



Fantomas como parte de sistemas sociotécnicos para el aseguramiento de la calidad de servicios mamográficos: una revisión global

Vila, Cinthia; Ruiz, Micaela; Britos Maximiliano
Sarasua, María
Botti, Horacio
Presentamos nuestro Trabajo Final de Grado (TFG)

Tabla de contenidos

1. Resumen
2. Introducción
3. Materiales y métodos
 - 3.1. Estrategia de búsqueda.
 - 3.2. Selección de estudios.
 - 3.3. Extracción de datos.
 - 3.4. Síntesis y análisis de información.
4. Resultados
 - 4.1. Optimización de la búsqueda
 - 4.2. Selección de estudios
 - 4.3. Características de lo estudios incluidos
5. Discusión.
6. Declaración de conflictos de interés.
7. Referencias bibliográficas.

1. Resumen

La utilización de fantasmas en el control de calidad de imagen es necesaria para evaluar el correcto funcionamiento de los equipos mamográficos y brindar un servicio adecuado. En este campo, las personas, organizaciones, las tecnologías y los procesos se conjugan en un sistema sociotécnico complejo con múltiples niveles interdependientes. Esta revisión global busca responder las siguientes preguntas: ¿Qué representa para los investigadores la calidad de las imágenes de fantasmas mamarios, por qué es importante para ellos y cómo la miden/determinan? ¿Cómo se obtienen, procesan y analizan las imágenes para extraer información útil? ¿Qué aspectos de las investigaciones secundarias y por ende de las primarias precisan ser críticamente evaluadas y eventualmente repensadas? Se utilizaron diversas bases de datos y buscadores bibliográficos. Se empleó la guía Cochrane de revisiones sistemáticas, AMSTAR 2 para evaluar la calidad de las revisiones y PRISMA en la presentación de resultados. Además, utilizamos Zotero, EndNote y Rayyan para gestionar las referencias bibliográficas. De 1301 trabajos identificados, luego de la eliminación de duplicados, la aplicación de filtros y criterios de inclusión y exclusión se obtuvieron dos trabajos que comprenden la evaluación de herramientas basadas en fantasmas físicos y computacionales que pueden aplicarse para evaluar la calidad de las imágenes mamográficas. Una de las revisiones se centra en la comparación de la calidad de las imágenes para la detección de las microcalcificaciones cuando estas son modificadas por ampliación digital u obtenidas mediante magnificación geométrica, incluyendo imágenes de fantasmas mamarios. La otra busca comparar métodos subjetivos (dependientes del juicio de expertos) con métodos automáticos para la evaluación de la calidad de las imágenes de fantasmas mamarios. Aunque ambas revisiones son cuantitativas en diseño, por razones de heterogeneidad, ninguna de las revisiones pudo realizar un metaanálisis, por lo que sus conclusiones son de tipo cualitativo. Las revisiones incluyen trabajos que usan una variedad de fantasmas (3 de 12 fantasmas fueron compartidos), evalúan la calidad de las imágenes de estos a través del estudio de la detección de microcalcificaciones simuladas e índices físicos de calidad (3 de 30 fueron compartidos). Ambas revisiones quedaron clasificadas según AMSTAR 2 como de calidad críticamente baja. No hay acuerdo en que es ni cómo se mide la calidad de imágenes de fantasmas mamarios. Las revisiones que aquí consideramos acuerdan en que la experiencia de los evaluadores humanos y la medición del CNR y SNR son factores relevantes para la evaluación de la calidad de las imágenes de fantasmas mamarios. Se recomienda un enfoque integral para evaluar la calidad de imágenes de fantasmas mamarios, que incluya métodos automáticos, índices físicos, modelos de observadores y evaluadores humanos. Estos métodos automáticos podrían ser parte de una herramienta de apoyo para evaluadores humanos o utilizarse para decisiones automáticas con revisiones periódicas por expertos.

2. Introducción

Según la Organización Mundial de la Salud (OMS), el cáncer de mama es el más frecuente en mujeres ¹. Uruguay presenta estadísticas desfavorables en el contexto mundial, siendo la primera causa de muerte por cáncer en mujeres en nuestro país ². Se ha estimado que 1 de cada 10 personas de sexo femenino tiene posibilidades de presentar cáncer de mama a lo largo de su vida ³. El tratamiento del cáncer de mama puede ser muy eficaz, en particular cuando la enfermedad es detectada a tiempo y para ello es de suma importancia los controles regulares, a través del principal método imagenológico el cual es la mamografía ².

La mamografía es una imagen de rayos X, ya sea de modalidad bidimensional o tridimensional (tomosíntesis) de las partes blandas de la mama, y se puede realizar en forma de tamizaje, para detección y/o diagnóstico de enfermedades mamarias ². El procedimiento en nuestro país para el tamizaje es la realización de dos enfoques, el cráneo-caudal (CC) y el medio lateral oblicuo (MLO) ². Para el diagnóstico se pueden agregar enfoques complementarios como son la magnificación, la compresión focal, imágenes laterales (mediolateral y lateromedial), cráneo-caudal extendida lateralmente o medialmente, cola axilar medio-lateral (mediolateral axillary tale), cráneo-lateral extendido ("Cleopatra"), tomosíntesis (DBT), cleavage, proyección pezón en perfil, cráneo-caudal invertida y proyecciones usando marcadores en piel para localización de lesiones ⁴.

Las microcalcificaciones son pequeños depósitos de calcio como oxalato de calcio o fosfato de calcio, que se pueden depositar en los distintos componentes estructurales de la mama como lo son los lóbulos mamarios, conductos, estroma, acinos mamarios o vasos, cuyo número, tamaño y forma son características importantes⁴. Estas microcalcificaciones mamarias se pueden clasificar en malignas (frecuentemente con un tamaño en el rango de 50 a 500 μm) y/o benignas (que son generalmente mayores a 1 milímetro) ⁵. Así, la identificación y caracterización de microcalcificaciones malignas pequeñas son las actividades del médico imagenólogo mamario que más se benefician del uso de la magnificación ⁶. Esta técnica consiste en un aumento geométrico de la imagen dado por un incremento de la distancia entre la mama y el detector ⁵. Se puede dar una pérdida de la nitidez que acompañe a la magnificación, lo que puede afectar la calidad de imagen, pero esto puede ser compensado con el aumento de la dosis y el uso de un detector de alta resolución ⁴.

De acuerdo con recomendaciones internacionales y nacionales, el logro y mantenimiento de la buena calidad en las imágenes mamográficas así como de la calidad de los equipos y de todos los procesos vinculados al tamizaje del cáncer de mama requiere de programas debidamente acreditados de aseguramiento de la calidad ⁷. Lamentablemente en nuestro país no contamos con tales programas. La calidad de las imágenes debe comprender un adecuado posicionamiento (que

depende de una compleja interacción entre personas y el equipamiento), una adecuada técnica (que incluye adecuada/o: a) compresión, b) técnica digital - exposición, contraste inherente y nitidez-, c) resolución espacial, d) resolución de contraste, e) ruido -de sistema o estructural-, f) dosis, g) calidad de la radiación - combinación ánodo/filtro que se ajuste mejor a cada condición-, h) restricción del movimiento, i) control de los artefactos), un adecuado registro (que incluye: a) identificación del estudio con el nombre completo de la/el paciente y otro identificador, b) fecha del examen, c) vista y lateralidad, d) institución donde se realiza el estudio, e) identificación del licenciado o tecnólogo que realiza el estudio, f) identificación del cassette/screen utilizado y g) identificador de la unidad mamográfica utilizada) y un adecuado etiquetado del mamograma (que incluye datos sobre factores técnicos como combinación ánodo/filtro, kVp, mAs, tiempo de exposición, fuerza de compresión, espesor comprimido de la mama, grado de oblicuidad) ^{7,4}. Estos son algunos de los parámetros que se utilizan para estudiar la calidad de imágenes, los que deben ser determinados o estimados para cada imagen o una muestra representativa de imágenes para hacer una correcta evaluación de la calidad de los servicios mamográficos, debiendo estar dentro de un rango definido, siendo que si estas imágenes están fuera este rango, sería necesario hacer un control de calidad del equipo u otra medida según corresponda ^{7,8}. Además, la evaluación integral de los servicios mamográficos requiere evaluar la calidad de las interpretaciones de las imágenes que realizan los médicos imagenólogos y evaluar la eficacia de los procesos de captación activa y pasiva de pacientes, el grado de satisfacción de los usuarios, la efectividad o costo-efectividad de las actividades miradas integralmente, incluyendo los tiempos entre el llamado y el diagnóstico o la referencia efectiva a un centro de estudios complementario o de mayor nivel ^{7,9,10}. El control de calidad del equipamiento se refiere a realizar pruebas específicas para asegurar el correcto funcionamiento del sistema imagenológico y que éste no se aleje de los valores que se consideran aceptables ⁸. En el caso de las mamografías, el control de calidad se basa en limitar la radiación a la mínima necesaria dentro del rango aceptable, encontrar problemas en el equipo que puedan afectar la calidad de las imágenes ¹¹. El rol de los licenciados en imagenología es de gran importancia, ya que gracias a su conocimiento de la práctica diaria, y sumada una capacitación adecuada, ellos pueden y deben realizar ciertos controles de calidad, permitiendo mantener estándares parciales de funcionamiento en rangos adecuados ¹¹. A través del uso de elementos externos a los necesarios para la adquisición de imágenes de uso clínico se puede establecer un control parcial de la calidad de las imágenes dado por una correcta funcionalidad de los equipos, para ello es necesaria la utilización de los fantasmas, que en general son objetos de prueba estandarizados que simulan partes de la realidad. Los fantasmas mamarios son objetos cuya forma o dimensiones y material simula características del cuerpo humano, así como estructuras patológicas ¹². Estos dispositivos pueden ser artefactos físicos o computacionales creados por el ser humano que se utilizan para facilitar los controles de la calidad técnica de los servicios mamográficos. A través del análisis de las imágenes de los fantasmas se pueden evaluar un conjunto de

tecnologías y capacidades humanas que se utilizan en la obtención, registro, procesamiento, visualización e interpretación de imágenes.¹² Algunos fantasmas contienen objetos que simulan estructuras fibrosas, microcalcificaciones y masas tumorales y sus imágenes pueden ser estudiadas directamente por los licenciados o médicos o pueden analizarse mediante técnicas computacionales de acuerdo con teorías de física de imágenes o por medio de aprendizaje automático⁴. Otros fantasmas están confeccionados para una evaluación únicamente computacional, de acuerdo con teorías de física de imágenes¹³. La inteligencia artificial (IA) es la habilidad de algunos sistemas computarizados, de procesar datos, aprender de ellos y elaborar respuestas en base a ello mediante el empleo de algoritmos (incluidos los de aprendizaje automático) y modelos matemáticos iniciales¹⁴. El aprendizaje automático (también llamado “machine learning”), es la forma en que puede ser utilizada una máquina de computación electrónica para descubrir patrones o clasificar instancias entre los datos (imágenes) suministrados a partir de modelos previos y programas que aprenden modelos⁴. El aprendizaje automático se puede utilizar tanto sobre imágenes clínicas (con múltiples fines posibles) como sobre imágenes de fantasmas (como forma de control de calidad de imagen). En cualquier caso, el avance de esta tecnología hace que puedan surgir herramientas de apoyo a decisiones de los profesionales o potencialmente sustituirlos en la toma de decisiones¹⁴.

El uso de fantasmas, como mencionamos antes, se ha recomendado que forme parte de programas de aseguramiento de calidad integrales¹⁵. Se ha propuesto que el desarrollo tecnológico y su adopción son mejor entendidos como parte sistemas sociotécnicos¹⁶. En el campo de la salud se conjugan diversos procesos como la educación, el trabajo, la investigación, la innovación tecnológica, la producción y el comercio, así como el control social expresado a través de regulaciones, formas de participación y gobernanza y conductas humanas, y presenta dinámicas de carácter global y local que trascienden al campo específico y por ende incluyen múltiples niveles de complejidad sociotécnica¹⁷. El estudio de las condiciones sociotécnicas para el desarrollo local de programas de aseguramiento de la calidad escapa a este trabajo. Este trabajo tiene por objetivo revisar la literatura sistemática (revisión global o revisión de revisiones sistemáticas) en el campo científico tecnológico y social que refiere al uso de los fantasmas para asegurar la calidad de las imágenes mamográficas. Las preguntas principales que se busca responder son las siguientes: ¿Qué representa para los investigadores la calidad de las imágenes de fantasmas mamarios, por qué es importante para ellos y cómo la miden/determinan? ¿Cómo se obtienen, procesan y analizan las imágenes para extraer información útil? ¿Qué aspectos de las investigaciones secundarias y por ende de las primarias precisan ser críticamente evaluadas y eventualmente repensadas?

