



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**  
FACULTAD DE MEDICINA

**TRABAJO DE INVESTIGACIÓN PARA OPTAR POR EL GRADO  
ACADÉMICO DE BACHILLER EN MEDICINA**

**TÍTULO:**

**RENDIMIENTO DE LAS PRUEBAS DE AMPLIFICACIÓN DE  
ÁCIDOS NUCLEICOS (PAAN) EN EL DIAGNÓSTICO DE  
TUBERCULOSIS GENITOURINARIA: UNA REVISIÓN  
SISTEMÁTICA Y METANÁLISIS**

**ALUMNO:**

**CARLOS ALONSO ALTEZ FERNANDEZ**

**ASESOR:**

**CESAR UGARTE GIL**

**2016**

## **Contenido**

1. Resumen.....	1
2. Introducción.....	5
3. Métodos.....	7
4. Resultados.....	10
5. Discusión.....	13
6. Conclusiones.....	17
7. Referencias bibliográficas.....	24
8. Apéndice.....	27

## **Resumen**

**Introducción** El diagnóstico de la tuberculosis genitourinaria (TBGU) es difícil por las manifestaciones clínicas no específicas y los problemas con las pruebas convencionales. El retraso en el diagnóstico tiene un impacto severo en las vías urinarias. A pesar de que las pruebas de ácidos nucleicos dan resultados rápidos, hay una falta de consensos con respecto a su uso. Por lo tanto, nuestro objetivo es determinar el rendimiento de estas pruebas en el diagnóstico de la TBGU y evaluar la heterogeneidad entre los estudios.

**Métodos** Se realizó una revisión sistemática y metanálisis de los artículos de que comparen el rendimiento de una referencia estándar y una PAAN para el diagnóstico de TBGU. Se buscaron artículos publicados entre el 1 de enero de 1990 y el 14 de abril del 2016 en Medline, EMBASE, Web of Science, LILACS, Cochrane Library y Scopus. Se excluyeron los estudios de tipo caso-control por su alto riesgo de sesgo. Dos investigadores identificaron los artículos elegibles y extrajeron la data independientemente. Se analizó la data en grupos con el mismo test índice. Luego, se generaron medidas de resumen para cada grupo, a través de un metanálisis de efectos aleatorios. Se calculó la sensibilidad y especificidad resumen con un intervalo de confianza (IC) al 95% cuando los estudios no eran heterogéneos. El protocolo fue registrado en la base de datos PROSPERO con el numero CRD42016039020

**Resultados** Se identificaron once estudios relevantes de diez artículos, que dieron información sobre las pruebas de PCR, LCR y Xpert MTB/RIF. Todos los estudios con PCR fueron “pruebas hechas en casa”. Estos utilizaron diferentes primers genéticos y además tenían problemas de calidad, por lo que no procedimos con el metanálisis. Solo

un estudio uso LCR. Los estudios Xpert fueron de buena calidad y no heterogéneos, con una sensibilidad resumen de 0.87 (0.66-0.96) y especificidad de 0.91 (0.84-0.95).

**Discusión** Los estudios de PCR fueron altamente heterogéneos. En los estudios que usaron Xpert MTB/RIF, la especificidad fue favorable, con intervalos de confianza aceptables. Sin embargo, futuros estudios pueden actualizar el metanálisis y obtener estimados más precisos. Se requiere estudios con mayor calidad con urgencia para mejorar el diagnóstico de tuberculosis urogenital.

**Financiamiento** Este estudio no recibió ningún tipo de financiamiento.

**Palabras clave** Urogenital tuberculosis, Nucleic acid amplification tests, systematic review

## **Abstract**

**Introduction** Diagnosis of genitourinary tuberculosis(TBGU) is difficult because of unspecific clinical manifestations and problems with conventional tests. Delayed diagnosis impacts the urinary tract severely. Eventhough, nucleic acid amplification tests (PAANs) yield fast results, there is lack of consensus regarding their use. We therefore aimed to assess the accuracy of PAANs in the diagnosis of TBGU and to evaluate the heterogeneity between studies.

**Methods** We did a systematic review and meta-analysis of research articles comparing the accuracy of a reference standard and a PAAN for diagnosis of TBGU. We searched Medline, EMBASE, Web of Science, LILACS, Cochrane Library, and Scopus for articles published between Jan 1, 1990, and Apr 14, 2016. We excluded case-control studies for their high risk of bias. Two investigators identified eligible articles and extracted data for individual study sites. We analyzed data in groups with the same index test. Then, we generated pooled summary estimates (95% CIs) for sensitivity and specificity by a random-effects meta-analysis when studies were not heterogeneous. The protocol was registered at PROSPERO database with the number CRD42016039020

**Findings** We identified eleven relevant studies from ten articles, giving information on PCR, LCR and Xpert MTB/RIF tests. All PCR studies were “in-house” tests, with different gene targets and had several quality concerns therefore we did not proceed with a pooled analysis. Only one study used LCR. Xpert studies were of good quality and not heterogeneous, pooled sensitivity was 0·87 (0·66-0·96) and specificity was 0·91 (0·84-0·95)

**Discussion** PCR studies were highly heterogeneous. Among Xpert MTB/RIF studies, specificity was favorable with an acceptable confidence interval, however new studies can update meta-analysis and get more precise estimates. Further high-quality studies are urgently needed to improve diagnosis of genitourinary tuberculosis

**Funding** This study did not receive any funding.

**Keywords** Urogenital tuberculosis, Nucleic acid amplification tests, systematic review

## Introducción

La Tuberculosis (TB) es todavía una de las más grandes amenazas mundiales a la salud pública. En el 2014, la tuberculosis mató a 1,5 millones de personas (1). La Organización Mundial de la Salud (OMS) ha estimado 9.6 millones de nuevos casos de tuberculosis en 2014, el 12% de los cuales eran VIH-positivos (1).

La tuberculosis extra pulmonar (TBE) representa aproximadamente el 10% de toda la población afectada de tuberculosis. La tuberculosis genitourinaria (TBGU) es el tercer tipo más común de TBE, después de la TB pleural y ganglionar (2). Además, la TBGU asociada a TB pulmonar ocurre entre un 15 a 20% de los casos en los países en vías de desarrollo (2).

La TBGU se define como la infección por *Mycobacterium tuberculosis* de las vías urinarias, los genitales masculinos o los genitales femeninos. Sin embargo, la mayoría de los autores usan el término de TBGU para referirse solo al compromiso de las vías urinarias (3-5). La enfermedad ocurre en la mayoría de los casos después de largos períodos de infección latente, seguido por diseminación hematógena y siembra en los riñones, el epidídimo o las trompas de Falopio. La siembra en la próstata también se ha reportado, pero es rara. El resto de órganos genitourinarios se ven afectados por expansión local (5-7).

Las manifestaciones clínicas de la TBGU son poco específicas. Estas dependen de los órganos afectados y la gravedad del compromiso, además muchas veces, imitan la presentación de enfermedades urológicas y ginecológicas (8). El diagnóstico de la TBGU es difícil y con frecuencia se pasa por alto. Las pruebas diagnósticas convencionales incluyen la microscopía y el cultivo de orina. Dado que la TBGU suele ser paucibacilar (9), la mayoría de las muestras evaluadas por microscopía son

negativas. A pesar de que el cultivo es el *gold-standard* para el diagnóstico de TBGU, los métodos disponibles en la actualidad, tanto sólidos como líquidos, son poco sensibles, y tienen una serie de inconvenientes, incluyendo el largo tiempo para el crecimiento bacilar, las altas tasas de contaminación y los costos asociados (10).

