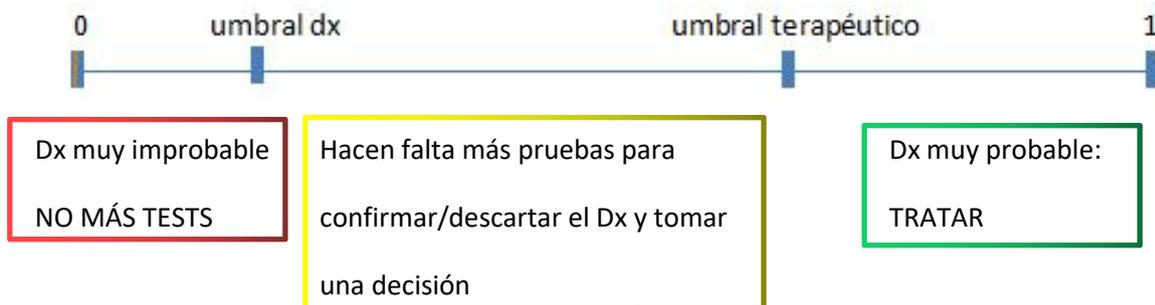


**Multicenter evaluation of BioFire FilmArray Meningitis/Encephalitis panel for detection of bacteria, viruses and yeast in cerebrospinal fluid. Leber AL et al. J Clin Microbiol 2016; 54 (9): 2251-61.**

Victoria Lobo, Alejandro Martí, Nieves Aparisi, Miguel Ángel García 8 - JUNIO - 2023

Una prueba diagnóstica es un procedimiento que nos permite confirmar o descartar un diagnóstico, o aumentar o disminuir su verosimilitud. No nos dirá si una persona está sana o enferma con certeza, su resultado permite aumentar o disminuir la probabilidad (p) de que un individuo esté sano o enfermo. Su utilidad es la capacidad para disminuir la p por debajo del umbral de la prueba (p por debajo de la cual es tan poco probable que esté enfermo que descartamos la enfermedad de forma razonable) o aumentar la probabilidad hasta superar el umbral de tratamiento (valor por encima del cual confirmamos razonablemente el diagnóstico e iniciamos el tratamiento).



Una prueba es válida si detecta la mayoría de personas con enfermedad, descarta a la mayoría de personas sin enfermedad, sus resultados positivos indican alta p de que la enfermedad esté presente y sus resultados negativos indican alta p de que la enfermedad esté ausente.

Características de una prueba diagnóstica:

FIABILIDAD o PRECISIÓN: capacidad de obtener los mismos resultados cuando se repite en condiciones similares. Influyen: variabilidad biológica, las características inherentes al test, influencia del observador.

Se podría valorar (sobre todo en pruebas que dependen del observador) la variabilidad intra e interobservadores con los índices kappa y correlación intraclase.

En pruebas cuya interpretación no depende del observador, los descriptores se valoran en sentido vertical (S y E) que indican **validez** de la prueba, y en sentido horizontal (VPP y VPN) que nos informan de la **seguridad** de la prueba. Otros índices son los cocientes de probabilidad positivo ( $CP+ = LHR+ = S / (1-E)$ ): indica cuánto más probable es obtener un resultado positivo en enfermos que en sanos, y negativo ( $CP- = (1-S)/E$ ) indica cuánto es más probable es obtener un resultado negativo en enfermos que en sanos.

### A/ ¿Son válidos los resultados del estudio?

#### Preguntas de eliminación

1. ¿Existió una comparación con una prueba de referencia adecuada?

*PISTA: ¿Es correcto el patrón oro? (no siempre se puede aplicar el mismo patrón otro a todos los pacientes?)*

Realiza comparación de cultivo de líquido cefalorraquídeo (LCR) bacteriano, viral y hongos (apartados en Material y Métodos) PERO NO PACIENTES CON SOSPECHA MAYOR O MENOR DE MENINGOENCEFALITIS (ME). Más que patrón oro, que podía construirse con una elevada sospecha clínica con una/unas pruebas complementarias de confirmación de esa sospecha, en

este estudio el resultado de esta prueba se compara con los aislamientos bacterianos obtenidos con procedimientos MICRO estándar validados y usados en cada lugar (distintos en cada centro: cultivos, fenotipos, proteómica según MALDI-TOF); se realizan tinciones de Gram con LCR citocentrifugado con muestras con cultivo positivo y con cultivo negativo.