3. Materiales y métodos

Este estudio sigue los lineamientos de la guía Cochrane para revisión de revisiones¹⁸. De esta forma se intenta abarcar la mayor cantidad de información disponible en el tema para su análisis y que el resultado sea lo más exacto y confiable posible sobre el tópico de investigación.

3.1. Estrategia de búsqueda.

Utilizamos mapas de términos y ecuaciones booleanas de búsqueda¹⁸.

Para intentar abarcar la mayor cantidad de revisiones sistemáticas relacionadas al tema incluimos los buscadores y las bases de datos correspondientes: PubMed¹⁹, BVS²⁰, EBSCOhost²¹, Scopus²², ScienceDirect²³, La Referencia²⁴, Silo²⁴, IOP science²⁵ y IEEE²⁶ accesibles a través de Google Académico²⁷ y del Portal Timbó²⁸. Las búsquedas se realizaron entre el 23/05/2023 y el 14/06/2023. Los términos elegidos para cada mapa (entre paréntesis) se unieron con conectores lógicos OR y entre mapas con conectores AND, por ejemplo:

(phantom OR ...) AND (mammography OR...) AND (quality OR ...) AND
("Systematic Review" OR ...)

El rango temporal de estudio corresponde al periodo 2012-2023 inclusive. Este rango fue elegido debido a que una búsqueda inicial en pubmed nos mostró que antes del año 2011 no hay revisiones sistemáticas con nuestra temática elegida, esto sucede porque la implementación de los primeros mamógrafos digitales se da a partir del 7 de febrero del 2000 cuando la Food and Drug Administration (FDA) aprueba la utilización de GE senographic 2000D²⁹, siendo este el primero (senógrafo). No aplicamos restricciones de idioma en las búsquedas. No hicimos búsqueda de literatura gris.

Los resultados de las búsquedas se transfirieron entre distintas plataformas de trabajo usando archivos de texto con formato Research Information Systems (RIS)³⁰. Los resultados se importaron en Endnote X9³¹ para la eliminación de duplicados. El archivo sin duplicados detectados por EndNote fue importado en Rayyan³² para reforzar la eliminación de duplicados y continuar con el screening y selección de estudios.

3.2. Selección de estudios:

Luego de eliminados los duplicados mencionados anteriormente, 4 revisores de manera independiente seleccionaron las referencias que restaban aún, siguiendo los criterios de inclusión y exclusión que se acordaron, seguido por la lectura de los textos completos de aquellos que no llegaron a tener unanimidad por parte de los revisores. Así mismo los desacuerdos de la elegibilidad mencionada se resolvieron mediante consenso. Los criterios de inclusión determinados fueron los siguientes: (1) Estudios que se realizaron siguiendo los criterios de una revisión sistemática; (2) Estudios que abordan la temática mamografía; (3) Estudios que involucren la

utilización de fantasmas mamarios; (4) Estudios publicados de texto completo. Los criterios de exclusión fueron los siguientes: (1) Estudios que no se realizaron bajo los criterios de una revisión, como son por ejemplo las investigaciones científicas; (2) Estudios que realizaron solo evaluaciones mediante el método de ultrasonido o radiómica (radiomics); (3) Estudios que no contuvieron resumen; (4) Congresos, posters o capítulos de libros.

3.3. Extracción de datos

Para la recuperación de la información de las revisiones seleccionadas se elaboraron planillas de extracción de datos ¹ generando preguntas específicas de respuesta corta a partir de las preguntas de investigación. Los datos extraídos son los siguientes: Tipos de fantasmas, usos de fantasmas, frecuencia de utilización, tipo de receptor del equipo, guías que utilizaron, existencia de protocolo, preguntas que busca responder, el período de revisión, métodos de búsqueda, bases de datos, otra literatura, superposición de referencias, de autores y de citas.

3.4. Síntesis y análisis de información

Dado que los estudios analizados en el proceso de revisión presentaron divergencias en cuanto a la forma en que midieron las métricas de calidad, los protocolos utilizados para obtener las imágenes y los tipos de fantasmas empleados, se encontró una falta de coherencia que impide llevar a cabo un metanálisis, valiosa herramienta en la investigación científica que puede permitir reunir y examinar información cuantitativa (estadística) de diversos estudios, lo que puede permitir llegar a conclusiones más contundentes y confiables sobre un tema en particular ¹⁸. No obstante, nuestro trabajo proporciona un análisis basado en cualidades de los diversos estudios y sintetiza la información recopilada.

Para evaluar la calidad de las revisiones seleccionadas ha sido utilizado la herramienta AMSTAR 2 (A Measurement Tool to Assess systematic Reviews) ³³.

¹ https://docs.google.com/spreadsheets/d/1u5KmkfNRqHA-oVZmYAI0AX653-IQxKF-S1Mtay_ONA/edit#gid=1893212337

4. Resultados

4.1. Optimización de la búsqueda para la identificación de estudios relevantes

Para cada término original se generó un mapa de términos alternativos y estos se combinaron para llegar a una ecuación o método de búsqueda que funcionara mejor en la mayoría de los motores de búsqueda para las bases de datos consultadas. Así, por aproximación sucesiva llegamos a la siguiente ecuación:

Búsqueda 1 = (phantom OR "breast model" OR manikin OR phantoms OR "physical model" OR "Imaging phantom" OR "Imaging phantoms" OR "radiologic phantoms" OR "radiologic phantom" OR "Radiographic Phantom" OR "radiographic phantoms") AND ("review Article" OR "review of literature" or "literature review" or "meta-analysis" or "systematic review" OR "literature synthesis" OR review OR "systematic synthesis") AND (mammography OR mammographic OR "breast imaging" OR mammogram OR "breast CT" OR "breast MRI" OR "breast 3D" OR "breast 2D" OR "magnetic resonance imaging" OR tomosynthesis OR "breast cancer" OR breast) AND ("imaging system performance" OR "image quality" OR "quality control" OR "quality assurance" OR "imaging service quality" OR "quality")

Esta ecuación se utilizó en los siguientes buscadores PubMed, BVS, EBSCOhost, Scopus y IEEE. En los buscadores La Referencia y Silo los términos fueron traducidos al español. En las bases Sciencedirect y IOP science hubo una limitación del número de términos y frases conectables, por lo que fue necesario simplificar la ecuación a la siguiente forma:

(mammography OR "breast imaging") AND (phantom OR "breast model" OR manikin OR phantoms OR "physical model" OR "physical phantom") AND ("systematic review")

4.2. Selección de estudios

La búsqueda permitió identificar 1301 resultados. Se eliminaron 352 publicaciones mediante detección de duplicados en EndNote, quedando en 949 ítems. EndNote detecta principalmente duplicados exactos, por lo tanto, recurrimos a Rayyan, con el que realizamos una eliminación manual de duplicados. Así se eliminaron 68 duplicados adicionales, quedando 881 ítems en nuestra base de datos. 50 referencias en Rayyan fueron eliminadas por no tener resúmenes o no corresponder con trabajos completos (listas de resúmenes, congresos, etc), llegando a 831 ítems. Aplicando en títulos, resúmenes y palabras claves, los términos "this review", "systematic", "systematic review", "review", "meta-analysis", "we review", "synthesis" y "were reviewed" identificamos 433 referencias. Desde este punto, 4 revisores (CV, MS, MR y MB) trabajaron independientemente en Rayyan en modo ciego. Elegimos

una serie de términos para una posible exclusión: “ultrasound”, “radiotherapy”, “radiomics”, “animal”, “pediatric”, “in vitro”, “animals”, “ultrasonographic”, “cadaver” y “rodents”. Las principales razones de exclusión fueron “wrong population” (población equivocada), “wrong technique”(técnica equivocada), “wrong theme”(tema equivocado), “wrong intervention”(intervención equivocada) y “wrong publication type”(tipo de publicación equivocada). De esta manera se excluyeron 396 ítems y se llegó a 37 publicaciones. Mediante la lectura de texto completo, tres investigadores (MS, MB y CV) por consenso resolvieron excluir 35 ítems y fueron seleccionados 2 revisiones sistemáticas para esta revisión global. *Ver figura 1*

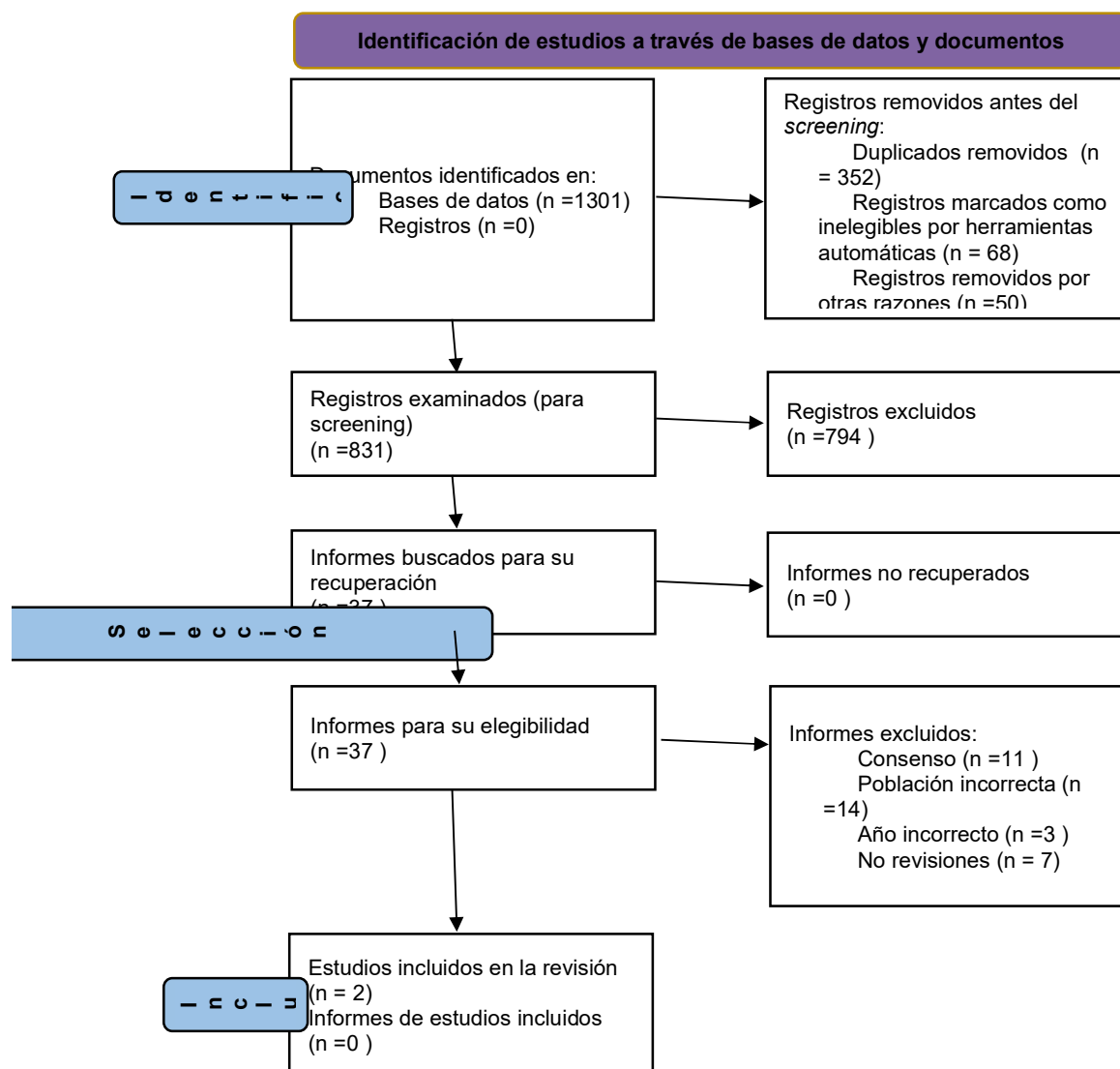


Figura 1. Diagrama de flujo de los resultados de los procesos de identificación, selección e inclusión de trabajos de revisiones sistemáticas. Seguimos la propuesta de la guía PRISMA de presentación de resultados de revisiones.

4.3. Características de los estudios incluidos

Siguiendo la guía Cochrane para revisiones de revisiones ¹⁸, consideramos primero las características generales de los reportes.

