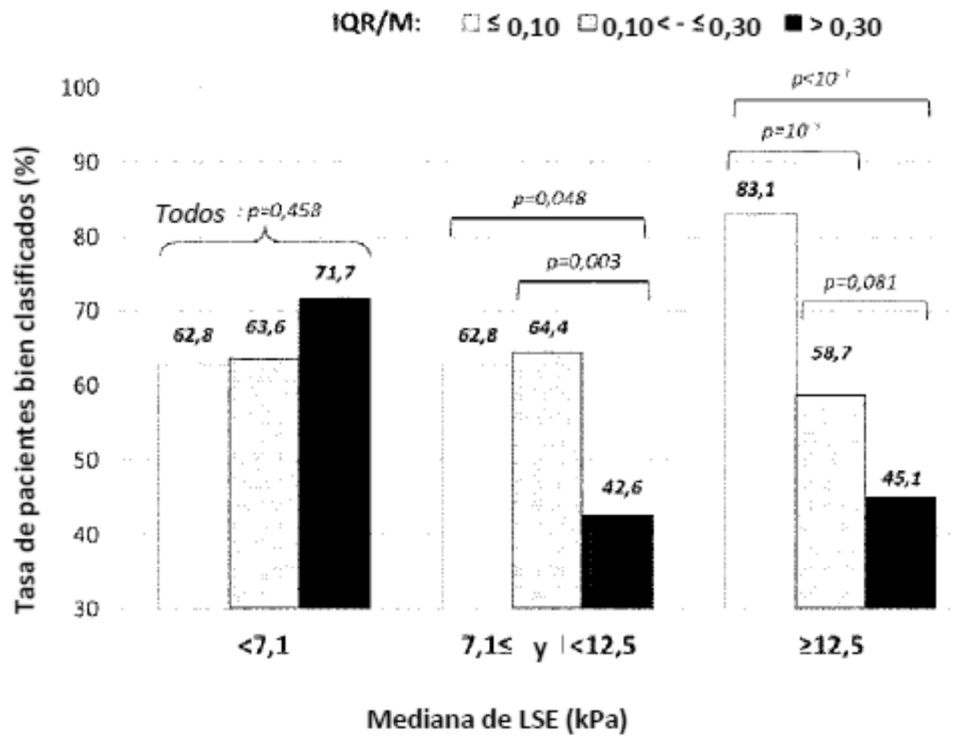


Figura 5

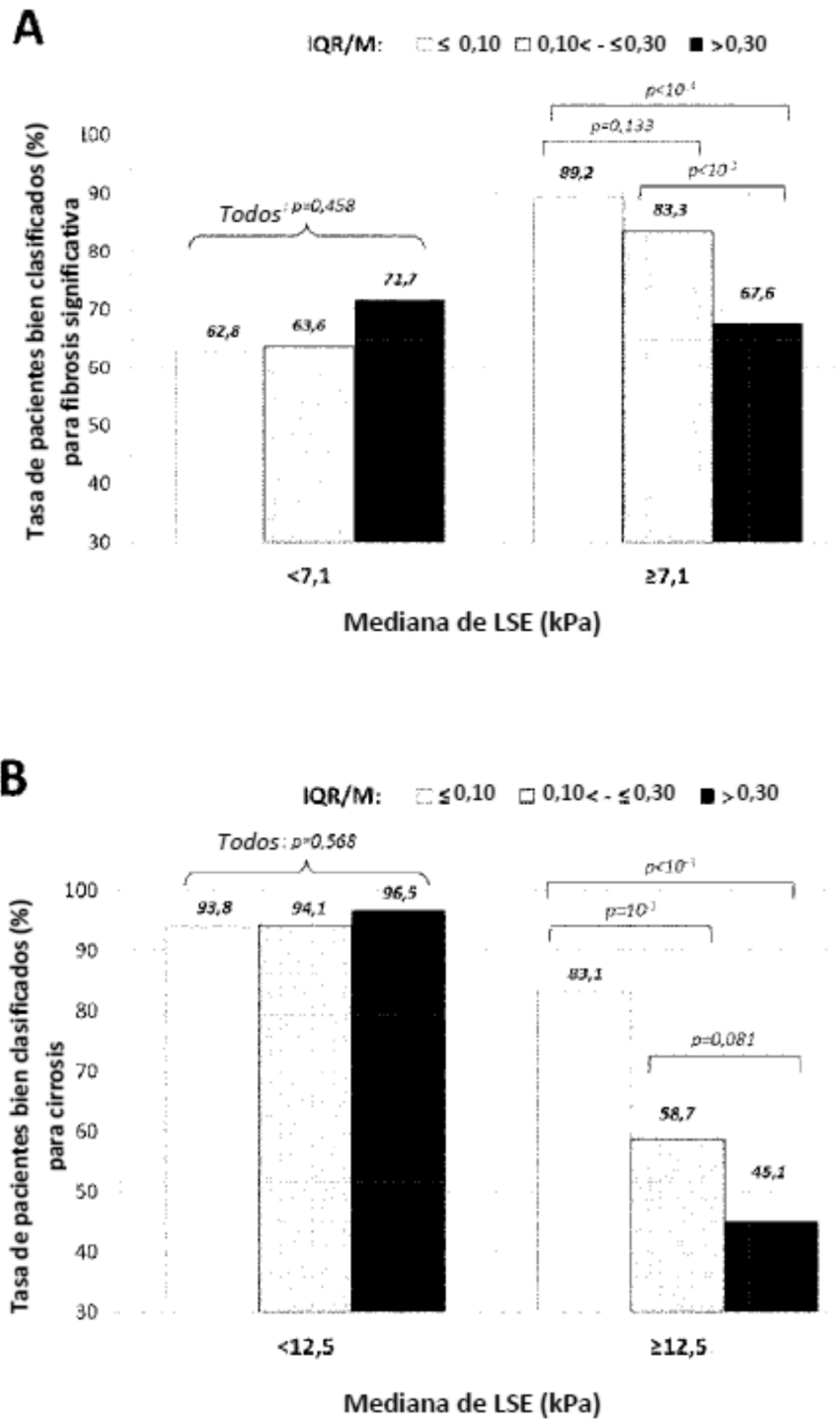


Figura 6