El retraso en el diagnóstico da lugar a complicaciones como infertilidad, riñón no-funcionante unilateral, contracción vesical e incluso falla renal (8). Una serie de casos reportó que, al momento del diagnóstico, el 26.9% de los pacientes tenían un riñón no funcionante, 7,4% insuficiencia renal y 10% vejiga contraída (2,4).

Las pruebas de amplificación de ácidos nucleicos (PAAN), se han desarrollado para superar las limitaciones de las pruebas convencionales para TB (11-15). Estas pueden dar resultados entre 2 a 48 horas (16). Las PAAN se clasifican en comerciales o "in-house" (16). Pueden clasificarse, además, por su mecanismo: reacción en cadena de la polimerasa (PCR), reacción en cadena de la ligasa (LCR) y tecnologías más recientes (variantes del PCR) como Xpert MTB / RIF y Genotype® MTBDRplus (17,18).

Mientras que el uso de las PAAN ha sido bien estudiado en la TB pulmonar, su uso en la TBGU carece de suficiente evidencia (11,19,20). Realizamos esta revisión sistemática para determinar el rendimiento de las PAAN realizadas en orina para el diagnóstico de TBGU y para identificar los factores relacionados con la heterogeneidad de los resultados entre los estudios.



## **Métodos**

### **Estrategia de búsqueda y criterios de selección**

Nuestro estudio se desarrolló de acuerdo a las directrices del grupo Cochrane para revisiones sistemáticas y metanálisis de pruebas diagnósticas (21). El protocolo fue registrado en la base de datos PROSPERO con el código CRD42016039020. Se realizaron búsquedas en Medline, EMBASE, Scopus, LILACS, Cochrane Library y en Web of Science para los estudios publicados en cualquier idioma, a partir del 1 de enero de 1990 al 14 de abril, 2016 (véase el apéndice 1 para los términos de búsqueda).

Se incluyeron todos los estudios que compararon una PAAN en orina para el diagnóstico de GUTB con una referencia estándar. Estos estudios deben haber presentado datos en forma de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN). En los casos en que faltaban datos, se contactaron a los autores por correo electrónico para obtener información adicional.

Se incluyeron estudios que utilizaron una referencia estándar microbiológica o una referencia estándar amplia. Definimos la referencia estándar microbiológica como un cultivo positivo de *Mycobacterium tuberculosis* en medio sólido o líquido para una o más muestras de cada paciente. Definimos a la referencia estándar amplia como aquella que involucra un cultivo positivo o manifestaciones clínicas de la enfermedad con respuesta clínica después del tratamiento. Se excluyeron los estudios con menos de diez pacientes, por el alto riesgo de sesgo. También se excluyeron los estudios que tenían un diseño de casos y controles, ya que sobreestiman el rendimiento de las pruebas diagnósticas.

Dos revisores examinaron de forma independiente las citas recogidas para determinar la pertinencia y revisaron los artículos completos con los criterios de elegibilidad especificados previamente; los desacuerdos sobre la selección de los estudios se resolvieron mediante consulta con un tercer revisor (CU-G). Los resultados de la búsqueda en la literatura se presentan en un diagrama de flujo siguiendo las directrices PRISMA (figura 1) (22).

### **Extracción de datos y evaluación de calidad**

Dos revisores extrajeron los datos y los desacuerdos se resolvieron por consenso (el apéndice incluye detalles adicionales de la extracción de datos). Se evaluó la calidad de los estudios con la Herramienta de Evaluación de la Calidad de Estudios de rendimiento diagnóstico (QUADAS-2) (23). QUADAS-2 consiste en cuatro dominios: selección de pacientes, prueba índice, referencia estándar, y flujo y tiempo. Cada dominio se evaluó para evaluar el riesgo de sesgo y los tres primeros para determinar si hay preocupación por la aplicabilidad. Utilizamos preguntas, que en la herramienta se describen como de “señalización” para formar juicios sobre el riesgo de sesgo. Como es recomendado, desarrollamos una guía sobre cómo evaluar cada pregunta de señalización e interpretar la información adaptada a esta revisión. Un revisor (CA) puso a prueba la herramienta con cuatro de los estudios incluidos. Sobre la base de la experiencia adquirida en el estudio piloto, finalizamos la herramienta (véase el apéndice 2 para la evaluación de la calidad). Dos revisores evaluaron de forma independiente la calidad metodológica de los estudios incluidos con la herramienta finalizada; los desacuerdos acerca de la calidad metodológica se resolvieron mediante consulta con un tercer revisor (CU-G).

### **Análisis estadístico**

Se utilizaron los métodos estándar recomendados para los metanálisis de los estudios diagnósticos (21). El análisis fue hecho usando el programa estadístico R versión 3 (24,25).

Para minimizar la heterogeneidad proyectada, decidimos a priori analizar los datos en subgrupos: los estudios que utilizan PCR, los estudios que utilizan LCR y los estudios que utilizan Xpert MTB/RIF. Por cada subgrupo, se evaluó la heterogeneidad estadísticamente con la prueba de  $\chi^2$  y visualmente con los diagramas tipo forest plot y curvas de resumen operador-receptor (SROC). Generamos estimaciones agrupadas y diferencias para la sensibilidad y especificidad con intervalos de confianza al 95% mediante el uso de una curva ROC resumen bivariada (26). Se asumió que la sensibilidad y especificidad varían entre los estudios debido a las diferencias en las poblaciones, los errores de muestreo, y las diferencias en los umbrales implícitos aplicados a los datos. Por lo tanto, se aplicó un modelo de efectos aleatorios para tomar en cuenta de la heterogeneidad entre los estudios.

## **Resultados**

Se identificaron 386 citas, de las cuales 54 fueron potencialmente relevantes en base al título y el resumen (figura 1). Después de la revisión de texto completo, diez artículos fueron incluidos en el análisis, los cuales proporcionaron datos de once estudios diferentes, con una población total de 1732 personas.

### *Características de los estudios*

De los once estudios seleccionados, ocho usaron PCR, uno LCR y, dos Xpert MTB/RIF. Todos los estudios de PCR fueron pruebas locales “in-house”. El tamaño medio de cada estudio fue 248 (especímenes o sujetos), con un rango de 20 a 1000. La Tabla 1 muestra las características principales de estos estudios.

En general, nueve estudios se realizaron en países de ingresos medios donde la carga de TB es alta. Sólo un estudio recogió los datos de manera retrospectiva. Además, se utilizó un estándar de referencia microbiológica en ocho estudios, uno utilizó un patrón de referencia amplio y los otros dos utilizaron diferentes combinaciones de manifestaciones clínicas y cultivos.

Cada uno de los estudios que utilizaron PCR tenía un protocolo distinto, con genes diana y cebadores diferentes. Seis estudios utilizaron la técnica de PCR simple y dos utilizaron una técnica de PCR anidada.