2. ¿Incluyó la muestra un espectro adecuado de pacientes?.....NO LO SÉ

*PISTAS: ¿están adecuadamente descritos los pacientes y cómo se seleccionaron?  
Casi cualquier prueba distingue entre sanos y gravemente enfermos*

No incluye pacientes, sospechamos que deben ser enfermos con sospecha de ME, con una prevalencia del 6,7%. Se trata de muestras de LCR de distintos puntos de EEUU durante aproximadamente 8 meses (febrero a septiembre 2014). Criterios de inclusión de las muestras:

- LCR recogido por punción lumbar con adecuado volumen residual tras centrifugar ( $\geq 0,5$  ml) que se deja para tests estándar de cultivo bacteriano
- la muestra se pudo incluir en los primeros 7 días tras su recolección para la realización de tests (o se congeló para un procesado posterior)

Se excluyeron muestras repetidas del mismo paciente.

Se identificaron de forma adecuada las alícuotas del LCR extraído de cada paciente. Se recogen datos clínicos y demográficos, como estado de hospitalización del sujeto en el momento de la recogida del LCR, datos bioquímicos (número de leucocitos/mm<sup>3</sup>, proteínas, glucosa), cualquier test adicional de LCR requerido, resultados del cultivo bacteriano final de LCR; estado inmune, tratamientos antimicrobianos administrados en las 24 h previas a la recogida del LCR y diagnóstico clínico final en las muestras con resultados positivos para gérmenes de ME (con tests estándar de laboratorio de MICRO, FilmArray o comparadores de PCR si la glucorraquia ( $\leq 45$  mg/dl), proteinorraquia ( $\geq 100$  mg/dl) o número de leucocitos ( $\geq 5$ /mm<sup>3</sup>) era indicativo de infección.

El espectro de pacientes se describe en el apartado Demographics de Resultados: 51% hombres, 93% de pacientes hospitalizados (59%) o en el servicio de Urgencias (34%); el resto fueron pacientes valorados en CC Externas. 59% adultos y el restante 41% < 18 años.

Suponemos que la PL se hace con una sospecha no pequeña de ME (aunque posiblemente en el 7% de pacientes en los que se remite la muestra desde CC Externas no se realice la PL con la sospecha de ME -por ejemplo cuadros infecciosos larvados o subagudos, estudios de demencia, etc-).

3. ¿Existe una adecuada descripción de la prueba?

*PISTAS: ¿Se define con claridad qué es un resultado positivo y qué es un resultado negativo? ¿Se especifica la reproducibilidad de la prueba? (éste puede ser un punto clave en pruebas que dependen del observador como las técnicas de imagen)*

Explica con términos técnicos cómo se realiza la multiPCR (sistema automatizado con pasos sucesivos: preparación, transcriptasa inversa, PCR y detección), que debe ser una técnica estándar y totalmente reproducible (con curva de aprendizaje necesaria, como cualquier técnica de cierta complejidad). También los comparadores que usa (el correspondiente al gold estándar) se explican de forma pormenorizada. NO TENGO CONOCIMIENTOS SUFICIENTES PARA VALORAR SI ES O NO LA INFORMACIÓN NECESARIA PARA CONFIRMAR QUE LA REPRODUCIBILIDAD ES ADECUADA. No es una prueba dependiente de observador.

#### Preguntas “de matiz”

4. ¿Hubo evaluación “ciega” de los resultados?

PISTA: ¿Las personas que interpretaron la prueba conocían los resultados del patrón oro (o viceversa)?