Tabla I. Población, intervención, comparación y desenlaces

	Øynes, et. al. ⁵	Alawaji et. al. ³⁴
Población	Imágenes digitales 2D diagnósticas e imágenes de fantomas físicos y digitales	Imágenes 2D digitales y analógicas de fantomas físicos y digitales hasta julio 2020
Intervención	Ampliación digital realizado por el médico que analiza la imagen	Análisis computarizado de la calidad las imágenes
Comparación	Magnificación geométrica realizada por el licenciado y analizada por el médico con o sin ampliación digital	Evaluación de la calidad por observadores humanos y análisis computarizado
Desenlace (outcome) (resultados medidos)	Detección de microcalcificaciones por observadores humanos	Eficiencia y reproducibilidad de la evaluación de la calidad de las imágenes

En la tabla I se describe los componentes PICO (población, intervención, comparación y desenlace- del inglés outcome-)

Tabla II. Fantomas incluidos en las revisiones.

Fantomas incluidos	Øynes, et. al.	Alawaji et. al.
ACR (156) ³⁵	Si	Si
MASTOS model ³⁶	Si	No
CDMAM ³⁷	Si	Si
PMMA ³⁸	Si	Si
TORMAS ³⁹	No	Si
TORMAM ⁴⁰	No	Si

TORMAX ⁴¹	No	Si
CIRS 011A ⁴²	No	Si
CIRS SP01 (es un versión del 011A)	No	Si
MTM 100/R ⁴³	No	Si
RANCON ⁴⁴	No	Si
Contrast-detail ⁴⁵	No	Si

En la tabla II se resumen los fantasmas utilizados por las revisiones seleccionadas, donde hay 12 fantasmas, 1 fantasmas utilizado en la revisión de Øynes et al. 8 fantasmas que solo son nombrados en la revisión de Alawaji et. al. y 3 fantasmas utilizados en ambas revisiones.

Tabla III AMSTAR 2 ³³.

		ART1	ART2
1	¿Las preguntas de investigación y los criterios de inclusión para la revisión incluyeron los componentes de PICO?	Si	Si
2	¿Contenía el informe de la revisión una declaración explícita de que los métodos de revisión se establecieron antes de realizar la revisión y justificaba el informe alguna desviación significativa del protocolo?	No	No
3	¿Explicaron los autores de la revisión su selección de los diseños de estudio para su inclusión en la revisión?	No	No
4	¿Utilizaron los autores de la revisión una estrategia integral de búsqueda de literatura?	Parcial Si	Parcial Si
5	¿Los autores de la revisión realizaron la selección de estudios por duplicado?	Si	Si
6	¿Los autores de la revisión realizaron la extracción de datos por duplicado?	Si	Si

7	¿Los autores de la revisión proporcionaron una lista de estudios excluidos y justificaron las exclusiones?	No	No
8	¿Los autores de la revisión describieron los estudios incluidos con suficiente detalle?	Parcial Si	Parcial Si
9	¿Utilizaron los autores de la revisión una técnica satisfactoria para evaluar el riesgo de sesgo (RoB) en los estudios individuales que se incluyeron en la revisión?	No	No
10	¿Los autores de la revisión informaron sobre las fuentes de financiación de los estudios incluidos en la revisión?	No	No
11	Si se realizó un metanálisis, ¿los autores de la revisión utilizaron métodos apropiados para la combinación estadística de resultados?	No M*	No M*
12	Si se realizó un metanálisis, ¿evaluaron los autores de la revisión el impacto potencial de RoB en estudios individuales sobre los resultados del metanálisis u otra síntesis de evidencia?	No M*	No M*
13	¿Los autores de la revisión tuvieron en cuenta RoB en estudios individuales al interpretar/discutir los resultados de la revisión?	Si	No
14	¿Los autores de la revisión proporcionaron una explicación satisfactoria y una discusión sobre cualquier heterogeneidad observada en los resultados de la revisión?	Si	No
15	Si realizaron una síntesis cuantitativa, ¿los autores de la revisión llevaron a cabo una investigación adecuada del sesgo de publicación (sesgo de estudio pequeño) y discutieron su probable impacto en los	No	No

	resultados de la revisión?		
16	¿Informaron los autores de la revisión alguna fuente potencial de conflicto de intereses, incluida la financiación que recibieron para realizar la revisión?	Si	Si

*No M= No Metaanálisis

En la revisión sistemática de Øynes et al. se reportó el uso de los fantasmas físicos y computacionales, que fueron utilizados en estudios de la detectabilidad de microcalcificaciones utilizando ampliación digital y/o magnificación geométrica en contexto de investigación sin aclarar ningún otro tipo de utilización ⁵. En esta revisión se consideraron exclusivamente estudios que incluyeron imágenes generadas por equipos de mamografía digital. En cuanto a los métodos de revisión, los autores reportaron haber realizado un protocolo de investigación previo que siguió las guías Cochrane y “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (guía PRISMA, del año 2009)^{5,46}. Los revisores realizaron una búsqueda bibliográfica “comprehensiva”, es decir, utilizaron diferentes buscadores y bases de datos para poder identificar, seleccionar y evaluar artículos potencialmente relevantes. No aplicaron ningún rango temporal así como tampoco hubo restricción de idioma. Utilizaron combinaciones de términos de vocabulario controlado y texto libre. Dos revisores de manera independiente realizaron la búsqueda, selección y extracción de información. En cuanto al riesgo de sesgo, para los estudios experimentales realizados en fantasmas solo nombran la utilización de una versión adaptada de Quadas 2 ⁴⁷. En relación con lo anterior, quizás lo más importante es que omitieron el dominio de “estándar de referencia”, es decir los autores de la revisión no reconocen en forma explícita ningún método de análisis de las imágenes como estándar de referencia, sin embargo, la comparación incluye la performance de los observadores humanos ⁵. No se llevó a cabo un metaanálisis sobre la detección de microcalcificaciones porque los estudios originales no compartían medidas de calidad y/o detección comparables entre sí. En esta revisión no se estudió la declaración de conflictos de intereses o datos de financiación de los artículos originales que se incluyeron ni explicación de heterogeneidad en los trabajos originales ⁵. Las personas involucradas en la revisión declararon no tener conflictos de interés, ni relación alguna con las compañías generadoras de productos o servicios que pueden estar implicados ^{5 2}.

² https://docs.google.com/spreadsheets/d/1u5KmkfNRqHA-oVZmYAI0AX653-IQxKF-S1Mtay_ONA/edit#gid=1893212337

Tabla IV. Bases de datos y buscadores utilizados en ambas revisiones

Bases de datos	Øynes, et. al. ⁵	Alawaji et. al. ³⁴
MEDLINE	Si	Si
EMBASE	Si	SI
CINAHL	Si	Si
Engineering Village	Si	No
Web of Science	Si	Si
Scopus	No	Si
Otra literatura		
Resúmenes de congresos	Si	No
Literatura gris	Si	Si

En su revisión sistemática, Øynes et al. además de comparar y resumir el desempeño de la magnificación geométrica y de amplificación digital en mamografías en las detección de microcalcificaciones simuladas usando fantasmas, incluyen estudios diagnósticos sobre imágenes mamográficas que están fuera del foco de esta revisión ⁵. De nuestro interés, el trabajo realizado reúne información retrospectiva sobre estudios experimentales de fantasmas mamarios físicos y de un fantoma computacional de Montecarlo, para evaluar la detección de las microcalcificaciones simuladas por evaluadores humanos comparando la ampliación digital versus magnificación geométrica, que es considerada el estándar de referencia ⁵. Los autores proponen que en ciertas condiciones en estudios de detección con fantasmas la ampliación digital es superior a la magnificación geométrica (Tabla V) ⁵. Las medidas de calidad para cada uno de los estudios incluidos en el trabajo son diferentes (Tabla VI). En la tabla V podemos observar de las revisiones que utilizan fantasmas, los distintos índices que obtienen los revisores. En ella podemos ver que no hay concordancia en la utilización de estos índices debido a la variabilidad de posibilidades de medición produciendo así una incapacidad de generar un correcto metaanálisis. Una de las revisiones utiliza Normalized performance index (PI), otra utiliza Image Quality Figure (IQF), dos revisiones utilizan Contrast-Detail Detection (CDD), la siguiente utiliza Contrast to Noise Ratio (CNR), otra Correct Observation Ratio (COR) y otra utiliza “RANK SUM score” ⁵. Estos índices a su vez pueden ser obtenidos de manera automática por el equipo o pueden ser calculados por los operadores. Dado que los trabajos originales no utilizan los mismos índices de calidad, los autores decidieron realizar un análisis cualitativo. Según Øynes et al. pequeñas microcalcificaciones (diámetro de 50-100 o < 200 μm) y la combinación ánodo-filtro de igual material Mo/Mo o Rh/Rh favorecen

el rendimiento de la magnificación geométrica en relación a la ampliación digital (Tabla V) ⁵. Al contrario, la combinación de ánodo y filtros de materiales distintos Molibdeno y Rhodio producen un mayor rendimiento de la ampliación digital con respecto a la magnificación geométrica (Tabla V), pero no se brindan explicaciones para ello ⁵, lo que podría suponer una fragilidad teórica para las conclusiones ⁵. Además, los autores incluyen un trabajo que muestra evidencias que equiparan en FFDM la magnificación geométrica con detectores convencionales y la detección usando detectores nuevos de conteo de fotones sin magnificación ⁵. Es importante notar, que dicho trabajo usó fantasmas con microcalcificaciones simuladas de 200 μm ⁵, por ende, no hay evidencias sobre el rendimiento de los nuevos detectores en pruebas con microcalcificaciones pequeñas (ej. 50-100 μm).

A partir de las evidencias que hemos mencionado, los autores concluyen (el subrayado es nuestro):

"In conclusion, zoom may be equivalent to magnification in many cases given that optimised procedures and newer detector technologies are now available. This finding has the potential to reduce AGD and improve examination workflow. Both diagnostic test studies and phantom studies using newer detectors would contribute additional knowledge on this topic."

Según los autores, las ventajas de la ampliación digital (igual rendimiento de detección de microcalcificaciones simuladas y menor exposición a radiaciones ionizantes) dependen de la posibilidad conjunta de optimizar equipamientos y de incorporar nuevas tecnologías de detectores. Sin embargo, reflexionamos que es criticable la conclusión en tres aspectos. En primer lugar, no encontramos en la revisión evidencia sobre el rendimiento del uso de detectores nuevos comparados con la ampliación geométrica si las microcalcificaciones simuladas en fantasmas físicos son pequeñas, por lo que debemos asumir que la ampliación geométrica continuaría siendo el estándar de referencia en la detección de microcalcificaciones malignas. En segundo lugar, las inequidades en la atención a la salud dependientes de las dificultades de acceso a las nuevas tecnologías de conteo de fotones, que no están disponibles en la mayoría de los países de medianos y bajos ingresos, ni en Uruguay, deben hacer dudar seriamente del potencial valor social del cambio hacia los nuevos detectores en combinación o no con la aplicación de la ampliación digital o geométrica ⁴⁸. La tercera crítica que destacamos sobre la conclusión arriba citada entre comillas y subrayada de que el resultado del estudio brinda evidencia que sugiere la equivalencia en el rendimiento entre la ampliación digital y la magnificación geométrica. El uso del término equivalencia no puede ser considerado en el contexto del trabajo un aspecto menor, ya que desde el punto de vista científico se haría referencia a evidencias obtenidas a partir de estudios con los diseños y métodos apropiados para estudiar equivalencias. Notamos que según plantean los revisores, los artículos originales incluidos (los trabajos primarios) utilizaron pruebas estadísticas de superioridad, es decir, las pruebas tradicionales

que como H0 proponen la ausencia de diferencias o igualdad y como HA proponen la superioridad de la tecnología nueva sobre la anterior ⁴⁹. Bajo este paradigma, si se rechaza la H0 se evidencia la superioridad de la nueva tecnología y cuando no se puede rechazar la H0, se debe concluir que no hay evidencia que respalde la superioridad, pero, no se puede asumir que las opciones son equivalentes ⁴⁹. Los estudios de equivalencia buscan generar evidencia de que una nueva tecnología no es ni mejor ni peor en rendimiento que una tecnología anterior, y los estudios de no inferioridad, que la nueva tecnología no tiene un peor rendimiento que la anterior ⁴⁹. Así, los estudios de equivalencia son el paradigma apropiado para los trabajos primarios y para el metaanálisis si las conclusiones que se quieren sacar son de equivalencia. Si el uso del término “equivalente” fuera coloquial y no científico, estaría fuera de lugar y podría conducir a errores de interpretación. Si el uso fuera científico, al reportar el hallazgo de una equivalencia entre la ampliación digital y la magnificación geométrica, Øynes et al. incurren en un error metodológico importante. A nuestro entender, la única conclusión que no constituye una falacia inferencial, que se funde en los trabajos incluidos en la revisión de Øynes et al. es que ningún trabajo previo mostró evidencia sobre la equivalencia de la ampliación digital y la magnificación geométrica, ya que los estudios primarios considerados por los autores tenían diseños de superioridad y que en el futuro deberían realizarse más estudios de superioridad y principalmente estudios de equivalencia y/o no inferioridad, para comparar la magnificación geométrica con tanto la ampliación digital como el uso de nuevos detectores con contadores de fotones.