### *Evaluación de calidad*

La Figura 2 muestra los resultados de la evaluación de la calidad con la herramienta QUADAS-2; 64% de los estudios (todos del grupo de PCR) fueron considerados con alto riesgo de sesgo en los dominios de prueba índice y referencia estándar (véase el Apéndice 2 y el cuadro suplementario para obtener información más detallada). Esto se

produjo porque los operadores no fueron cegados a los resultados de la prueba índice antes de realizar la prueba de referencia o viceversa. Además, los protocolos utilizados en las pruebas PCR “in-house” no fueron estandarizados y las variaciones en aspectos técnicos podrían haber introducido sesgos.

También por esta última razón, la aplicabilidad de todos los estudios de PCR fue considerada como de alta preocupación en el dominio de prueba de índice. Además, la aplicabilidad de tres estudios se consideró como de alta preocupación en el dominio de referencia estándar por que se utilizó una referencia estándar distinta a la microbiológica o la amplia .

#### *Medidas de resumen*

La Figura 3 muestra el rendimiento diagnóstico de los estudios distribuidos en tres grupos, PCR, LCR y Xpert MTB/RIF, con un intervalo de confianza (IC) de 95%. No todos los resultados coinciden con las medidas reportadas ya que se utilizó una corrección de continuidad estadística del 5% que afectó a los estudios con mayores IC y valores extremos.

Dentro de los estudios que utilizaron PCR, una heterogeneidad importante pudo ser observada mediante la evaluación del forest plot (figura 2). Debido a la baja calidad de estos estudios, los diferentes protocolos y variantes de PCR utilizados, y la heterogeneidad visual en los forest plot, se decidió no proceder con la estimación de medidas de resumen.

Los dos estudios que utilizaron Xpert MTB/RIF mostraron buena calidad, y la heterogeneidad fue evaluada mediante la prueba de  $\chi^2$ . La Tabla 2 muestra las medidas de resumen de estos estudios. También se trazó una curva SROC (figura 4). La

sensibilidad y especificidad resumen fue 0.87 (0.66-0.96) y 0.91 (0.84-0.95),  
respectivamente.

## **Discusión**

Esta revisión sistemática indica que el rendimiento de las PAANs en muestras de orina para el diagnóstico TBGU produce una sensibilidad y especificidad que van desde 0.83 a 0.95 y 0.79 a 0.99, respectivamente. Si bien estos resultados muestran un rendimiento diagnóstico aceptable, una correcta interpretación de los resultados debe tener en cuenta los amplios intervalos de confianza, las diversas fuentes de heterogeneidad, tales como problemas de calidad, diferentes genes diana y protocolos, y distinto número de muestras analizadas.

Dentro de los diversos protocolos seguidos para las pruebas de PCR, las posibles fuentes de resultados falsos positivos y falsos negativos son la contaminación y la inhibición, respectivamente. La contaminación es causada generalmente por productos de amplificación aerosolizados que interfieren con los análisis posteriores. Esto es especialmente cierto para los métodos de PCR anidada, en la cual se realizan múltiples amplificaciones (27). Si no se controla, en un tiempo relativamente corto, reactivos de laboratorio, equipos, e incluso sistemas de ventilación se contaminan con estos productos (27). Por otro lado, ciertos metabolitos, fármacos y fluidos corporales pueden producir la inhibición del PCR. En la orina, el inhibidor más crítico es la urea, que puede degradar la polimerasa de una manera concentración dependiente (28). De esta manera, cuando se realiza un PCR, se necesita un protocolo controlado que debe seguirse cuidadosamente para evitar inhibidores y contaminantes, mientras que con la prueba Xpert MTB/RIF, uno no se enfrenta a los mismos problemas (29).

La prueba Xpert MTB/RIF tiene la ventaja de hacerse en un cartucho cerrado, automatizado, con mínimos aspectos técnicos manuales (29). En nuestro análisis, las estimaciones combinadas de los estudios de Xpert MTB/RIF en mostraron una sensibilidad y especificidad satisfactorias: 0.87 y 0.91, respectivamente. Sin embargo,

sólo la especificidad tenía un IC estrecho. Clínicamente se puede traducir en que la prueba informa correctamente el 91% de los pacientes sanos como verdaderos negativos. Debido a que este metanálisis sólo comprendía dos estudios, se requiere más investigación para mejorar y obtener estimaciones más precisas. Además, los valores predictivos positivos (VPP) y valor predictivo negativo (VPN) que dependen de la prevalencia de la enfermedad deben ser reportados para diferentes situaciones clínicas, ya que tienen mayor utilidad práctica.

La información de la resistencia a los fármacos en los pacientes con TBGU es escasa (1,2,30,31). En este contexto, ninguno de los estudios Xpert MTB/RIF analizó la resistencia a la rifampicina, posiblemente debido a que se realizaron en países de altos ingresos, donde la tuberculosis multidrogo resistente (MDR) no es una preocupación. Un estudio realizado en China utilizó una prueba diferente, Genotype® MTBDRplus, que evalúa la isoniazida y la resistencia a la rifampicina, en un contexto ambulatorio (32). Se encontró que un tercio de sus pacientes tenían TBGU resistente a algún medicamento, y una cuarta parte tenía TBGU-MDR. Futuros estudios con Xpert MTB/RIF o Genotype® MTBDRplus deben ayudar a caracterizar la epidemiología de la resistencia a los medicamentos, evaluando la utilidad clínica de estas pruebas a diferentes niveles de prevalencia de MDR.

Se identificaron dos problemas importantes en los estudios: la referencia estándar y el número de muestras analizadas no eran igual en todas las investigaciones, y la falta de reporte sobre la variación del rendimiento de las pruebas en pacientes con comorbilidades.

En los estudios que utilizaron el cultivo positivo como patrón de referencia, existe la posibilidad de que casos verdaderamente positivos con solo respuesta clínica después del tratamiento, hayan sido etiquetados como falsos positivos; produciendo una



especificidad menor. Además, algunos estudios analizaron más de un espécimen por paciente. Si bien existe un consenso basado en opinión de expertos para el uso de tres muestras consecutivas de la mañana para cultivos o PAANs, no hay ningún estudio específico que aborde este tema (33). Aumentar el número de especímenes puede incrementar la sensibilidad a costa de la reducción de la especificidad, por otra parte, se desconoce si esta estrategia es costo-efectiva.

Ninguno de los artículos seleccionados informó sobre las comorbilidades de sus pacientes. Estamos especialmente preocupados por los pacientes con VIH, ya que se ha demostrado que el rendimiento de las pruebas de diagnóstico varía en esta población. En los pacientes infectados por el VIH, los que presentan tuberculosis pulmonar suelen tener patología paucibacilar y por ende menor rendimiento de las PAANs (34,35). Sin embargo, en la tuberculosis extrapulmonar y especialmente la meningitis tuberculosa, dos estudios informaron un mayor rendimiento de Xpert MTB/RIF en aquellos pacientes infectados con VIH (36,37). Sugerimos que los autores consideren determinar el estado de infección por VIH en futuras investigaciones en pacientes con TBGU.