No se aclara este apartado. Pero da la impresión de que se manejan las muestras de LCR de forma independiente del resultado de la prueba comparadora. Sí describe que el trabajo y la validación del resultado se hizo de forma ciega. El estudio FilmArray se hizo simultáneamente al cultivo y Gram en todos los casos. En algunos casos con FilmArray y cultivo negativos, el Gram se revisó y se confirmó el FilmArray. Hay motivos para que el cultivo sea negativo.

5. ¿La decisión de realizar el patrón oro fue independiente del resultado de la prueba problema?

PISTAS: considerar si

- i. Se incluyeron preferentemente los resultados positivos en la prueba a evaluar;
- ii. Se utilizaron diferentes patrones oro en los positivos y en los negativos..

Es complicado de realizar, pero NO SE VALORA LA POSIBILIDAD DE ME CON CULTIVOS ESTÁNDAR NEGATIVOS (que quizá pueden ser limitados, con FFNN -infecciones que no se detectan- o FFPP -gérmenes que “aparecen” en LCR pero que no son responsables de la infección-). Lo aclara en las últimas líneas del apartado Material y Métodos: “el diagnóstico clínico final sólo se obtuvo de muestras positivas (positivo para FilmArray o positivo para los parámetros de LCR sugestivos de infección) pero no de todas las muestras, no fue posible calcular valor predictivo positivo o negativo. La biología molecular parece superar al cultivo en S y E, pero es necesario el cultivo para el aislamiento de bacterias poco frecuentes en este tipo de paneles, como en infecciones nosocomiales, para estudios de sensibilidad, etc. Se sigue manteniendo como prueba de oro.

Tabla de hallazgos en LCR en pacientes con meningitis bacteriana aguda (fuente: Enfoque clínico de los grandes síndromes infecciosos. 2ª edición. J Gómez, M Gobernado).

**TABLA V. Hallazgos en LCR en pacientes con MAB**

Presión LCR	200-500 mm H <sub>2</sub> O (media: 370 ± 130)
Apariencia	Turbio
Pleocitosis	1.000-5.000/mm <sup>3</sup> (intervalo: 100 a >10.000)
Tanto por ciento de neutrófilos	≥80%
Proteínas	100-500 mg/dl (media: 490 ± 450)
Relación glucosa LCR/glucemia (glucorraquia)	≤0,4 (<40 mg/dl) (media: 0,2 ± 0,2)
Tinción de Gram positiva	60-90%
• <i>Haemophilus influenzae</i> tipo b	25-65%
• <i>Neisseria meningitidis</i>	30-89%
• <i>Streptococcus pneumoniae</i>	69-83%
• <i>Listeria monocytogenes</i>	75%
• <i>Streptococcus agalactiae</i>	80-90%
• Otras <i>Streptococcus</i> spp.	50-73%
• <i>Staphylococcus aureus</i>	22-44%
Cultivo positivo	70-85%
• <i>Haemophilus influenzae</i> tipo b	25-90%
• <i>Neisseria meningitidis</i>	40-60%
• <i>Streptococcus pneumoniae</i>	60-90%
• <i>Listeria monocytogenes</i>	35%
• <i>Streptococcus agalactiae</i>	80-85%
• Otras <i>Streptococcus</i> spp.	50-65%
• <i>Staphylococcus aureus</i>	75-100%
Detección de antígenos bacterianos	50-100%
• <i>Haemophilus influenzae</i> tipo b	78-100%
• <i>Neisseria meningitidis</i>	50-93%
• <i>Streptococcus pneumoniae</i>	67-100%
• <i>Streptococcus agalactiae</i>	69-100%
Reacción en cadena de la polimerasa específica	
• <i>Haemophilus influenzae</i> tipo b	Sensibilidad 94%, especificidad 96%
• <i>Neisseria meningitidis</i>	Sensibilidad 91-97%, especificidad 91-99,6%
• <i>Streptococcus pneumoniae</i>	
• <i>Listeria monocytogenes</i>	
• <i>Escherichia coli</i>	
Reacción en cadena de la polimerasa semi-anidada	
• <i>Haemophilus influenzae</i>	Sensibilidad 94%, especificidad 96%
• <i>Neisseria meningitidis</i>	
• <i>Streptococcus</i> spp.	
Reacción en cadena de la polimerasa de amplio espectro	Sensibilidad 94%, especificidad 96%
• <i>Haemophilus influenzae</i> tipo b	67-100%      97-100%
• <i>Neisseria meningitidis</i>	89-100%      95-100%
• <i>Streptococcus pneumoniae</i>	79-100%      95-100%
Determinación de lactato	Sensibilidad 96%, especificidad 100%

**B/ ¿Cuáles son los resultados?**

6. ¿Se puede calcular los cocientes de probabilidad (likelihood ratios)?