Tabla V. Comparación cualitativa del rendimiento de la ampliación digital y magnificación geométrica en estudios de detección con fantasmas según Øynes et al.⁵

	Magnificación geométrica superior	Ampliación digital superior
MicroS* entre 50–100 μm (no importa la combinación ánodo-filtro)	Si	No
Ánodo- filtro Mo/Rh (MicroS 160–500 μm)	No	Si
Ánodo-filtro Mo/Mo (MicroS 160–500 μm)	Si	No
Ánodo-filtro Rh/Rh (MicroS 160–500 μm)	Si	No

*MicroS= Microcalcificaciones simuladas

Tabla VI- Índices de calidad de imagen utilizados

	Alkhalifah et al.	Egan et al.	Vahey et al.	Koutalonis et al.	Hermann et al.
PI (Normalised performance index* ¹)	No	Si ^{\$}	No	No	No
IQF (image quality figure* ²)	No	No	Si [#]	No	No
CDD (contrast-detail detection* ³)	No	No	Si [#]	No	Si [#]
CNR (contrast-noise ratio* ⁴)	No	No	No	Si ^{\$}	No
COR (Correct observation ratio* ⁵)	No	No	No	No	Si [#]
RANK SUM SCORE* ⁶	Si ^{\$}	No	No	No	No

^{\$}Análisis Computacional

[#]Observación humana

*¹ En castellano Índice de Rendimiento Normalizado. Que se calcula como el múltiplo de la razón de contraste-ruído y el número de mediciones dividido entre la dosis glandular promedio: $PI = (CNR \cdot n) / AGD$

*² Figura de calidad de imagen. Que se calcula como la sumatoria para cada columna de objetos (discos de oro) en el fantoma CDMAM del múltiplo del espesor del disco i y el diámetro del disco i min que es el de menor diámetro detectado por el observador (y en algunos trabajos, modelos de observador): $IQF = \sum_{i=1}^{i=16} C_i \cdot D_i^{min}$

*³ Detectabilidad de detalle de contraste

*⁴ Relación contraste-ruído

*⁵ Razón de observación correcta

*⁶ Puntuación de suma de rango. Se utiliza el test de Kruskal–Wallis, que es un método no paramétrico útil en la investigación para comparar grupos y determinar si hay diferencias significativas entre ellos sin suponer que los datos se distribuyen de una manera específica.

La revisión sistemática de Alawaji et al. reporta que los fantasmas fueron utilizados para evaluar, desarrollar y realizar controles de calidad, incluyendo la investigación y cuantificación de la dosis glandular media ³⁴. Esta revisión utilizó vocabulario controlado así como texto libre en la búsqueda sistemática de artículos en inglés que se complementó con búsquedas manuales para agregar artículos relevantes que no fueron identificados desde un inicio; se identificaron trabajos desde el comienzo de los registros en las bases de datos consultadas (ver Tabla IV) hasta julio del 2020 inclusive, así los receptores utilizados en los trabajos originales incluidos son tanto analógicos (trabajos publicados entre 1995 y 2012) como digitales (2008-2019). Los autores hicieron el reporte de sus resultados de búsqueda, identificación, filtrado e inclusión de artículos, siguieron la guía Preferred

Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), del año 2009 ^{34,46}. En esta revisión no se estudia la declaración de conflictos de intereses o datos de financiación de los artículos originales que se incluyeron y no hay una evaluación del riesgo de sesgo, ni explicación de heterogeneidad en los trabajos originales ³⁴. Las personas involucradas en la revisión declaran no tener conflicto de interés, ni relación alguna con las compañías generadoras de productos o servicios que pueden estar nombrados ³⁴.³

Los autores de esta revisión relevaron la variedad de fantomas, índices de medición de calidad (ver tablas VII y VIII) y programas de análisis de imágenes que se emplearon en los estudios incluídos, variedad que contribuye a una heterogeneidad de resultados que no permite hacer una metaanálisis, y limita a los autores a la realización de un análisis cualitativo ³⁴.

Tabla VII -Métodos de estudio de la calidad empleados en los estudios de mamografía digital en trabajos revisados por Alawaji et al.

Estudio	Fantoma	Métodos
Sundell et al., 2019 ⁵⁰	ACR (model-156)	ACR guideline scoring
Gennaro et al., 2017 ⁵¹	Leeds TORMAS	15 IQIs
Nelson et al., 2016 ⁵²	ACR (model-156) / 3D Anthropomorphic	ACR guideline scoring, line pair resolution bars, and noise
Kamiya et al., 2015 ⁵³	ACR (model-156)	Mean pixel value, contrast, and CNR
Barufaldi et al., 2015 ⁵⁴	ACR (model-156)	Mean pixel value, contrast, and CNR
Donini et al., 2014 ⁵⁵	Unspecified	MTF, NPS, DQE, Bad Pixel, Uniformity, Dark analysis, Lag, AGD, CNR, and Response curve
Perez-Ponce et al., 2011 ⁵⁶	CDMAM	IQF, contrast, and contrast-details scoring
Mayo et al.,2010 ⁴⁴	RACON (developed by authors) / CIRS SP01	Absolute contrast (Ca), relative contrast (Cr %), SNR, size for microcalcification zone, horizontal and vertical resolution as lp/mm, IQF, and uniformity test

³ https://docs.google.com/spreadsheets/d/1u5KmkfNRqHA-oVZmYAI0AX653-IQxKF-S1Mtay_ONA/edit#gid=1893212337

Yip, Chukwu et al., 2009 ³⁷	CDMAM	Contrast, SNR, MTF, and contrast-detail scoring
Prieto et al.,2009 ⁵⁷	CDMAM	Contrast-detail scoring
Mayo et al.,2008 ⁵⁸	CIRS (SP01) / CDMAM	IQI (microcalcification group and resolution pattern bar scoring) for CIRS SP01, and IQF for CDMAM 3.4
Gennaro et al., 2007 ⁵⁹	Leeds TORMAS	CNR, relative contrast, objects count

Tabla VIII - Métodos de estudio de la calidad empleados en los estudios de mamografía analógica en trabajos revisados por Alawaji et al.

Estudio	Fantoma	Parámetros medidos
Asahara and Kodera,2012 ⁶⁰	ACR (model-156)	ACR guideline scoring
Álvarez et al.,2012 ⁶¹	ACR (model-156)	ACR guideline scoring, AGD
Lee et al.,2010 ⁶²	ACR (model-156)	ACR guideline scoring
Hamza et al.,2009 ⁶³	ACR (model-156)	ACR guideline scoring
Patrocínio et al., 2009 ⁶⁴	CIRS 011	SNR, pixel intensity variation, contrast, ESAK
Adel et al.,2005 ⁴³	MTM 100/R	Scoring details
Mayo et al.,2005 ⁶⁵	CIRS (model 11A, SP01 version)	IQI: score the microcalcification, and line pairs per millimeter with score horizontal and vertical lines
Mayo et al.,2004 ⁶⁶	CIRS (model 11A, SP01 version)	SNR, contrast, No. of microcalcification, Line pairs per millimeter
Blot et al.,2002 ⁶⁷	Leeds TOR(MAM)	Scoring details
Jansen and Zoetelief, 2000 ⁶⁸	Contrast-detail phantom	Scoring details
Dougherty, 1998 ⁶⁹	ACR (model-156)	Contrast, SNR
Castellano Smith et al.,1998 ⁷⁰	TOR(MAX)	Scoring details
Brooks et al.,1997 ¹⁵	ACR (model-156)	ACR guideline scoring
Chakraborty and Eckert,1995 ⁷¹	ACR (model-156)	ACR guideline scoring

Como se observa en las Tablas VII y VIII los métodos más utilizados fueron A) la guía de puntaje de la ACR para evaluadores humanos (7 estudios); B) contraste (5 estudios); C) CNR (4 estudios); D) IQF, SNR y “contrast-details scoring”, fueron usados en dos (3 estudios). Por otro lado fueron utilizados en al menos 2 estudios el

MTF, uniformity, AGD, relative contrast, y el IQI; el resto de los parámetros no se repitieron entre los estudios.

En esta revisión sistemática los autores hacen una comparación cualitativa entre métodos computarizados de análisis y métodos basados en observadores humanos (llamados métodos subjetivos) en la evaluación de la calidad de imagen, para detectar desviaciones y problemas subyacentes que pueden tener un impacto en la calidad de imagen. Sin embargo, también se hace una comparación cualitativa entre métodos computarizados, varios de los cuales fueron de carácter no automático y requirieron etapas de preprocesamiento ³⁴. De hecho los autores reportan que el método de “template matching”⁴ (que se podría traducir como comparación o pareo de plantillas) fue el más aplicado en los 26 estudios incluidos, hecho que en opinión de Alawaji et al. se debió a que fue el método más exacto en el reconocimiento automático de patrones con alto contraste ³⁴. Aunque este algoritmo permite detectar tanto patrones de alto como de bajo contraste y los resultados en fantasmas físicos muestran una buena correlación entre observadores humanos y este método en la detección de microcalcificaciones simuladas, así como objetos de bajo contraste como fibras y masas, sin embargo, en uno de 5 estudios el método tuvo un rendimiento notoriamente inferior para el caso de las masas y fibras ³⁴. Los autores notan que en el estudio de estos tipos de patrones de bajo contraste los trabajos incluidos mostraron resultados variables, variaciones que se asociaron a una mayor exploración en cuanto a técnicas de análisis y que podrían estar influidas por el uso de equipamientos de diferente marca y el empleo de diferentes técnicas de adquisición, lo que en conjunto evidenciaría una dificultad mayor en dichas condiciones para aplicar el método de template matching, lo que está de acuerdo con lo reportado previamente por dos artículos revisados por los autores y también generan incertidumbre y una mayor dificultad para establecer recomendaciones precisas sobre qué método computarizado a emplear sea automatizado o no ³⁴. Sin embargo, el procesamiento preparativo de imágenes usando la transformada discreta de ondículas (del inglés *Discrete Wavelet Transform*, de aquí en mas DWT) permitió mejorar el acuerdo entre el método de pareo de plantillas y la detección basada en observaciones humanas en condiciones de bajo contraste y en particular un trabajo revisado mostró que en el caso de masas y microcalcificaciones simuladas, el método computacional fue sutilmente más sensible que el método subjetivo ³⁴.

Los autores afirman que los estudios analizados permiten concluir que los métodos de control de calidad de imágenes basados en análisis computarizados son factibles, confiables (“reliable”), de bajo costo y más eficientes y reproducibles en comparación con los evaluadores humanos ³⁴. Según los autores son ampliamente conocidas las desventajas de los sistemas subjetivos basados en observadores humanos, ya que para los mismos se han demostrado bajas confiabilidades inter e intra observadores y les es difícil la consistencia (es decir la falta de contradicciones lógicas) debido a la influencia de factores externos a la tarea en sí como: la fatiga, la

luz ambiental, el nivel de experiencia y las interpretaciones que realizan de las instrucciones ³⁴. De hecho, los autores reconocen que, para la tarea específica, los observadores humanos solo requieren mínimas señales aún en entornos ruidosos. Sin embargo, según proponen los autores, dichas inconsistencias se manifiestan a la hora de evaluar la calidad del sistema mamográfico haciendo que sea más difícil la detección de fallos en la calidad de imagen y en el desempeño del sistema ³⁴. Los autores proponen que medidas cuantitativas adicionales como el CNR y SNR son útiles para evaluar directamente la calidad de las imágenes y para evaluar la visibilidad de las señales ³⁴. Dado que el SNR se determina sobre la región de interés (ROI), que depende de la forma y tamaño del patrón derivado del objeto a detectar, el espectro de frecuencia espacial es más probable que sea apropiado ³⁴. En cambio, el CNR es más confiable que el SNR a la hora de hacer controles semanales ya que se pueden observar pequeños cambios en la calidad de imagen en el largo plazo y dar resultados más reproducibles ³⁴.