#### *Fortalezas y debilidades del estudio*

Se identificaron algunas debilidades en nuestra revisión. No hemos podido evaluar la tuberculosis genital masculina y femenina debido a la falta de referencia estándar clara y por tanto, la posibilidad de generar grandes sesgos en el análisis (20). No se evaluó el efecto de la adición PAANs a otras pruebas, y el efecto neto de PAANs en los valores predictivos positivos o valores predictivos negativos ya que éstos dependen de la prevalencia de la enfermedad. Tampoco pudimos explorar el efecto de cuestiones tales como la experiencia con el equipamiento para realizar PAANs, y la infraestructura de laboratorio debido a las carencias en el reporte de los estudios analizados .

Por otro lado, nuestro estudio tiene varios puntos fuertes: una estrategia de búsqueda exhaustiva (incluyendo resúmenes de congresos, y sin restricciones de idioma), y una profunda evaluación crítica de la calidad, la exploración de la heterogeneidad conforme con las directrices publicadas (15), y el reporte siguiendo las recomendaciones PRISMA para revisiones sistemáticas (16).

### *Limitaciones*

La falta de calidad en el reporte de los estudios de rendimiento diagnóstico interfiere con la evaluación crítica y la replicación de los estudios (38). Sugerimos que los autores sigan las normas “Standards for Reporting Diagnostic Accuracy“ (STARD) con el fin de lograr transparencia, integridad y la excelencia en los reportes (39). Estas normas fueron actualizada en el 2015 (40), y comprenden una lista de 30 elementos esenciales para estudios de rendimiento de pruebas de diagnóstico de alta calidad; también ha incorporado pruebas sobre fuentes de sesgo y la variabilidad en la precisión de diagnóstico (42).

## **Conclusión**

La TBUG es una enfermedad ignorada; el cuadro clínico inespecífico y las pruebas convencionales limitadas son las causantes de este problema. El diagnóstico es difícil y con frecuencia tardío, llevando a un mayor impacto en el tracto urinario.

Todos los estudios de PCR analizados fueron heterogéneos, ya que involucran diferentes genes diana, protocolos y tienen varios problemas de calidad. En consecuencia, no se pudo hacer un meta análisis con este grupo.

Los estudios con Xpert MTB/RIF, en contraste, fueron de alta calidad y no heterogéneos. Nuestro análisis mostró una especificidad favorable. Se necesitan urgentemente más estudios de calidad para mejorar el diagnóstico de la TBUG.

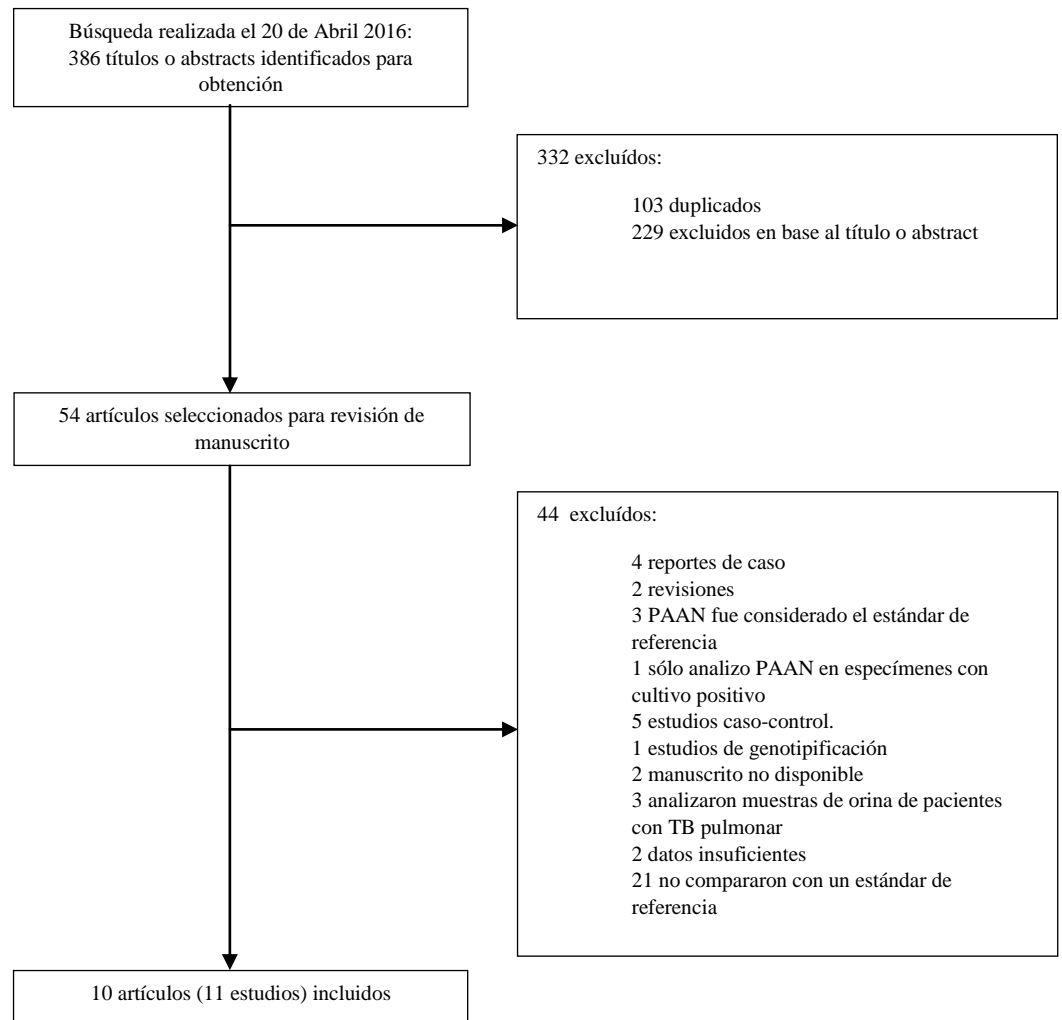
## **Conflictos de interés**

Declaramos que no tenemos conflicto de interés.

## **Agradecimientos**

El equipo quiere agradecer especialmente a Seungseo Choi, un estudiante de medicina de la Facultad de Medicina Alberto Hurtado de la UPCH, por su ayuda en la traducción de los manuscritos en coreano, y a María Rueda, MD por sus importantes sugerencias sobre el manuscrito.