PISTAS: ¿Se han tenido en cuenta los pacientes con resultados “no concluyentes”? ¿Se pueden calcular los cocientes de probabilidad para distintos niveles de prueba, si procede?

7. ¿Cuál es la precisión de los resultados?

PISTA: Hay que buscar o calcular los intervalos de confianza de los cocientes de probabilidad

Los resultados se pueden montar en la calculadora CASPE. Al sumar los resultados positivos de la tablas 2 y 3 encontramos 104 resultados positivos del film array y 141 totales positivos a las pruebas consideradas patrón oro.

INDICE	CARACTERÍSTICA EVALUADA		
	PRESENTE (Prueba de referencia +)	AUSENTE (Prueba de referencia -)	
PRUEBA DIAGNÓSTICA +	98	43	141
PRUEBA DIAGNÓSTICA -	6	1413	1419
	104	1456	1560

		IC 95%
Sensibilidad	94,2%	88,0% a 97,3%
Especificidad	97,0%	96,0% a 97,8%
Valor predictivo positivo	69,5%	61,5% a 76,5%
Valor predictivo negativo	99,6%	99,1% a 99,8%
Proporción de falsos positivos	3,0%	2,2% a 4,0%
Proporción de falsos negativos	5,8%	2,7% a 12,0%
Exactitud	96,9%	95,9% a 97,6%
Odds ratio diagnóstica	536,72	222,99 a 1291,87
Índice J de Youden	0,9	
CPP o LR(+)	31,91	23,68 a 43,00
CPR o LR(-)	0,06	0,03 a 0,13
Probabilidad pre-prueba (Prevalencia)	6,7%	

En el apartado SUMMARY OF FILMARRAY ME PANEL FINDINGS, se citan 136 de las 1560 muestras con algún patógeno encontrado. Se podría obtener una tabla similar a la previa, en la que el total de resultados positivos fueran 136 con los mismos FN (43) y FP (6); los resultados de S, E, VPP y VPN son análogos,

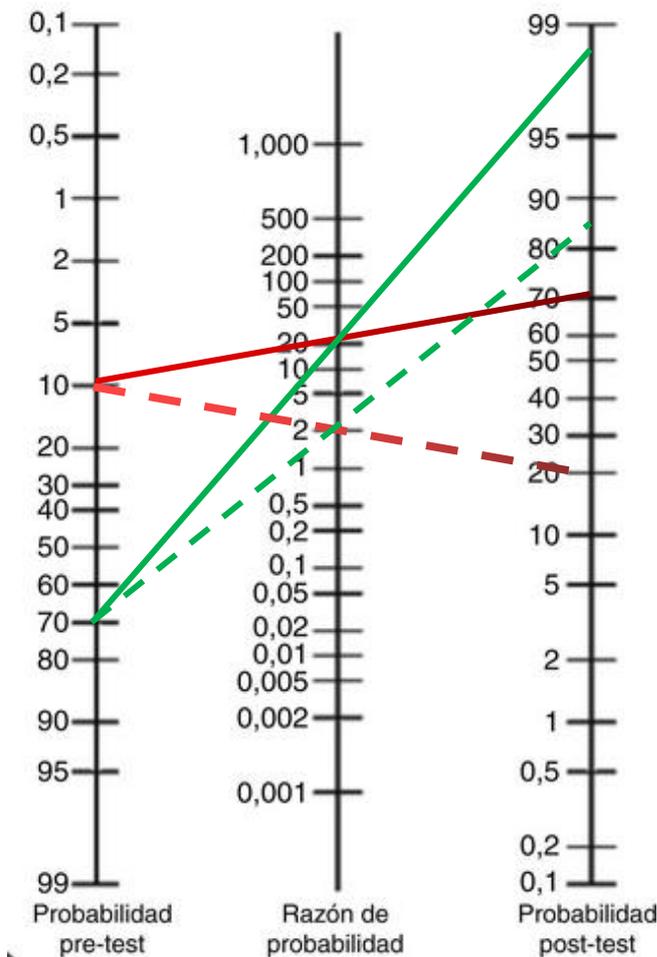
	enfermos	sanos	totales
Test +	93	43	136
Test -	6	1418	1424
Total	99	1461	1560