Por todas las limitaciones reportadas, incluyendo la incertidumbre de causa múltiple asociada a los resultados de los trabajos revisados (que impidió la realización de un metaanálisis), el esperable menor rendimiento (menor que el mejor rendimiento humano) de los métodos computacionales más confiables, como el pareo de patrones, en condiciones de bajo contraste y las inconsistencias humanas en respuesta a factores externos a la tarea, Alawaji et al. proponen que los métodos de control de la calidad mamográfica basados en imágenes de fantasmas físicos que sigan recomendaciones de sociedades científicas o de los fabricantes de los equipos, deben ser híbridos, con tres componentes: físico (al menos SNR y CNR), de detección computacional de patrones (pareo de plantillas con) y de observadores humanos ³⁴. Cabe destacar que el rol que los autores asignan a los asesores humanos es de revisores de informes automáticos que contengan resultados de los métodos físicos de evaluación de la calidad de las imágenes y de la detección computacional de patrones ³⁴.

5. Discusión

El uso de los fantasmas permite elaborar respuestas a algunas preguntas relevantes en la imagenología mamaria, sin embargo, tenemos interrogantes sobre ellos mismos, preguntas que aún no están respondidas en la literatura, como queda claro de la lectura de los trabajos revisados. Por un lado, existen múltiples fantasmas comerciales y no comerciales y múltiples procesos de empleo de los mismos, no hay por tanto un estándar de referencia mundial o ampliamente acordado en distintos continentes, y es dudoso que esto sea posible o beneficioso, si las tecnologías permanecen heterogéneas. De hecho, las nuevas tecnologías de receptores han exigido el desarrollo de nuevos fantasmas. La diversidad de opciones de fantasmas y procesos, la dinámica del desarrollo de nuevas tecnologías y las desigualdades de intereses y posibilidades económicas, profesionales y sociales en distintas localizaciones dificultan la racionalidad científica en el uso de fantasmas y en la toma de decisiones de aceptación y abandono de tecnologías. Sin embargo, como veremos también existen dificultades debidas a la calidad de las investigaciones secundarias revisadas en este trabajo y de las investigaciones primarias recogidas por ellos. Dichos problemas de calidad no fueron suficientemente analizados y discutidos en Alawaji et al. y de Øynes et al. aquí revisados ^{5,34}.

Se requiere disponer de equipamiento con tecnologías de alta confiabilidad que generen una buena calidad de imagen con dosis glandulares tan bajas como sea posible e implementar programas de aseguramiento de calidad con controles periódicos, ya que en caso contrario muchas de las alteraciones patológicas como las microcalcificaciones o estructuras fibrosas o nodulares o asimetrías, pueden no ser detectadas en imágenes de screening del cáncer de mama⁴. Para mejorar la detección de alteraciones patológicas asociadas al cáncer de mama en etapa diagnóstica se han implementado distintas técnicas complementarias físicas que utilizan los licenciados en imagenología como la magnificación y la compresión focal ⁴, pero también se han creado técnicas computacionales como es la ampliación digital que utilizan los médicos imagenólogos ⁵. Para el control rutinario de calidad de las imágenes en las etapas de tamizaje y diagnóstica se utilizan guías de aseguramiento de calidad que emplean procedimientos de revisión físicos sobre instrumentos e imágenes a cargo de físicos médicos y procedimientos de revisión por expertos médicos ²⁵, pero se han creado también procedimientos computacionales semiautomáticos o automáticos que usan algoritmos de reconocimiento automático de patrones que podrían emplear los informáticos médicos o personal especializado en imágenes mamográficas.

De los 831 resultados únicos identificados en este trabajo, sólo 2 artículos fueron finalmente incluidos como revisiones sistemáticas que estudiaran la calidad de las imágenes usando como dispositivo a fantasmas mamarios en el período de estudio (Fig. 1). Resumidamente, Øynes et al. son particularmente optimistas sobre el uso

futuro de la amplificación digital en la medida en que se adopten las nuevas tecnologías de receptores de conteo de fotones ⁵. Mientras la magnificación geométrica implica exponer nuevamente la persona a radiaciones ionizantes, utilizar solamente amplificación electrónica, no. Esta ventaja de la amplificación disminuye en la medida en que se mejoren las condiciones de adquisición de imágenes, ya que las dosis disminuirán correspondientemente. Øynes et al. reportan que el zoom podría evitar hoy la mamografía de magnificación en ciertas condiciones ⁵. Sin embargo, son muy específicas las condiciones en que el zoom parece ser superior a la magnificación (ver Tabla IV), posibilidad que no se pudo confirmar en un metaanálisis, por lo que la evidencia de dicha superioridad es probablemente débil. Por otro lado, Alawaji et al. son particularmente optimistas sobre el uso futuro del aprendizaje profundo en métodos automáticos de evaluación de la calidad de imágenes mamográficas, en la medida de que los juegos de imágenes etiquetados se hacen más grandes y diversos ³⁴. Ellos reportan que la inconsistencia de los juicios humanos debidas a influencias ajenas a la tarea específica hace que los controles subjetivos rutinarios de la calidad de las imágenes mamográficas no sean confiables; a diferencia de ello, los métodos de reconocimiento automático o semiautomáticos son internamente consistentes y en condiciones específicas podrían rendir mejor que la evaluación subjetiva ³⁴. Alawaji et al. proponen que hoy el pareo de plantillas en conjunto con la transformación discreta de ondículas y el uso de indicadores físicos de la calidad de la imagen como SNR y CNR serían suficientes para producir controles rutinarios de calidad basados en informes automáticos auditables por asesores expertos ³⁴.

Dejando de lado las especulaciones a futuro, a nuestro juicio, ninguna de las dos revisiones sistemáticas arroja evidencia empírica suficiente para apoyar la realización de los cambios tecnológicos que se proponen para el día de hoy. En contra del peso científico de los cambios propuestos, hacemos notar que ambas revisiones son de calidad críticamente baja de acuerdo con AMSTAR 2 (Tabla III). De hecho, aunque cualquier aspecto de una tecnología puede ser objeto de un estudio de equivalencia, superioridad o no inferioridad, más frecuentemente se hace referencia principalmente a la comparación de eficacias de intervenciones o de exactitudes de métodos de detección. Los controles rutinarios de calidad de servicios mamográficos son intervenciones y les corresponde por tanto al menos estudios comparativos de eficacia. Por otro lado, la detección computacional de patrones y la ampliación electrónica son métodos a los que deben corresponderles estudios comparativos de exactitud. Podemos concluir de este razonamiento, que ninguna de las revisiones tomó en cuenta aspectos de exactitud o de eficacia comparativa en términos cuantitativos metaanalíticos, por razones que los autores atribuyen a los trabajos primarios. Por supuesto, los estudios comparativos de seguridad han quedado por fuera de las revisiones, y este tipo de investigaciones secundarias serían muy bienvenidas en nuestra opinión.

Del análisis de los resultados de nuestra revisión surgen dos argumentaciones en contra de la conclusión principal del trabajo de Øynes et al. de que la ampliación digital es probablemente equivalente a la magnificación geométrica en muchas condiciones ⁵:

“In conclusion, zoom may be equivalent to magnification in many cases given that optimised procedures and newer detector technologies are now available. This finding has the potential to reduce AGD and improve examination workflow. Both diagnostic test studies and phantom studies using newer detectors would contribute additional knowledge on this topic.”

“En conclusión, la ampliación electrónica puede ser equivalente a la magnificación en muchos casos dado que procedimientos optimizados y nuevas tecnologías de detectores están ahora disponibles. Este hallazgo tiene el potencial de reducir la DGM y mejorar el flujo de trabajo de los estudios. Tanto los estudios de pruebas diagnósticas como de fantasmas que usen nuevos detectores contribuirían conocimiento adicional en este tópico”

La primera argumentación es de carácter lógico: ninguno de los trabajos sobre fantasmas incluidos la revisión de Øynes et al. empleó pruebas de equivalencia; algunos realizaron pruebas de superioridad, otros no realizaron prueba estadística alguna. Por tanto, la conclusión es falaz, no debería usarse el concepto de equivalencia porque ella nunca fue puesta a prueba (ver más abajo). Dado que además de estudios sobre fantasmas, Øynes et al. incluyeron en otra rama de la revisión, un metaanálisis de estudios primarios sobre imágenes clínicas cabría la posibilidad de que la conclusión se basara en esta parte de la revisión. Sin embargo, el estudio metaanalítico mostró limitaciones serias por problemas de heterogeneidad que deberían impedir a los autores extraer conclusiones, por ende la afirmación, aunque se muestra cautelosa es en realidad una especulación sesgada. Sería recomendable en lugar de realizar tal afirmación, emplear el principio de precaución, que presume un sesgo hacia la incorporación tecnológica. Esto se refiere a considerar y eventualmente moderar la inclinación a adoptar tecnologías que podrían ofrecer ventajas si las ventajas previstas son mínimas o si carecen de evidencia sólida ⁷². En este sentido, vemos que Øynes et al. afirman la posibilidad de que ambas técnicas puedan ser equivalentes, pero al no ser definitiva la evidencia, lo más correcto sería destacar su potencial (por ejemplo que la ampliación electrónica disminuye la exposición a RX), dejar la pregunta sobre la equivalencia abierta para futuras investigaciones, sin sacar conclusiones riesgosas favorables sobre la tecnología de la ampliación digital, y considerar especialmente las falencias y posibles mejoras en el diseño de los trabajos primarios. Desde el punto de vista de otros autores y desde el punto de vista físico e imagenológico, la magnificación geométrica y la ampliación digital no son cosas equivalentes. De acuerdo con Mercer et al., la magnificación es el estándar de referencia en cuanto a la correcta detección de microcalcificaciones para el diagnóstico de patologías

mamarias y la ampliación digital tendría un lugar relevante en la mamografía de screening ⁴. De acuerdo con estos autores, basar el diagnóstico en la reexploración de la imagen de screening usando el ampliación digital no sería adecuado para el diagnóstico mamario porque: a) el zoom incrementa el tamaño de la imagen pero no genera una mejor definición de las microcalcificaciones, lo que puede llevar a una pérdida de la nitidez de los bordes de las estructuras y, b) no siempre quedan demostradas todas las microcalcificaciones en esta primera mamografía (sobre la que se hace el zoom), siendo ventajoso realizar una segunda toma de datos ⁴, donde además generalmente se aplica una técnica complementaria (magnificación, compresión, etc.). Las investigaciones que podrían permitir dilucidar la interrogante sobre la equivalencia clínica de la ampliación digital y la magnificación geométrica para el diagnóstico deben incluir el estudio sistemático sobre los equipamientos, los fantasmas, las medidas integrales de calidad de imagen computacionales y basadas en expertos y centrarse en los desenlaces, resultados e impactos clínicos relevantes a través de diseños experimentales randomizados, controlados y doble ciego donde las pruebas estuvieran orientadas a confirmar la no inferioridad o equivalencia propuesta. Pensamos que publicar un protocolo con estas características sería una contribución interesante en el campo. Mientras esos estudios no se realicen es de gran importancia perfeccionar y asegurar la calidad de las técnicas que hoy se consideran de referencia. Ni la ampliación digital o la magnificación geométrica pueden solucionar los problemas potenciales de calidad de imagen de tamizaje donde la detección de microcalcificaciones es sumamente importante para la prevención secundaria del cáncer de mama. Se debería mejorar la tecnología de detección y producción de rayos X y mejorar la calidad de las imágenes mamográficas, implementando programas de aseguramiento de la calidad basados en evidencias científicas, tema que se estudia en la revisión de Alawaji et al.

Øynes et al. proponen que usar la ampliación digital en el marco tecnológico del conteo de fotones rinde similares resultados a usar magnificación en el marco de las tecnologías de receptores de paneles planos ⁷. Øynes et al. no incluyeron trabajos que comparen la magnificación geométrica y la ampliación electrónica en el marco del conteo de fotones, siendo una pregunta de investigación interesante futura. El tamaño del píxel del detector, es un punto importante de destacar a la hora de la detección de las microcalcificaciones, ya que, cuanto menor sea su tamaño es mayor la necesidad de la magnificación, por lo cual, también la tecnología de conteo de fotones en los nuevos detectores (con píxeles más pequeños) sería beneficiosa ⁷³. Continuar aumentando la resolución espacial, que podría ser una ventaja si microcalcificaciones menores a 50 μm pudieran alertar de lesiones malignas, no sería la única ventaja del conteo de fotones: los estudios espectrales mamográficos posibles en este marco tecnológico, permiten también distinguir la naturaleza química de las microcalcificaciones ⁷³. Esto es así porque se ha llegado a un acuerdo amplio en que las microcalcificaciones mamarias se pueden clasificar mayoritariamente en dos tipos: tipo I, de oxalato de calcio, asociadas calcificaciones intraductales por lesiones benignas; y tipo II, con hidroxapatita, asociadas a

lesiones extraductales relacionadas con procesos de muerte y proliferación celular malignos o benignos y porque existen evidencias recientes de que el conteo de fotones según su energía permite distinguir estos tipos de microcalcificaciones ⁷³. Preveamos que en el futuro se realizarán estudios que investiguen cómo estas cuatro cosas (mejoras en la resolución espacial de los detectores, ampliación digital, magnificación geométrica y conteo espectral de fotones) interaccionan en el screening del cáncer de mama y en el diagnóstico de patologías mamarias.