**Figura 1: Selección de estudios que reportan el uso de PAAN en orina para el diagnóstico de tuberculosis urogenital**



**Tabla 1: Descripción de los estudios en la revisión y las medidas de rendimiento reportadas**

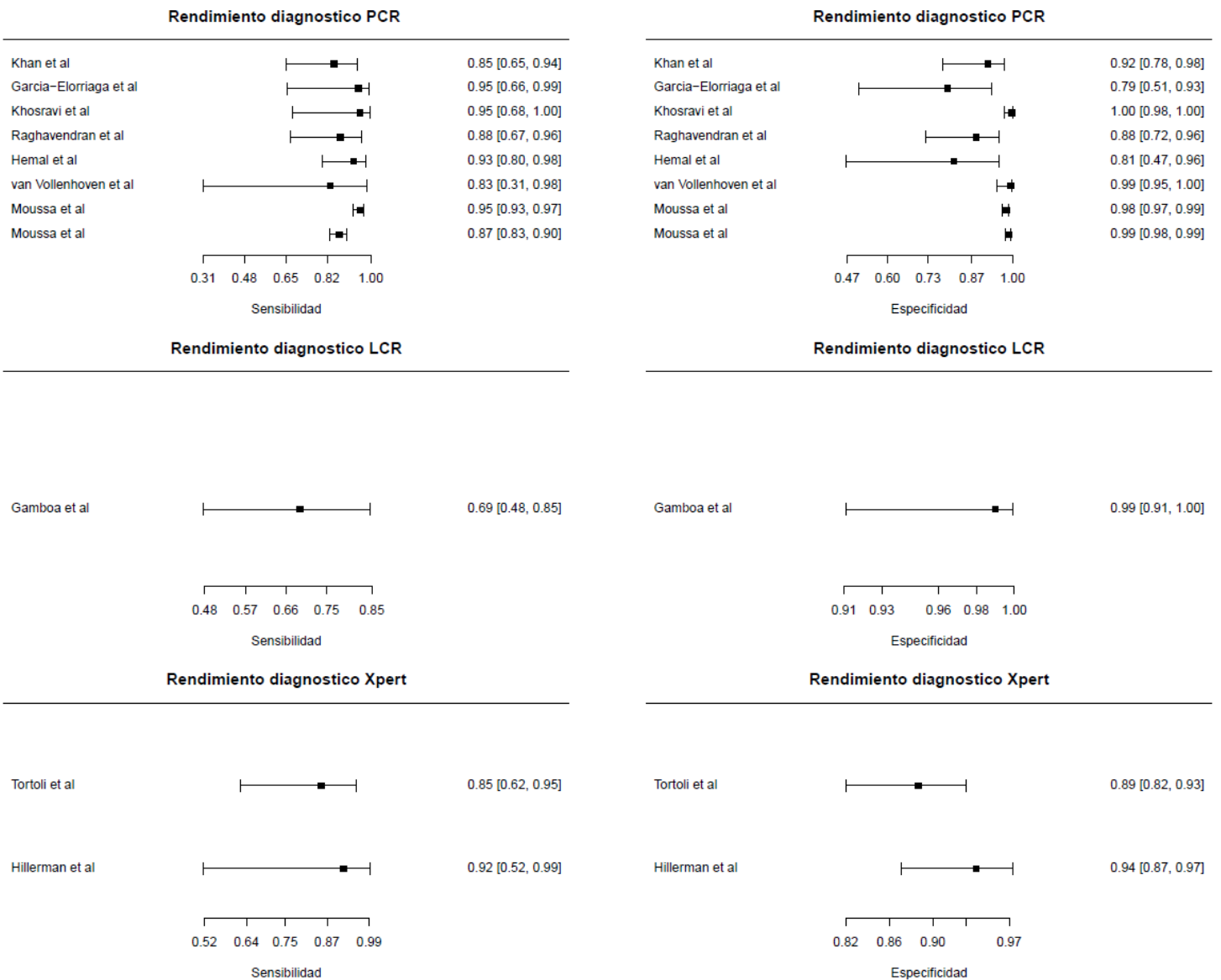
Estudio	Año	País	Recolección Prospectiva	Estudios con cegamiento	Prueba índice utilizada	Detalles específicos de la prueba índice	Número de pacientes	Especímenes por paciente	Referencia estándar	Sensibilidad	Especificidad
Khan et al (41)	2013	Pakistán	Si	No	Local	Real Time PCR IS6110, MPB-64, 16rRNA	50	3 MCU	A	88.6	96.5
García-Elorriaga et al (42)	2009	México	No	No	Local	Nested PCR 32-kDa, MTP40 and IS6110	20	1	*	100	82
Khosravi et al (43)	2010	Irán	Si	No	Local	Nested PCR IS6110	200	1	A	100	100
Raghavendran et al (44)	2016	India	Si	No	Local	PCR (gene target nor reported)	48	1	A	89.5	89.6
Hemal et al (45)	2000	India	Si	No	Local	PCR MPB-64	42	Desconocido	†	94.3	85.7
van Vollenhoven et al (46)	1996	Sudáfrica	Si	No	Local	PCR M13 mp8	82	Desconocido	A	100	100
Moussa et al (47)	2000	Egipto	Si	No	Local	PCR 16S rRNA	1000	3 CMUS	A	87.05	98.9
Moussa et al (47)	2000	Egipto	Si	No	Local	PCR IS6110	1000	3 CMUS	A	95.59	98.11
Gamboa et al (48)	1998	Colombia	Si	No	Comercial	LCx M. Tuberculosis Assay	69	Desconocido	A	70	100
Hillerman et al (49)	2011	Alemania	Si	Si	Comercial	Xpert MTB/RIF	91	1	A	100	98.6
Tortoli et al (50)	2012	Italia	Si	Si	Comercial	Xpert MTB/RIF	130	1	B	87.5	99.1

MCU: Muestra continua de primera orina de la mañana; A: estándar de referencia microbiológico (cultivo positivo); B: Estándar de referencia amplio (cultivo positivo o manifestaciones clínicas con adecuada respuesta al tratamiento después de un mes como mínimo); \*Decisión final del médico tratante considerando el cultivo y la respuesta al tratamiento; †hallazgos radiológicos avanzados y típicos, cultivo o microscopía de orina positivos, y examen histológico de una biopsia o de un espécimen extraído quirúrgicamente.



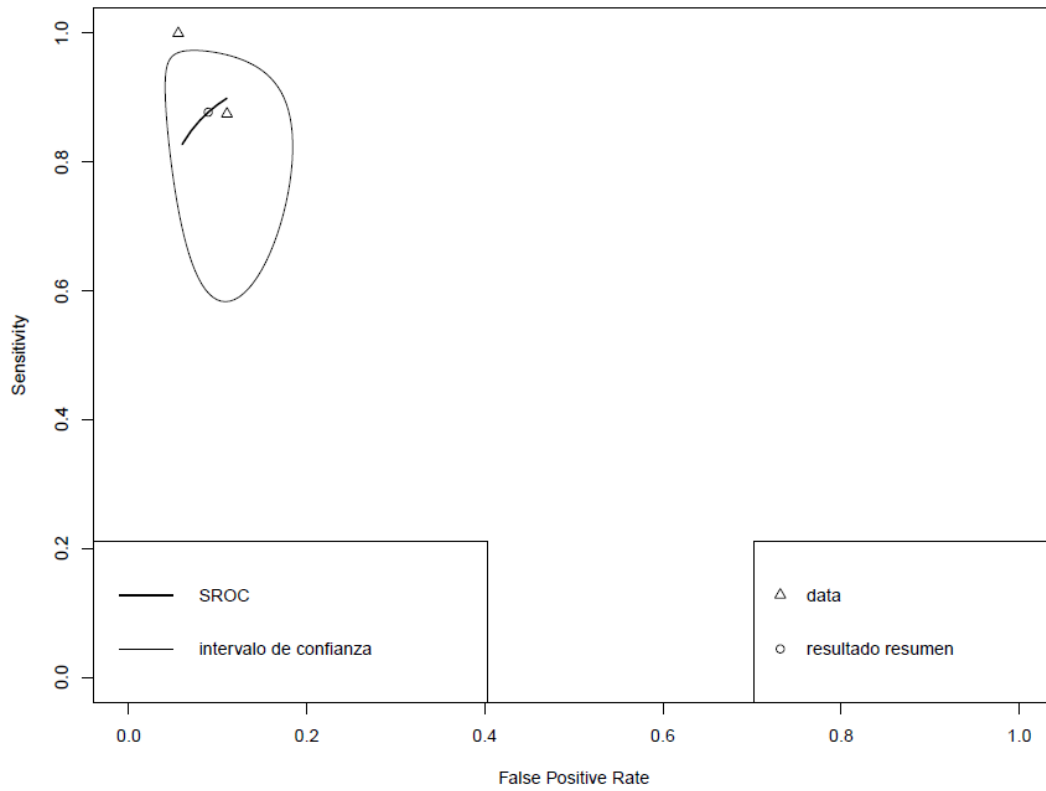
**Figura 2: Resumen de la Herramienta de Evaluación de la Calidad de Estudios de rendimiento diagnóstico (QUADAS-2**

**Figura 3: Forest plot del rendimiento de PCR, LCR y Xpert MTB/RIF en el diagnóstico de TBUG**



Las estimaciones puntuales de sensibilidad y especificidad de cada estudio se muestran en los círculos sólidos. Las barras de error corresponden al intervalo de confianza al 95%. Una corrección de la continuidad estadística de 5% fue aplicada.