La diferencia entre ambas tablas radica en que hay 5 muestras con doble detección de germen (que se considera FP).

Con estos resultados se pueden conseguir valores de cocientes de probabilidad positivos y negativos (CPP y CPN) muy adecuados.

Los resultados de aislamientos de meningitis bacterianas son tan escasos que es muy complicado obtener conclusiones categóricas (22 de 1560 muestras). Y es llamativo que no se aísle ningún meningococo ni ninguna Lysteria.

La forma de entender el CPP es el siguiente:



LÍNEA ROJA CONTINUA: Asumimos probabilidad pre-prueba (p test) del 10%. trazamos línea por el nomograma desde ese punto pasando por el cociente de probabilidad positivo que obtenemos (nos quedamos con una estimación cercana a 20, que es cercana al límite inferior del intervalo de confianza que hemos obtenido) = corresponde a una probabilidad postprueba cercana al 70%!!! Un test con un CPP muy elevado hace que, a partir de una probabilidad pretest muy baja la convirtamos en una probabilidad posttest elevada, hace que la sospecha diagnóstica sea elevada (sea superior al umbral terapéutico) y nos planteemos en serio iniciar tratamiento específico

Podemos hacer esa misma deducción con cálculos numéricos con 2 fórmulas aplicando el teorema de Bayes:

\*ODDS es una forma de expresar el riesgo de una forma distinta a la probabilidad =

$$\text{ODDS} = P (\text{probabilidad}) / (1 - P)$$

$$\text{La conversión de ODDS a PROBABILIDAD es: } p = \text{ODDS} / (1 + \text{ODDS})$$

\*ODDS POST-PRUEBA = COCIENTE DE PROBABILIDAD X ODDS PRE-PRUEBA

En ese caso: p pretest 10%= 0,1 -> odds pretest = 0,1111

Odds post-test = 20 x 0,1111 = 2,2222 -> p post-test = 2,2222 / 3,2222 = 68,96% !!!!

Para el ejemplo de la línea roja discontinua, con un cociente de probabilidad mucho más discreto, de 2, los resultados no son tan brillantes: un resultado de p posttest cercano al 20% que se confirma en el cálculo numérico:

Odds post-test = 2 x 0,1111 = 0,2222 -> p post-test = 0,2222 / 1,2222 = 18,18%

La probabilidad posttest es algo mayor, pero aún inferior al umbral terapéutico.

Si partimos de una p pretest mucho mayor, 70% (líneas verdes), el resultado es distinto:

Para cociente probabilidad alto (20), la p posttest “visual” es superior al 95%. Numéricamente:

p pretest 0,7 -> odds pretest 2,3333

Odds post-test = 20 x 2,3333 = 46,6666 -> p posttest = 46,6666 / 47,6666 = 97,7%!!!!

Nos podemos hacer un planteamiento. Si la p pretest ya es lo suficientemente elevada, superior al umbral terapéutico, quizá no tenga sentido hacer un test diagnóstico de confirmación. O sí, si con un 70% dudamos de poner tratamiento, porque es caro o tóxico, pero con >90% parece que el diagnóstico es altamente probable, deberemos iniciar el tratamiento.

Si repetimos el proceso con cociente de probabilidad modesto (2):

- línea discontinua, resultado algo mejor que el pretest, algo superior al 80%

- odds posttest = 2 x 2,3333 = 4,6666 -> p posttest = 4,6666 / 5,6666 = 82,35%!!!!