Por otro lado, se puede hipotetizar que la utilización de inteligencia artificial/aprendizaje automático a través de algoritmos de superresolución podría evitar la realización de la mamografía de magnificación, disminuyendo la borrosidad causada por el ampliación digital ⁷⁴, una potencial mejora que el trabajo de Øynes et al. no estudia ni considera. Estos algoritmos podrían aumentar la resolución espacial de la imagen lo suficiente para no perder la nitidez de la imagen de los bordes de las estructuras en la ampliación digital ⁷⁴. Las tecnologías basadas en aprendizaje automático dedicadas al mejoramiento de imágenes mamarias han sido poco testeadas en la clínica pero tienen tanto beneficios como riesgos y dificultades esperados. Una dificultad radica en el entrenamiento de los algoritmos y la generalizabilidad (por ejemplo para cualquier equipo en cualquier condición, o población) o adaptabilidad (por ejemplo para refinar el entrenamiento con parámetros locales ajustables) de los modelos aprendidos ⁷⁵. No vamos a desarrollar más comentarios sobre esto aquí, dado que no es una temática abordada por la revisión de Øynes et al., pero señalamos que podría resultar interesante estudiar la equivalencia de la ampliación digital con superresolución y la magnificación geométrica, claro está, con diseños adecuados.

Siguiendo con el análisis propuesto (ver más arriba) consideramos que los autores nos brindan un pensamiento falaz, donde la ampliación digital y la magnificación geométrica podrían ser equivalentes según los autores ⁵. Esto nos llevó a revisar los artículos originales y algunos de ellos resultaron ser muy escuetos en evidencia estadística. Los artículos de Egan et. al.⁷⁶, Koutalonis et. al.⁷⁷ y Hermann et. al.⁷⁸ en su sección de métodos no describen estadísticas que hayan sido utilizadas, a sí mismo, en sus resultados no presentan medidas de resumen de tendencia de dispersión como el desvío estándar, rango intercuartil u otros. El artículo de Alkhalifah et al.⁷⁹ utiliza la prueba de H de Kruskal-Wallis, un test no paramétrico, para los resultados agrupados según las diferentes combinaciones ánodo filtro obteniendo el score de cada uno para ampliación digital y magnificación geométrica, evaluando su superioridad. El trabajo de Vahey et. al. ⁸⁰ al igual que el anterior, es un artículo que utiliza estadística de superioridad realizando la prueba Mann–Whitney U-test (versión no paramétrica de la prueba t de Student) comparando muestras independientes. Por estas razones la idea de equivalencia no debería estar presente en sus conclusiones ya que no está respaldado por las pruebas apropiadas. Este tipo de afirmaciones falaces en las conclusiones de trabajos científicos no es un problema desconocido para la

comunidad científica imagenológica. Park et al. en 2014 reportaron que de 38 trabajos primarios estudiados donde se realizaban afirmaciones de no inferioridad o similitud en las conclusiones sólo 1 había realizado las pruebas estadísticas correspondientes a la afirmación y 29 realizaron test convencionales de superioridad que no corresponden con la afirmación ⁸¹. May et al. en 2018 reportaron que en el área de la pediatría radiológica el fenómeno no es poco frecuente tampoco: 36 de 112 trabajos primarios que realizaron comparaciones afirmaron falazmente que los métodos estudiados eran equivalentes cuando ninguno de los 112 estudios desarrolló un diseño adecuado de estudio de equivalencia o no inferioridad ⁸². Esto ocurre a pesar de publicaciones que en 2013 alertaban de la necesidad de diseñar cuidadosamente estudios imagenológicos comparativos que buscan estudiar equivalencia o no inferioridad ⁸³. En 2019, la Revista Británica de Educación en Anestesia (BJA Education) publicó un trabajo que resume las características de los ensayos de no inferioridad, cuyo diseño debe ser cuidadoso ya que el riesgo con estudios de baja potencia es aceptar tratamientos o métodos inferiores como si fueran no inferiores al estándar de referencia ⁴⁹. Por todo lo anterior parece ser necesario en el área de la imagenología médica reforzar la educación en el diseño de estudios adecuados, incluyendo consideración a los ensayos de equivalencia y no inferioridad.

El reporte de Alawaji et al. sobre el uso de fantasmas en la evaluación automática de la calidad de imágenes diferencia a la evaluación computacionales (automáticas o semiautomáticas) de las formas de evaluación llamadas subjetivas (dependientes del juicio de los humanos)³⁴. Los autores jerarquizan, sin aportar definiciones, las siguientes características negativas de los enfoques subjetivos: baja fiabilidad (del inglés “reliability”) inter e intra observadores y dificultad para mantener la consistencia ” inter e intra observadores en función de factores externos (rango de 89-97% y aprox. 76% en estudios revisados por los autores)³⁴. Como característica positiva de los observadores humanos Alawaji et al. mencionan su capacidad para detectar objetos con mínimo contraste y para identificar artefactos ³⁴. A favor de los enfoques automáticos comparados con los enfoques subjetivos Alawaji et al. mencionan las siguientes características: eficiencia alta, alta realizabilidad (“feasible”), costo bajo, fiabilidad alta ⁸⁴, alta consistencia y alta reproducibilidad ³⁴. En contra de los enfoques automáticos comparados con los enfoques subjetivos, Alawaji et al. incluyen la dependencia en el contraste y la identificación errónea de artefactos como microcalcificaciones ³⁴

Según la Sociedad Americana para la Calidad (“American Society for Quality”, ASQ) la fiabilidad es ⁸⁴:

“la probabilidad de que un producto, sistema o servicio vaya a rendir o proporcionar la función para la que fue creado en forma adecuada por un período de tiempo especificado, o vaya a operar en un entorno definido sin fallar”

Esta definición de arriba está basada en literatura que incluye la obra de Birolini ⁸⁵ y otros libros del campo de la ingeniería de la fiabilidad. De acuerdo con dicha literatura, el fallo no implica que no se detecten y reparen errores/problemas, sino que se refiere a que el proceso se detenga a causa de ellos, ergo, la fiabilidad se puede mejorar a través de la ingeniería del diseño del sistema ⁸⁵. La ASQ también define fiabilidad como una medida que utiliza distintas variables indicadoras: la probabilidad de éxito, la durabilidad, la confianza, la calidad a lo largo del tiempo y la disponibilidad para realizar una función ⁸⁴. Según Birolini, en el campo de la fiabilidad informática, la consistencia y fiabilidad de un programa se definen respectivamente como el *“grado de uniformidad, estandarización y ausencia de contradicciones dentro de la documentación o partes de un paquete de software”* y el *“grado en el que un paquete de software puede ejecutar la función requerida sin causar fallas en el sistema”* ⁸⁵. Aplicadas estas definiciones a los programas basados en juicios subjetivos se puede concluir que Alawaji et al. afirman que los sistemas de aseguramiento de calidad subjetivos detendrían su operación con una probabilidad demasiado alta para ser aceptable o que la calidad de su operación sería menor que la adecuada, y que los procesos deben tener soluciones o formas de reparación para errores debidos a la baja consistencia de las opiniones de expertos que son altamente probables. Por el contrario, los análisis y la detección computacionales de las imágenes aportarían a los programas de aseguramiento de calidad una mayor probabilidad de mantenimiento de la operativa y un funcionamiento más adecuado, con inconsistencias nulas o despreciables. Estas afirmaciones, no son más que hipótesis que se pueden construir a partir de las propuestas contenidas en la revisión de Alawaji et al. y que se deberían poner a prueba en forma experimental usando diseños adecuados, como los de los ensayos de no inferioridad ⁴⁹. Es importante notar que existen varias comparaciones a realizar si consideramos el paradigma propuesto por los autores: las formas subjetivas (o manuales) y las objetivas (o automáticas) de extraer información de la calidad de las imágenes a partir del conteo de objetos o de la medición de índices de calidad físicos:

	Subjetiva/manual	Objetiva/automática
Detección y conteo de objetos/patrones		
Medición de índices físicos de calidad		

Este conjunto de 4 comparaciones posibles subyace al trabajo de Alawaji et al. si aceptamos la premisa de la separación de lo subjetivo versus lo objetivo y la asociación de lo subjetivo con lo manual y de lo objetivo con lo automático. Definir los conceptos de objetividad, subjetividad, manualidad y automaticidad es por ende fundamental para discutir el trabajo de Alawaji et al. Objetividad es entendida como un valor, algo objetivo es de mayor importancia y goza de mayor aprobación que algo que difiere de lo anterior solo por su subjetividad y se refiere a generar una mirada desde ningún lugar, una mirada sin perspectiva particular ⁸⁶. Es posible que el aspecto que más importe de la objetividad en relación al tema en estudio es su potencial de mayor aprobación, ya que en definitiva se debe llegar a elaborar y aceptar una decisión como firme. No hay una definición objetiva y unánime de objetividad. La Enciclopedia de Filosofía de Stanford reporta 4 formas principales distintas de comprender la objetividad (en vinculación con la ciencia): como fidelidad a los hechos, como ausencia de compromisos normativos y libertad de valores, como ausencia de prejuicios personales o idiosincrasias y cuarto, como una forma de unidad conceptual ⁸⁶. La primera, la forma más ingenua, parece ser la forma de comprensión que propone implícitamente el trabajo de Alawaji et al. Vale la pena considerar las limitaciones que puede significar esta forma de comprensión a la hora de conocer y entender los fenómenos relacionados con la calidad de las imágenes de fantomas mamarios. En esa dirección, comenzamos por preguntarnos, cuáles son los supuestos hechos y sus explicaciones deterministas: 1) El fantoma contiene objetos radio absorbentes de propiedades conocidas porque el objeto fue fiablemente fabricado bajo estándares previamente establecidos; 2) El objeto está representado en la imagen porque el objeto tiene las propiedades adecuadas y porque las condiciones de obtención de la misma son de calidad aceptables; 3) La imagen tiene propiedades características que permiten medir la calidad de la imagen porque tenemos las teorías e instrumentos necesarios; 4) Un humano experimentado detecta la imagen del objeto en la imagen completa porque se dan las condiciones en 2 y porque existen condiciones aceptables de calidad de visualización, visión y procesamiento perceptual. En este sentido, es lógico aceptar que tanto los enfoques manuales como automáticos pueden estar orientados a intentar verificar los hechos 1, 2, 3 y/o 4 y sus determinantes como partes objetivas de la realidad. La diferencia está en el hecho en sí mismo y también en los determinantes: el caso 4 incluye en el sistema al evaluador de las imágenes. La objetividad entonces es un valor y/o aspiración que no excluye a la evaluación por expertos, la que también genera hechos. Otra observación importante es que las condiciones en las cuales se recogen los resultados o se evalúan los hechos son sumamente importantes, las que en sí pueden ser vistas como hechos previos y necesarios, desde un punto de vista determinista. A su vez, considerar la objetividad como ausencia de prejuicios personales o idiosincrasias nos permite ver la cualidad opuesta, la subjetividad, que puede ser observada también en múltiples hechos que podemos reconocer en los procesos que llevan a la definición de los algoritmos informáticos o a los procesos de elaboración y selección de teorías y algoritmos que se utilizaron por ejemplo en la detección de patrones en los trabajos revisados por

Alawaji et al. Las ideas en este párrafo mantienen un conjunto importante de acuerdos con trabajos previos ⁸⁷, pero es importante reconocer y el lector lo comprobará fácilmente, que este aspecto criticable del trabajo de Alawaji et al. no es más que un problema de encuadre o sesgo implícito muy frecuente.

La automaticidad y el holismo son dos conceptos que también merecen análisis al momento de discutir el reporte de Alawaji et al. Los autores proponen la realización de métodos holísticos para evaluar la calidad de imagen a partir del uso de cualquiera de los fantasmas acreditados ³⁴, métodos que consideramos necesariamente como semiautomáticos de acuerdo a su propuesta de incluir procesos computacionales y revisiones humanas, es decir que requieren intervención humana directa restringida a la validación de informes con alertas automáticas; y mixtos, ya que como se propone estarían basados tanto en la detección de patrones como en la medición de índices cuantitativos de calidad de imagen posiblemente usando un nuevo fantoma o cualquier fantoma acreditado a través del empleo combinado de algoritmos de pareo de plantillas, mediciones de SNR y CNR y transformada discreta de ondículas. De este modo introducimos una corrección en la forma de entender y/o comunicar la propuesta de Alawaji et al.