Curva SROC para estudios con Xpert



Cada triángulo representa un estudio en el metanálisis, y el círculo es la medida de estimación resumen. La línea delgada denota el intervalo de confianza.

**Figura 4: Curvas receptor-operador resumen (SROC) para la prueba Xpert MTB/RIF.**



**Tabla 2: Mediciones resumen de rendimiento de la prueba Xpert MTB/RIF.**

<b>Propiedad de la prueba</b>	<b>Medida resumen * (95% CI)</b>	<b>Valor p de prueba de heterogeneidad †</b>
<b>Sensibilidad</b>	0.87 (0.66-0.96)	1
<b>Especificidad</b>	0.91 (0.84-0.95)	0.27
<b>Odds ratio diagnóstico</b>	58.16 (15.75- 214.92)	0.43
<b>Likelihood ratio +</b>	9.66(4.12-19.2)	0.31
<b>Likelihood ratio -</b>	0.15(0.04-0.40)	0.36

\*Modelo de efectos aleatorizados. †  $\chi^2$  para heterogeneidad de la prueba

## REFERENCIAS BIBLIOGRAFICAS

- 1 World Health Organization (WHO). Global Tuberculosis Report 2015. Ginebra: WHO; 2015.
- 2 Figueiredo AA, Lucon AM, Junior RF, Srougi M. Epidemiology of urogenital tuberculosis worldwide. *Int J Urol* 2008; 15: 827–32.
- 3 Kulchavenya E. Urogenital tuberculosis: definition and classification. *Ther Adv Infect Dis* 2014; 2: 117–22.
- 4 Figueiredo AA, Lucon AM. Urogenital Tuberculosis: Update and Review of 8961 Cases from the World Literature. *Rev Urol* 2008; 10: 207–17.
- 5 Molina RL, Diouf K, Nour NM. Tuberculosis and the Obstetrician-Gynecologist: A Global Perspective. *Rev Obstet Gynecol* 2013; 6: 174–81.
- 6 Abbara A, Davidson RN. Etiology and management of genitourinary tuberculosis. *Nat Rev Urol* 2011; 8: 678–88.
- 7 Bultitude MF. Campbell-Walsh Urology Tenth Edition. *BJU Int* 2012; 109: E10–E10.
- 8 Kapoor R, Ansari MS, Mandhani A, Gulia A. Clinical presentation and diagnostic approach in cases of genitourinary tuberculosis. *Indian J Urol IJU J Urol Soc India* 2008; 24: 401–5.
- 9 Tostmann A, Kik SV, Kalisvaart NA, et al. Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands. *Clin Infect Dis Off Publ Infect Dis Soc Am* 2008; 47: 1135–42.
- 10 World Health Organization. Policy framework for implementing new tuberculosis diagnostics. Geneva: World Health Organization, 2011.
- 11 Ling DI, Flores LL, Riley LW, Pai M. Commercial nucleic-acid amplification tests for diagnosis of pulmonary tuberculosis in respiratory specimens: meta-analysis and meta-regression. *PloS One* 2008; 3: e1536.
- 12 Adelman MW, Kurbatova E, Wang YF, et al. Cost analysis of a nucleic acid amplification test in the diagnosis of pulmonary tuberculosis at an urban hospital with a high prevalence of TB/HIV. *PloS One* 2014; 9: e100649.
- 13 Kobayashi M, Ray SM, Hanfelt J, Wang YF. Diagnosis of tuberculosis by using a nucleic acid amplification test in an urban population with high HIV prevalence in the United States. *PloS One* 2014; 9: e107552.
- 14 Taegtmeier M, Beeching NJ, Scott J, et al. The clinical impact of nucleic acid amplification tests on the diagnosis and management of tuberculosis in a British hospital. *Thorax* 2008; 63: 317–21.
- 15 Marks SM, Cronin W, Venkatappa T, et al. The health-system benefits and cost-effectiveness of using Mycobacterium tuberculosis direct nucleic acid amplification testing to diagnose tuberculosis disease in the United States. *Clin Infect Dis Off Publ Infect Dis Soc Am* 2013; 57: 532–42.
- 16 Centers for Disease Control and Prevention (CDC). Updated guidelines for the use of nucleic acid amplification tests in the diagnosis of tuberculosis. *MMWR Morb Mortal Wkly Rep* 2009; 58: 7–10.
- 17 Celik C, Gozel MG, Bakici MZ, Berk S, Ozsahin SL, Gulturk E. Applicability of Xpert MTB/RIF assay for routine diagnosis of tuberculosis: a four-year single-center experience. *Turk J Med Sci* 2015; 45: 1329–34.

- 18 Theron G, Peter J, Richardson M, et al. The diagnostic accuracy of the GenoType® Mtbdrsl assay for the detection of resistance to second-line anti-tuberculosis drugs. *Cochrane Database Syst Rev* 2014; 10: 4–123.
- 19 Shah HN, Badlani GH. Genitourinary Tuberculosis; An Update. *Curr Bladder Dysfunct Rep* 2013; 8: 186–96.
- 20 Bose M. Female genital tract tuberculosis: How long will it elude diagnosis? *Indian J Med Res* 2011; 134: 13–4.
- 21 Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks J, Bossuyt P, Gatsonis C, eds. *Cochrane handbook for systematic reviews of diagnostic test accuracy*. Version 0.9.0. London: The Cochrane Collaboration, 2010.
- 22 Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 2009; 6: e1000097.
- 23 Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155: 529–36.
- 24 R Core Team (2016). R Foundation for Statistical Computing: Vienna, Austria, R: A Language and Environment for Statistical Computing. Available at <http://www.R-project.org> (accessed April 17, 2016).
- 25 Doebler P (2012): Meta-Analysis of Diagnostic Accuracy with mada. Available at <https://cran.r-project.org/package=mada> (accessed April 17, 2016).
- 26 Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58: 982–90.
- 27 Aslanzadeh J. Preventing PCR Amplification Carryover Contamination in a Clinical Laboratory. *Ann Clin Lab Sci* 2004; 34: 389–96.
- 28 Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors – occurrence, properties and removal. *J Appl Microbiol* 2012; 113: 1014–26.
- 29 Banada PP, Sivasubramani SK, Blakemore R, et al. Containment of bioaerosol infection risk by the Xpert MTB/RIF assay and its applicability to point-of-care settings. *J Clin Microbiol* 2010; 48: 3551–7.
- 30 Zajackowski T. Genitourinary tuberculosis: historical and basic science review: past and present. *Cent Eur J Urol* 2012; 65: 182–7.
- 31 Merchant S, Bharati A, Merchant N. Tuberculosis of the genitourinary system-Urinary tract tuberculosis: Renal tuberculosis-Part I. *Indian J Radiol Imaging* 2013; 23: 46–63.
- 32 Ye Y, Hu X, Shi Y, et al. Clinical Features and Drug-Resistance Profile of Urinary Tuberculosis in South-Western China: A Cross-sectional Study. *Medicine (Baltimore)* 2016; 95: e3537.
- 33 Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999. *Am J Respir Crit Care Med* 2000; 161: 1376–95.
- 34 Swaminathan S, Padmapriyadarsini C, Narendran G. HIV-Associated Tuberculosis: Clinical Update. *Clin Infect Dis* 2010; 50: 1377–86.