Es el momento de plantear qué definimos como umbral de tratamiento. A PARTIR DE QUÉ PROBABILIDAD DE ENFERMEDAD NOS DECIDIMOS A INICIAR EL TRATAMIENTO, sin tener la confirmación 100% de que el paciente tiene la enfermedad (algo que ocurre muy a menudo en Medicina). Si es una enfermedad grave (como la ME) en el que no podemos demorar el inicio de un tratamiento que no es tóxico, esa probabilidad no debe ser excesivamente alta (quizá 40-60%). Si la enfermedad no es tan grave, y/o el tratamiento es tóxico, quizá podríamos tener un umbral de tratamiento mayor (70-80%)

### C/ ¿Son los resultados aplicables al escenario?

8. ¿Serán satisfactorios en el ámbito del escenario la reproducibilidad de la prueba y su interpretación?

*PISTA: Considera el ámbito de la prueba es demasiado diferente al del escenario*

La prueba está estandarizada, y es el pan de cada día en el trabajo de un laboratorio de MICRO. La reproducibilidad, tan una adecuada curva de aprendizaje, es buena. Sin embargo, la bajísima prevalencia de ME en el conjunto de muestras trabajadas hace que los resultados de este estudio no sean extrapolables a nuestro medio (conjunto de pacientes en los que se realiza PL con una probabilidad media o alta de ME).

9. ¿Es aceptable la prueba en este caso?

PISTA: Considera la disponibilidad de la prueba, los riesgos /molestias de la prueba y los costes.

La prueba es cara, pero se está incluyendo cada vez más dentro de la cartera de servicio de los servicios de MICRO. Un FilmArray cuesta 132 euros sin IVA. Pero su realización no supone una medida agresiva para el paciente. Y puede ahorrar tiempo en el diagnóstico microbiológico del cuadro de ME.

La prueba tiene sus limitaciones: no detecta S pyogenes o enterobacterias distintas de E coli K1, que pueden ser frecuentes en ciertos segmentos de edad.

10. ¿Modificarán los resultados de la prueba la decisión sobre cómo actuar?

PISTA: Desde la perspectiva del escenario, si la actitud no va a cambiar, la prueba es (al menos) inútil. Considera el umbral de acción y la probabilidad de enfermedad antes y después de la prueba.

ESTE TRABAJO NO ES UNA COMPARACIÓN CON EL GRAM Y CULTIVO DE LCR, con lo que no es planteable como alternativa a la realización de esas 2 técnicas, básicas y sencillas, para el manejo de pacientes con sospecha de ME. Como bien dicen en el apartado Discusión, debiera ser una prueba complementaria a valorar a realizar dentro del tratamiento de un líquido biológico tan valioso como el LCR. Es una interpretación errónea y simplista decir que a S y E de estas técnicas es superior a la del Gram y cultivo; el Film Array no incluye la detección de al menos un 15-20% de las bacterias causantes de ME (FN) y pueden detectarse restos de material genético de bacterias con dudoso papel en el desarrollo de ME (FP por contaminación en el procesado de la muestra). Además con el cultivo obtenemos la susceptibilidad a antibióticos. En un hospital comarcal con medios limitados este panel tiene utilidad sobre todo por su especificidad. CUIDADO CON EL VIRUS HERPES SIMPLE que puede ser negativo en los 3 primeros días. IMPORTANCIA en pacientes con clínica neurológica pero bioquímica de LCR normal.

La validez de la prueba para detectar cuadros víricos parece adecuada. Para cuadros bacterianos, la escasez de resultados nos da que pensar; no hay ningún caso de cultivo positivo a meningococo ni a Listeria !!!! (quizá más frecuentemente positivo en hemocultivos), con lo que debería recopilarse información sobre la utilidad de esta prueba para mirar a esos gérmenes.

Más que plantear este trabajo en términos de S y E, quizá podría haberse valorado la concordancia entre pruebas con un índice KAPPA .