	Manual	Automática	Semiautomática
Detección y conteo de objetos/patrones en imágenes originales y con TDO			
Medición de índices físicos de calidad en imágenes originales			
Evaluación mixta			Método holístico propuesto por Alawaji et al.

Esta revisión sistemática toma en cuenta 26 artículos que refieren al control de calidad de servicios mamográficos con variaciones grandes recursos y una heterogeneidad tal que impiden la extracción de conclusiones cuantitativas ³⁴. Las variaciones en los sistemas de mamografía son amplias ^{88,89}, por lo que un programa sólido de control de calidad debe incluir a todas estas variaciones ⁹⁰. La necesidad

de establecer un grupo de índices físicos y psicofísicos (por ejemplo los índices de detectabilidad) a través de un acuerdo internacional precisa de revisiones sistemáticas y probablemente nuevos ensayos comparativos bien diseñados de los referidos índices obtenidos a partir de un conjunto amplio de fantasmas acreditados para medir la calidad de imagen para distintos servicios y equipos. Pensamos que los índices psicofísicos ⁹¹ podrían otorgar mayor fiabilidad a los sistemas de control de la calidad de las imágenes, incluso en el enfoque holístico que proponen Alawaji et al., quienes reconocen que los análisis computarizados por sí solos no son de gran aporte en imágenes de objetos de bajo contraste, por lo que propone agregar medidas cuantitativas SNR y CNR ⁹²⁻⁹⁴ que aporten información más consistente y factible y confiable para valorar los pequeños cambios que hay en las imágenes, esto contribuye a mejorar los servicios de mamografía ^{63,95}. La evaluación de la calidad de las imágenes mamográficas debe hacerse a través de un programa de aseguramiento de calidad, el cual cuenta con tres vertientes 1) observadores humanos 2) análisis computacional y 3) físicos médicos especializados en equipos mamográficos según guías internacionales actuales y revisiones recientes ^{90,94,96}. Reis et al. han reportado una falta en las mencionadas guías de definiciones sobre la formación y capacidades que debe tener el experto que participa del proceso de evaluación de la calidad de las imágenes ⁹⁰. Tampoco lo hacen Alawaji et al. ³⁴. Esto es interesante desde la perspectiva de la carrera de Licenciatura en Imagenología, donde varios equipos están trabajando en problemas relacionados con el control y aseguramiento de la calidad de imágenes mamográficas. Es entonces posible y deseable avanzar en un diálogo interdisciplinario que permita identificar capacidades que deban enriquecerse para la formación de licenciados con habilidades suficientes en el control de calidad. En Uruguay, la variación tecnológica es amplia, contamos con equipos mamográficos con lectura indirecta (CR), lectura directa (DR) y lectura directa con tomosíntesis ⁴. El control de calidad utilizando fantasmas mamográficos para la detección de pequeñas variaciones en las imágenes hasta donde sabemos no está instalado en nuestro medio, aunque hay esfuerzos para llegar a ello. Uruguay no cuenta con un programa de aseguramiento de la calidad de los servicios mamográficos. Existen controles de forma sistemática en algunos centros de atención pero no siguen una normativa específica de aseguramiento para evaluar la calidad de las imágenes en un entorno clínico ⁹⁰. Consideramos que esta problemática no involucra solamente a nuestro país, existen otros países de medianos ingresos que enfrentan esta situación, desde nuestro lugar buscamos un programa de aseguramiento del control de calidad que involucre todas las tecnologías disponibles, los fantasmas mamográficos validados, los profesionales con dedicación a la imagenología mamaria y las herramientas de análisis computarizado disponibles para solucionar este problema que define la calidad de imágenes mamográficos que detectan el cáncer de mama en nuestra población como una problemática de salud mundial ⁸⁹.

En conclusión, existen en la literatura revisada evidencia sesgada y débil empíricamente, con problemas de diseño y que enfrenta dificultades grandes para ser realizada correctamente dada la diversidad de intereses que existen, que es favorable a los cambios tecnológicos que se estudian. Cabe reconocer que la aceptación de una nueva tecnología y el abandono y/o no aceptación de otra no es siempre una decisión racional y científica, más bien es un proceso psicosociotecnológico, demográfico y económico complejo ^{97,98}. Según algunos autores, el reemplazo puede ocurrir por distintos disparadores, como por ejemplo que sea convincentemente mostrado que una es superior a la otra o porque existen cuestionamientos a la eficacia de una de ellas, reportándose también que estas influencias parecen estar fuertemente moderadas por las misiones de las organizaciones de la salud donde se tomen las decisiones ⁹⁹. También los riesgos, su probabilidad y la magnitud del daño potencial entran en juego en la toma de estas decisiones. Este es un mensaje particularmente relevante para las organizaciones académicas de la salud hoy en un tiempo en el que se especula sobre los beneficios y peligros de la salud digital y de la inteligencia artificial ¹⁰⁰, así como en otros momentos históricos se debatió sobre el uso de la energía nuclear como fuente de energía para los hogares, industrias y servicios ¹⁰¹. Claramente, las nuevas tecnologías pueden producir mejores soluciones a problemas previos, pero también pueden producir nuevos problemas, ser indiferentes ante viejos problemas o incluso potenciarlos ¹⁰².

6. Conflicto de interés

Los autores de esta revisión declaran no tener conflicto de interés con alguna empresa, productos o servicios que podrían estar relacionados a la revisión.

7. Referencias bibliográficas

1. Castrezana Campos M del R. Geografía del cáncer de mama en México. *Investig Geográficas Bol Inst Geogr.* 2017;2017(93):140-157. doi:10.14350/rig.56879
2. Milans S, Lavista F. Introducción a la Imagenología Mamaria. *Tend En Med Rev Actual Médica Contin.* 2017;50:133-140.
3. Ruiz Lorente R, Rubio Hernández MC, Hernández Durán D, Tamayo Lien G. Experiencia en mujeres con cáncer de mama durante un trienio. *Rev Cuba Obstet Ginecol.* 2011;37(3):349-358.
4. Mercer C, Hogg P, Kelly J. *Digital Mammography: A Holistic Approach.* Springer Nature; 2022.
5. Øynes M, Strøm B, Tveito B, Hafslund B. Digital zoom of the full-field digital mammogram versus magnification mammography: a systematic review. *Eur Radiol.* 2020;30(8):4223-4233. doi:10.1007/s00330-020-06798-6
6. Ikeda D, Miyake KK. *Breast Imaging: The Requisites: Breast Imaging: The Requisites E-Book.* Elsevier Health Sciences; 2016.
7. Feigin K. Quality assurance in Mammography: An overview. *Eur J Radiol.* 2023;165:110935. doi:10.1016/j.ejrad.2023.110935
8. Bushong SC. *Manual de radiología para técnicos: Física, biología y protección radiológica.* Elsevier Health Sciences; 2017.
9. Dos Santos Andrade LS, De Melo Santos TT, Case de Oliveira ME, et al. Shorter delay to treatment by integrated diagnostic services and NGO-provided support among breast cancer patients in two Brazilian referral centres. *J Public Health Res.* 2021;10(3). doi:10.4081/jphr.2021.1880
10. Marcotte LM, Deeds S, Wheat C, et al. Automated Opt-Out vs Opt-In Patient Outreach Strategies for Breast Cancer Screening: A Randomized Clinical Trial. *JAMA Intern Med.* Published online September 11, 2023. doi:10.1001/jamainternmed.2023.4321
11. Galván-Espinoza HA. La necesidad e importancia del control de calidad en mamografía. 2012;11.
12. Carrasco Cazco MZ. Simulación de un fantoma de mama para dosimetría, utilizando Montecarlo. Published online August 28, 2020. Accessed October 7, 2023. <http://dspace.esPOCH.edu.ec/handle/123456789/14304>
13. Ay MR, Sarkar S, Shahriari M, Sardari D, Zaidi H. Assessment of different computational models for generation of x-ray spectra in diagnostic radiology and mammography. *Med Phys.* 2005;32(6Part1):1660-1675. doi:10.1118/1.1906126
14. Actis AM. Consideraciones bioéticas en relación con el uso de la inteligencia artificial en mastología. *Rev Médica Urug.* 2021;37(4). doi:10.29193/rmu.37.4.12
15. Brooks KW, Trueblood JH, Kearfott KJ, Lawton DT. Automated analysis of the American College of Radiology mammographic accreditation phantom images. *Med Phys.* 1997;24(5):709-723.
16. Ropohl G. Philosophy of Socio-Technical Systems. *Techné Res Philos Technol.* 1999;4(3):186-194.
17. Unver MB. Governing fiduciary relationships or building up a governance model for trust in AI? Review of healthcare as a socio-technical system. *Int Rev Law Comput Technol.* 2023;37(2):198-226. doi:10.1080/13600869.2023.2192569
18. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions.* John Wiley & Sons; 2019.
19. Trueba-Gómez R, Estrada-Lorenzo JM. La base de datos PubMed y la búsqueda de información científica. *Semin Fund Esp Reumatol.* 2010;11(2):49-63. doi:10.1016/j.semreu.2010.02.005
20. Packer A. Biblioteca virtual en salud. Published online 2000.
21. Cáceres Gini JM. EBSCO Host: base de datos de información académica y

- multidisciplinaria. *Univ Peru Cienc Apl UPC*. Published online September 25, 2020. Accessed October 7, 2023. <https://repositorioacademico.upc.edu.pe/handle/10757/652924>
22. Burnham JF. Scopus database: a review. *Biomed Digit Libr*. 2006;3(1):1. doi:10.1186/1742-5581-3-1
 23. Hunter K. ScienceDirect™. *Ser Libr*. Published online March 25, 1998. doi:10.1300/J123v33n03_07
 24. Seroubian M. Acceso abierto y ciencia abierta. Experiencia desde la gestión del repositorio institucional COLIBRI de la Universidad de la República. *Informatio*. 2022;27(1):284-308.
 25. Ehrenpreis M. IOPscience. *Charlest Advis*. 2018;20(1):45-47.
 26. Durniak A. Welcome to IEEE Xplore. *IEEE Power Eng Rev*. 2000;20(11):12.
 27. Cecchino NJ. Google Scholar. *J Med Libr Assoc JMLA*. 2010;98(4):320-321. doi:10.3163/1536-5050.98.4.016
 28. Rubianes E. La investigación y práctica médica y el portal TIMBÓ. *Rev Médica Urug*. 2009;25(1):3-4.
 29. Stephens TW, Wang I. Image acquisition. *Digit Mammogr Pract Approach*. 2012;6(10):18.
 30. RIS (file format). In: *Wikipedia*. ; 2023. Accessed September 10, 2023. [https://en.wikipedia.org/w/index.php?title=RIS_\(file_format\)&oldid=1165815812](https://en.wikipedia.org/w/index.php?title=RIS_(file_format)&oldid=1165815812)
 31. EndNote. In: *Wikipedia, la enciclopedia libre*. ; 2023. Accessed October 3, 2023. <https://es.wikipedia.org/w/index.php?title=EndNote&oldid=151088805>
 32. INFORMATION - Rayyan. Published April 20, 2022. Accessed October 3, 2023. <https://www.rayyan.ai/cite/>
 33. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. Published online September 21, 2017;j4008. doi:10.1136/bmj.j4008
 34. Alawaji Z, Tavakoli Taba S, Rae W. Automated image quality assessment of mammography phantoms: a systematic review. *Acta Radiol Stockh Swed* 1987. 2023;64(3):971-986. doi:10.1177/02841851221112856
 35. Mammography Compliance Phantoms (Gammex™ Technology) - Sun Nuclear. Accessed November 2, 2023. <https://www.sunnuclear.com/products/mammography-compliance-phantoms>
 36. Spyrou GP, Tzanakos G. MASTOS: mammography simulation tool for design optimization studies. *Med Inform Internet Med*. 2000;25(4):275-293.
 37. Yip M, Chukwu W, Kottis E, et al. Automated scoring method for the CDMAM phantom. In: Vol 7263. SPIE; 2009:422-431.
 38. Pro-MAM AEC PMMA > Mammography Phantoms > Nuclemed. Accessed November 2, 2023. <https://www.nuclemed.be/product.php?cat=136&prod=786>
 39. TOR MAS. Leeds Test Objects. Accessed October 31, 2023. <https://leedstestobjects.com/index.php/phantom/tor-mas/>
 40. TOR MAM. Leeds Test Objects. Accessed October 31, 2023. <https://leedstestobjects.com/index.php/phantom/tor-mam/>
 41. TOR MAX. Leeds Test Objects. Accessed October 31, 2023. <https://leedstestobjects.com/index.php/phantom/tor-max/>
 42. Tissue-Equivalent Phantom for Mammography. CIRS. Accessed October 31, 2023. <https://www.cirsinc.com/products/mammography/tissue-equivalent-phantom-for-mammography/>
 43. Adel M, Guis VH, Rasigni M, Bourennane S. Feasibility study of automating breast phantom scoring using image processing techniques. In: Vol 5672. SPIE; 2005:201-211.
 44. Mayo P, Rodenas F, Campayo JM, Verdú G. Quality assurance applied to mammographic equipments using phantoms and software for its evaluation. *Nucl Instrum Methods Phys Res Sect Accel Spectrometers Detect Assoc Equip*. 2010;619(1-3):372-374.
 45. Lu ZF, Nickoloff EL, So JC, Dutta AK. Comparison of computed radiography and