- 35 Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev* 2014; 1: CD009593.
- 36 Patel VB, Theron G, Lenders L, et al. Diagnostic Accuracy of Quantitative PCR (Xpert MTB/RIF) for Tuberculous Meningitis in a High Burden Setting: A Prospective Study. *PLoS Med* 2013; 10. DOI:10.1371/journal.pmed.1001536.
- 37 Nhu NTQ, Heemskerk D, Thu DDA, et al. Evaluation of GeneXpert MTB/RIF for Diagnosis of Tuberculous Meningitis. *J Clin Microbiol* 2014; 52: 226–33.
- 38 Korevaar DA, Enst WA van, Spijker R, Bossuyt PMM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med* 2014; 19: 47–54.
- 39 Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138: W1-12.
- 40 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; 351: h5527.
- 41 Khan F, Cheema F, Khan M. Accuracy of urinary PCR as compared with urine culture for early diagnosis of genitourinary tuberculosis. *Pak J Med Health Sci* 2013;7: 675-678.
- 42 Garcia-Elorriaga G, Gracida-Osorno C, Carrillo-Montes G, Gonzalez-Bonilla C. Clinical usefulness of the nested polymerase chain reaction in the diagnosis of extrapulmonary tuberculosis. *Salud Publica Mex* 2009; 51: 240–5.
- 43 Khosravi AD, Hashemzadeh M, Ghorbani A, Seghatoleslami S. Comparison of Nested-PCR technique and culture method in detection of *Mycobacterium tuberculosis* from patients suspected to genitourinary tuberculosis. *Afr J Biotechnol* 2010; 9: 2151–5.
- 44 Raghavendran M, Venugopal A. MP24-02 PCR IN GENITO URINARY TUBERCULOSIS. *J Urol* 2016; 195: e270–1.
- 45 Hemal AK, Gupta NP, Rajeev TP, Kumar R, Dar L, Seth P. Polymerase chain reaction in clinically suspected genitourinary tuberculosis: comparison with intravenous urography, bladder biopsy, and urine acid fast bacilli culture. *Urology* 2000; 56: 570–4.
- 46 van Vollenhoven P, Heyns CF, de Beer PM, Whitaker P, van Helden PD, Victor T. Polymerase chain reaction in the diagnosis of urinary tract tuberculosis. *Urol Res* 1996; 24: 107–11.
- 47 Moussa OM, Eraky I, El-Far MA, Osman HG, Ghoneim MA. Rapid diagnosis of genitourinary tuberculosis by polymerase chain reaction and non-radioactive DNA hybridization. *J Urol* 2000; 164: 584–8.
- 48 Gamboa F, Dominguez J, Padilla E, et al. Rapid diagnosis of extrapulmonary tuberculosis by ligase chain reaction amplification. *J Clin Microbiol* 1998; 36: 1324–9.
- 49 Hillemann D, Rusch-Gerdes S, Boehme C, Richter E. Rapid molecular detection of extrapulmonary tuberculosis by the automated GeneXpert MTB/RIF system. *J Clin Microbiol* 2011; 49: 1202–5.
- 50 Tortoli E, Russo C, Piersimoni C, et al. Clinical validation of Xpert MTB/RIF for the diagnosis of extrapulmonary tuberculosis. *Eur Respir J* 2012; 40: 442–7.

# APÉNDICE

## APÈNDICE 1

### Estrategia de búsqueda completa

#### PubMed:

("Nucleic Acid Amplification Techniques"[Mesh] OR "Polymerase Chain Reaction"[Mesh] OR "Real-Time Polymerase Chain Reaction"[Mesh] OR "Reverse Transcriptase Polymerase Chain Reaction"[Mesh] OR "Multiplex Polymerase Chain Reaction"[Mesh] OR "genexpert"[tw] OR "genotype"[tw])

AND

((("Tuberculosis, Urogenital"[Mesh] OR "Tuberculosis, Renal"[Mesh]) OR ("Urogenital System"[Mesh] AND "Tuberculosis"[Mesh]) OR "Genitourinary tract tuberculosis"[tw] OR ("Extrapulmonary tuberculosis"[tw] AND "Urine"[tw]))

#### Embase (a través de los comandos de Ovid):

1. exp \*urogenital tuberculosis/
2. polymerase chain reaction/ or exp \*gene amplification/ or exp \*nucleic acid analysis/
3. exp kidney tuberculosis/
4. Urogenital system.mp.
5. exp extrapulmonary tuberculosis/
6. exp urine/
7. 5 and 6
8. exp \*female genital system/
9. Genexpert.mp.
10. (MTBDR\* or "Genotype MTBDR\*").mp.
11. tuberculosis.mp. or Mycobacterium tuberculosis/
12. 4 and 11
13. 8 and 11
14. 2 or 9 or 10
15. 1 or 3 or 7 or 12 or 13
16. 14 and 15

#### Lilacs:

(tw:(tuberculosis urogenital)) OR (tw:(tuberculosis genitourinaria)) AND (tw:(pcr)) OR (tw:(pruebas amplificación nucleica)) OR (tw:(genexpert)) AND (instance:"regional") AND ( db:("LILACS"))

#### Scopus:

TITLE-ABS-KEY (genitourinary tuberculosis ) AND TITLE-ABS-KEY ( Nucleic acid amplification test)

#### Web of Science:

Tema: (Nucleic acid amplification test) AND Tema: (Genitourinary Tuberculosis)

**Cochrane Library:**

Genitourinary Tuberculosis

## APÉNDICE 2

### Reglas de QUADAS-2 e interpretación

#### Dominio 1: Selección de pacientes

**El riesgo de sesgo: ¿Puede la selección de los pacientes haber introducido sesgo?**

**Pregunta de señalización 1: ¿Fue una muestra consecutiva o al azar de pacientes incluida?**

Se marca "sí" si el estudio reclutó una muestra consecutiva o al azar de los pacientes elegibles; "no" si el estudio seleccionó pacientes por conveniencia, e "incierto" si el estudio no informó la forma de selección de los pacientes o no se informó claramente.

**Pregunta de señalización 2: ¿Se evitó un diseño de casos y controles?**

Vamos a marcar "sí" si el estudio incluyó sólo pacientes con sospecha de TBUG; "No" si el estudio se reclutó pacientes con TBUG confirmada o explícitamente señaló un diseño tipo casos y controles; e "incierto" para todos los demás escenarios o si no se informó claramente.