- film/screen combination using a contrast-detail phantom. *J Appl Clin Med Phys*. 2003;4(1):91-98. doi:10.1120/jacmp.v4i1.2548
46. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
 47. 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(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009
 48. World Health Organization. *Formulación de Políticas Sobre Dispositivos Médicos*. Organización Mundial de la Salud; 2012.
 49. Walker J. Non-inferiority statistics and equivalence studies. *BJA Educ*. 2019;19(8):267-271. doi:10.1016/j.bjae.2019.03.004
 50. Sundell VM, Mäkelä T, Meaney A, Kaasalainen T, Savolainen S. Automated daily quality control analysis for mammography in a multi-unit imaging center. *Acta Radiol*. 2019;60(2):140-148.
 51. Gennaro G, Ballaminut A, Contento G. A multiparametric automatic method to monitor long-term reproducibility in digital mammography: results from a regional screening programme. *Eur Radiol*. 2017;27:3776-3787.
 52. Nelson JS, Wells JR, Baker JA, Samei E. How does c-view image quality compare with conventional 2D FFDM? *Med Phys*. 2016;43(5):2538-2547.
 53. Kamiya N, Hirata R, Shinohara N. Development of a quality control tool for phantoms of breast cancer screening using ImageJ. *Int J New Innov Eng Technol*. 2015;3:6-14.
 54. Barufaldi B, Lau KC, Schiabel H, Maidment D. Computational assessment of mammography accreditation phantom images and correlation with human observer analysis. In: Vol 9416. SPIE; 2015:15-24.
 55. Donini B, Rivetti S, Lanconelli N, Bertolini M. Free software for performing physical analysis of systems for digital radiography and mammography. *Med Phys*. 2014;41(5):051903.
 56. Perez-Ponce H, Daul C, Wolf D, Noel A. Computation of realistic virtual phantom images for an objective lesion detectability assessment in digital mammography. *Med Eng Phys*. 2011;33(10):1276-1286.
 57. Prieto G, Chevalier M, Guibelalde E. Automatic scoring of CDMAM using a model of the recognition threshold of the human visual system: R. In: IEEE; 2009:2489-2492.
 58. Mayo P, Rodenas F, Verdú G, Campayo J, Villaescusa J. Analysis of digital image quality indexes for CIRS SP01 and CDMAM 3.4 mammographic phantoms. In: IEEE; 2008:418-421.
 59. Gennaro G, Ferro F, Contento G, Fornasin F, Di Maggio C. Automated analysis of phantom images for the evaluation of long-term reproducibility in digital mammography. *Phys Med Biol*. 2007;52(5):1387.
 60. Asahara M, Kodera Y. Computerized scheme for evaluating mammographic phantom images. *Med Phys*. 2012;39(3):1609-1617.
 61. Alvarez M, Pina DR, Miranda JRA, Duarte SB. Application of wavelets to the evaluation of phantom images for mammography quality control. *Phys Med Biol*. 2012;57(21):7177. doi:10.1088/0031-9155/57/21/7177
 62. Lee Y, Tsai D, Shinohara N. Computerized quantitative evaluation of mammographic accreditation phantom images. *Med Phys*. 2010;37(12):6323-6331.
 63. Hamza A, Alport M, Rae W. Automatic scoring of the test objects in ACR (American College of Radiology) phantom images using image processing techniques. *J Sci Technol*. 2009;10:140-156.
 64. Patrocinio AC, Filha ER, Angelo MF, Schiabel H, Medeiros RB. Computational analysis of simulated images generated in public health centers in sao paulo-Brazil. In: Springer; 2009:181-184.
 65. Mayo P, Rodenas F, Verdú G, Villaescusa J, Campayo J. A comparative study of computer assisted assessment of image quality index for mammographic phantom images. *Radiat Prot Dosimetry*. 2005;116(1-4):620-623.

66. Mayo P, Rodenas F, Verdú G, Villaescusa JI, Campayo JM. Automatic evaluation of the image quality of a mammographic phantom. *Comput Methods Programs Biomed.* 2004;73(2):115-128.
67. Blot L, Davis A, Holubinka M, Martí R, Zwiggelaar R. Automated quality assurance applied to mammographic imaging. *EURASIP J Adv Signal Process.* 2002;2002:1-10.
68. Jansen JTM, Zoetlief J. Computer aided assessment of image quality for mammography using a contrast detail phantom. *Radiat Prot Dosimetry.* 2000;90(1-2):181-184.
69. Dougherty G. Computerized evaluation of mammographic image quality using phantom images. *Comput Med Imaging Graph.* 1998;22(5):365-373.
70. Castellano Smith A, Castellano Smith I, Dance D. Objective assessment of phantom image quality in mammography: a feasibility study. *Br J Radiol.* 1998;71(841):48-58.
71. Chakraborty DP, Eckert MP. Quantitative versus subjective evaluation of mammography accreditation phantom images. *Med Phys.* 1995;22(2):133-143.
72. Linares Salgado JE. Hacia una ética para el mundo tecnológico. *Artefactos.* 2018;7(1):99-120.
73. Ghamraoui B, Bader S, Thuerling T, Glick SJ. Classification of breast microcalcifications with GaAs photon-counting spectral mammography using an inverse problem approach. *Biomed Phys Eng Express.* 2023;9(3):035007. doi:10.1088/2057-1976/acb70f
74. Rangarajan K, Gupta A, Dasgupta S, et al. Ultra-high resolution, multi-scale, context-aware approach for detection of small cancers on mammography. *Sci Rep.* 2022;12(1):11622. doi:10.1038/s41598-022-15259-7
75. Vega MÁ, Mora LMQ, Badilla MVC. Inteligencia artificial y aprendizaje automático en medicina. *Rev Medica Sinerg.* 2020;5(8):e557-e557. doi:10.31434/rms.v5i8.557
76. Egan G, Keavey E, Phelan N. Comparison of contact spot imaging on a scanning mammography system to conventional geometric magnification imaging. In: Springer; 2012:165-172.
77. Koutaloni M, Delis H, Pascoal A, Spyrou G, Costaridou L, Panayiotakis G. Can electronic zoom replace magnification in mammography? A comparative Monte Carlo study. *Br J Radiol.* 2010;83(991):569-577. doi:10.1259/bjr/21753020
78. Hermann KP, Obenauer S, Funke M, Grabbe E. Magnification mammography: a comparison of full-field digital mammography and screen-film mammography for the detection of simulated small masses and microcalcifications. *Eur Radiol.* 2002;12(9):2188-2191. doi:10.1007/s00330-002-1356-8
79. Alkhalifah K, Brindhaban A, Asbeutah A. Comparison between image quality in electronic zoom and geometric magnification in digital mammography. *J X-Ray Sci Technol.* 2016;24(5):681-689.
80. Vahey K, Ryan E, McLean D, Poulos A, Rickard M. A comparison between the electronic magnification (EM) and true magnification (TM) of breast phantom images using a CDMAM phantom. *Eur J Radiol.* 2012;81(7):1514-1519. doi:10.1016/j.ejrad.2011.03.056
81. Quality of reporting noninferiority/similarity in research studies of diagnostic imaging - PubMed. Accessed January 14, 2024. <https://pubmed.ncbi.nlm.nih.gov/24029642/>
82. May LA, Guillerman RP, Zhang W, Orth RC. Unfounded conclusions of equivalence in diagnostic accuracy studies: a pervasive fallacy of inference in pediatric radiology scientific abstracts. *Pediatr Radiol.* 2018;48(13):1861-1866. doi:10.1007/s00247-018-4222-9
83. Ahn S, Park SH, Lee KH. How to Demonstrate Similarity by Using Noninferiority and Equivalence Statistical Testing in Radiology Research. *Radiology.* 2013;267(2):328-338. doi:10.1148/radiol.12120725
84. What is Reliability? Quality & Reliability Defined | ASQ. Accessed October 18, 2023. <https://asq.org/quality-resources/reliability>
85. Birolini A. *Reliability Engineering.* Vol 5. Springer; 2007.
86. Scientific Objectivity (Stanford Encyclopedia of Philosophy). Accessed January 14, 2024. <https://plato.stanford.edu/entries/scientific-objectivity/#ObjeFaitFact>

87. Tadrous PJ. On the concept of objectivity in digital image analysis in pathology. *Pathology (Phila)*. 2010;42(3):207-211. doi:10.3109/00313021003641758
88. Faj D, Posedel D, Stimac D, et al. Survey of mammography practice in Croatia: equipment performance, image quality and dose. *Radiat Prot Dosim*. 2008;131(4):535-540. doi:10.1093/rpd/ncn283
89. D. E. Souza Sabino SM, Silva TB, Watanabe AH, Syrjanen K, Carvalho AL, Mauad EC. Implementation of a clinical quality control program in a mammography screening service of Brazil. *Anticancer Res*. 2014;34(9):5057-5065.
90. Reis C, Pascoal A, Sakellaris T, Koutaloni M %J I into imaging. Quality assurance and quality control in mammography: a review of available guidance worldwide. 2013;4:539-553.
91. Kalyvas N, Chamogeorgaki A, Michail C, et al. A Novel Method to Model Image Creation Based on Mammographic Sensors Performance Parameters: A Theoretical Study. *Sens Basel*. 2023;23(4). doi:10.3390/s23042335
92. Alukic E, Homar K, Pavic M, Zibert J, Mekis N. The impact of subjective image quality evaluation in mammography. *Radiogr Lond*. 2023;29(3):526-532. doi:10.1016/j.radi.2023.02.025
93. Castellano Smith AD, Castellano Smith IA, Dance DR. Objective assessment of phantom image quality in mammography: a feasibility study. *Br J Radiol*. 1998;71(841):48-58. doi:10.1259/bjr.71.841.9534699
94. Sundell VM, Makela T, Meaney A, Kaasalainen T, Savolainen S. Automated daily quality control analysis for mammography in a multi-unit imaging center. *Acta Radiol*. 2019;60(2):140-148. doi:10.1177/0284185118776502
95. Gennaro G, Ferro F, Contento G, Fornasin F, di Maggio C. Automated analysis of phantom images for the evaluation of long-term reproducibility in digital mammography. *Phys Med Biol*. 2007;52(5):1387-1407. doi:10.1088/0031-9155/52/5/012
96. Galli V, Pini M, De Metrio D, de Bianchi PS, Bucchi L. An image quality review programme in a population-based mammography screening service. *J Med Radiat Sci*. 2021;68(3):253-259. doi:10.1002/jmrs.487
97. Reyes-Santias F, Cordova-Arevalo O, Dominguez IB, Antelo M. Factors influencing medical imaging technology uptake by private hospitals. *Health Technol*. 2023;13(6):937-945. doi:10.1007/s12553-023-00774-y
98. Khairat S, Marc D, Crosby W, Sanousi AA. Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis. *JMIR Med Inform*. 2018;6(2):e8912. doi:10.2196/medinform.8912
99. Greenwood BN, Agarwal R, Agarwal R, Gopal A. The When and Why of Abandonment: The Role of Organizational Differences In Medical Technology Life Cycles. *Manag Sci*. 2017;63(9):2948-2966. doi:10.1287/mnsc.2016.2490
100. Guidance W. Ethics and governance of artificial intelligence for health. *World Health Organ*. Published online 2021.
101. Barben D. Analyzing acceptance politics: Towards an epistemological shift in the public understanding of science and technology. *Public Underst Sci*. 2010;19(3):274-292. doi:10.1177/0963662509335459
102. Howard DH, Shen YC. *Comparative Effectiveness Research, COURAGE, and Technological Abandonment*. National Bureau of Economic Research, Inc; 2011. Accessed January 14, 2024. <https://econpapers.repec.org/paper/nbrnberwo/17371.htm>