**Pregunta de señalización 3: ¿El estudio evita exclusiones inapropiadas?**

Una exclusión inapropiada podría ocurrir si, después de que el técnico de laboratorio ejecuta las pruebas de índice y de referencia, él o ella no registra los resultados de la prueba en el estudio. Esto podría ocurrir si hubo restricciones de recursos como uno podría encontrar en la práctica, pero no esperamos que esto ocurra en los estudios de investigación incluidos en esta revisión. Vamos a marcar "sí" para todos los estudios, ya que no anticipamos exclusiones inapropiadas.

**Aplicabilidad: ¿Hay preocupación de que la aplicación o la interpretación de la prueba que está siendo evaluada no coincidan con la pregunta de la revisión?**

Vamos a juzgar como preocupación "baja" si los especímenes seleccionados se ajustan a la pregunta de la revisión, que refleja la forma en que la prueba se utiliza en la práctica. Hemos de juzgar como "alta" preocupación si las muestras seleccionadas no representan aquellos para los que se utiliza la prueba en la práctica, como en los individuos que no son sospechosos de tener TBGU. Vamos a juzgar como preocupación "incierto" si no podemos determinar.

#### Dominio 2: Prueba de índice

**El riesgo de sesgo: ¿Podría la realización o la interpretación de la prueba del índice haber introducido un sesgo?**

**Pregunta de señalización 1: ¿Fueron los resultados de las pruebas de índice interpretados sin conocimiento de los resultados de la prueba de referencia?**

Vamos a dar el puntaje de "sí" si el lector del ensayo fue cegado a los resultados de las pruebas de referencia. Vamos a marcar "no" si el lector del ensayo no fue cegado a los resultados de las pruebas de referencia. Si las muestras eran de un banco biológico compuesto por ejemplares con TBGU y la identidad de estas muestras fue conocido por el lector de ensayo, también vamos a responder "no". Vamos a marcar "incierto" si no se mencionó en el documento o si los autores no dieron respuesta a esta pregunta.



**Aplicabilidad: ¿Hay preocupación de que la prueba del índice, su conducta, o su interpretación difieran de la pregunta de la revisión?**

Debido a que nuestra pregunta crítica es amplia, las variaciones en la tecnología de prueba, ejecución o interpretación pueden afectar a las estimaciones de la exactitud diagnóstica de una prueba. Primero vamos a agrupar tipos similares de pruebas y para después analizar la variación entre ellas. Se juzgara estos problemas como motivo de preocupación "baja" si las especificaciones y destino de genes son los mismos. Si no es así, vamos a evaluar estas cuestiones como preocupación "alta".

### **Dominio 3: Referencia estándar**

**El riesgo de sesgo: ¿Podría la referencia estándar, su conducción o su interpretación haber introducido un sesgo?**

**Pregunta de señalización 1: ¿Es probable que la referencia estándar valore adecuadamente la condición diana?**

Los cultivos de *Mycobacterium tuberculosis* son considerados actualmente como el estándar de oro para el diagnóstico de TBGU. Sin embargo, su sensibilidad no es alta en las muestras de orina, muestras de tejido, y el semen, por lo tanto, muchos estudios utilizan una combinación de otros resultados como el estándar de referencia: ya sea un cultivo positivo o una microscopía positiva; o una combinación de características clínicas y resolución de los signos o síntomas después de un tratamiento adecuado. Si el estándar de referencia del estudio se define como cualquiera de los anteriores, vamos a responder "sí"; cuando el patrón de referencia es diferente, vamos a responder "no".

**Pregunta de señalización 2: ¿Fueron los resultados la referencia estándar interpretados sin conocimiento de los resultados de la prueba índice?**

Calificaremos "sí" si el cegamiento se establece explícitamente, o estaba claro que la prueba de referencia se llevó a cabo en un laboratorio independiente, o si fue realizado por diferentes personas, o ambas cosas. Vamos a marcar "no" si el estudio indicó que el resultado de la referencia estándar fue interpretado con conocimiento del resultado de la prueba de índice. Vamos a marcar "incierto" si no se mencionó en el documento o si los autores no dieron respuesta a esta pregunta.

**Aplicabilidad: ¿Hay preocupación de que la condición del estudio según la definición de la norma de referencia no coincide con la pregunta?**

Juzgamos la aplicabilidad a ser de "baja preocupación" para todos los estudios que utilizan la referencia estándar arriba indicado, y de "alta preocupación" a aquellos estudios que han definido sus propias normas de referencia estándar.

### **Dominio 4: Flujo y tiempo**

**El riesgo de sesgo: ¿Podría el flujo de pacientes haber introducido sesgo?**

**Pregunta de señalización 1: ¿Hubo un intervalo apropiado entre la prueba de índice y patrón de referencia?**

Esperamos que la prueba estándar de referencia se lleve a cabo al mismo tiempo que la prueba de índice (es decir, cada uno realizado en una muestra emparejada para la mayoría de los estudios). Sin embargo,

esperamos que algunos estudios incluyan muestras de pacientes que han recibido una prueba de referencia en una muestra anterior. Vamos a responder "sí" si las pruebas fueron emparejados o estaban separados por unos pocos días. Vamos a responder "no" si las pruebas de referencia y de índice no se realizaron en muestras pareadas y se separaron por varios meses. Vamos a marcar "incierto" si no se mencionó en el documento o si los autores no dieron respuesta a esta pregunta.














































































**Pregunta de señalización 2: ¿Todos los pacientes recibieron la misma referencia estándar?**

Vamos a responder "sí" si el mismo patrón de referencia se aplica a todos los pacientes o una muestra aleatoria de los pacientes, "no" si el estándar de referencia solamente se aplicó a un grupo selectivo de los pacientes e "incierto" si se no se indica en el papel o si los autores no pudieron responder a esta pregunta.

**Pregunta de señalización 3: ¿Todos los pacientes fueron incluidos en el análisis?**

Vamos a determinar la respuesta a esta pregunta mediante la comparación del número de participantes enrolados con el número de pacientes incluidos en tablas de dos por dos.


Vamos a marcar "sí" si el número de participantes inscritos se estableció claramente y es igual al número presentado en el análisis o si las exclusiones se describieron adecuadamente. Vamos a marcar "no" si hubo participantes que faltan o excluidos del análisis y no había explicación dada. Vamos a marcar "incierto" si no hay suficiente información para evaluar si los participantes fueron excluidos del análisis.

Estudio	RISGO DE SESGO				PREOCUPACIÓN DE APLICABILIDAD		
	SELECCIÓN DE PACIENTES	PRUEBA A INDICE	ESTANDAR DE REFERENCIA	FLUJO Y TIEMPO	SELECCIÓN DE PACIENTES	INDEX TEST	ESTANDAR DE REFERENCIA
Khan et al							
Garcia-Elorriaga et al							
Khosravi et al							
Raghavendran et al							
Hemal et al							
van Vollenhoven et al							
Moussa et al							
Moussa et al							
Gamboa et al							
Hillerman et al							
Tortoli et al							

**Tabla suplementaria: Resultados detallados de la evaluación QUADAS-2**

 Bajo Riesgo

 Alto Riesgo

 Riesgo no-definido