

# La evaluación de la calidad de las imágenes médicas es más efectiva utilizando modelos de observador y objetos de prueba que con observadores reales e imágenes clínicas



**Gisella Gennaro**

Física médica/radiofísica,  
Departamento de Radiología  
del Instituto de Oncología (RCCS)  
de Padua (Italia)



**Wouter J.H. Veldkamp**

Físico médico/radiofísico,  
Departamento de Radiología  
del Hospital Universitario  
de Leiden (LUMC, Holanda)

En este debate intervienen, por orden alfabético, Gisella Gennaro y Wouter J.H. Veldkamp.

Gisella Gennaro trabaja como física médica/radiofísica en el Departamento de Radiología del Instituto de Oncología (RCCS) de Padua (Italia). Su investigación está centrada en mamografía digital (dosimetría, control de calidad, energía dual y tomosíntesis de mama). Es la investigadora principal de un ensayo clínico cuyo objetivo es comparar el rendimiento de la tomosíntesis digital de mama o *mamografía 3D* con el de la mamografía digital convencional. Es gestora del proyecto regional de control de calidad del programa de detección precoz de cáncer de mama e investigadora principal de un ensayo clínico para determinar la densidad de mama en mamografía.

Wouter J.H. Veldkamp estudió (1988-1994) Ingeniería Electrónica en la Universidad Tecnológica de Delft (Holanda). En 1995 comenzó sus estudios de doctorado en el Departamento de Radiología del Hospital de la Universidad de Radboud en Nijmegen (Holanda). Doctor en el año 2000 con una tesis doctoral sobre caracterización asistida por ordenador de microcalcificaciones en mamografía. Tras dos años de residencia en la misma institución en el período 1998-2000, obtuvo el título de físico médico/radiofísico. Desde el año 2002, el Dr. Veldkamp trabaja en el Departamento de Radiología del Hospital Universitario de Leiden (LUMC, Holanda). Es también asesor científico del centro holandés de formación y de expertos en *screening* de cáncer de mama (LRCB). Su investigación está centrada en el procesado de la imagen médica y estimación de la calidad de imagen en tomografía computarizada y radiografía convencional.

## Motivación de la controversia

Entre las actividades propias de los radiofísicos / físicos médicos la evaluación de la calidad de imagen constituye una de las tareas más críticas por las implicaciones que entraña para la salud de los pacientes. Habitualmente, esta tarea se lleva a cabo en las pruebas de aceptación y/o estado\* así como en los controles de calidad rutinarios. En las primeras, es necesario verificar si la calidad de la imagen proporcionada por

un nuevo sistema cumple las especificaciones exigidas. En las pruebas rutinarias hay que asegurar que la calidad se mantiene. Al margen de las pruebas habituales de control de calidad, la evaluación es también crucial en las investigaciones orientadas a optimizar el uso del sistema de imagen o a comparar distintos sistemas. Durante las pruebas de aceptación, los radiólogos valoran la calidad de imagen "clínica" y se suelen llevar a cabo ajustes para adaptarla a sus preferencias. En todos los casos se asume que el sistema o la tecnología asociada han sido probados previamente en estudios clínicos que garantizan que la imagen es adecuada.

Diferentes autores han publicado trabajos en los que se evidencia que, independientemente de su carácter subjetivo o cuantitativo, las medidas físicas o técnicas de la calidad no son siempre coincidentes con los resultados de los estudios clínicos. Existen factores tales como el ruido anatómico presente en las imágenes clínicas

\* Las pruebas de aceptación se realizan para asegurar entre otras la adecuación del equipo a las especificaciones del contrato de compra. Las pruebas de estado se realizan para establecer el estado de referencia inicial de acuerdo con las pruebas de aceptación y de aquellas pruebas adicionales que sean necesarias. En algunos países estas pruebas coinciden mientras que en otros están separadas y son realizadas por distinto personal.

Traducción: M. Chevalier

que influyen de forma determinante en la percepción de los signos clínicos en las radiografías. La relación entre la “calidad técnica de la imagen” (determinada por los radiofísicos) y la “calidad clínica de la imagen” (evaluada en estudios clínicos por radiólogos) es una cuestión que todavía no está completamente explicada. Un paso esencial para entender esta relación fue encontrar una definición adecuada para la calidad de imagen en el contexto médico. Robert F. Wagner (1938-2008) aborda en uno de sus primeros trabajos<sup>1</sup> la definición que ha logrado un mayor consenso en los siguientes términos:

*“The question of image quality has been an elusive one to define. One thing, however, is certain – that it must be defined in terms of the task the image is destined to perform.”*

Con esta frase, además de resaltar el concepto de calidad de imagen, se pretende rendir un pequeño homenaje a la figura de Robert F. Wagner, a quién debemos valiosas aportaciones en el campo de la imagen médica fruto de su gran inteligencia y de su brillante y persuasiva personalidad.

El concepto de calidad de imagen ligado a una tarea concreta es el punto de partida de las dos contribuciones de esta controversia. Las evaluaciones de la calidad de imagen más directamente vinculadas con este concepto son las basadas en imágenes clínicas (imágenes de pacientes) y observadores reales (radiólogos). Sin embargo, tal y como se refleja en el debate, la puesta en marcha de estas evaluaciones presenta problemas importantes, lo que ha originado la búsqueda de alternativas tales como los algoritmos de computación que simulan observadores reales (modelos de observador humano). La utilización de maniqués más realistas y representativos de la anatomía humana es la opción más contemplada para sustituir a los pacientes reales, posibilitando la adquisición de imágenes en muy diversas condiciones sin los riesgos asociados con la irradiación a pacientes. Tal y como se debate en la presente controversia, es importante tener presente en que medida los resultados asociados al uso de modelos de observador y maniqués en estudios de calidad de imagen pueden ser extrapolados a las situaciones reales.

## En contra de la proposición

**Gisella Gennaro**  
**Instituto Oncológico de Veneto (IRCCS)**  
**(Padua, Italia)**

“Calidad de imagen” es un término genérico aplicable a cualquier tipo de imagen y su medida debe referirse a la tarea o propósito con el que la imagen va a utilizarse. Por ejemplo, las imágenes fotográficas se hacen

generalmente por diversión y, como tales, el factor más valorado habitualmente es su apariencia. Por el contrario, las imágenes médicas son utilizadas por expertos y, para ellas, el aspecto fundamental a considerar es su adecuación a la tarea de toma de decisiones<sup>2</sup>. Por lo tanto, es preciso clarificar el propósito de las medidas de la calidad de imagen y el concepto de “efectividad” en este contexto antes de entrar en cualquier debate. Únicamente si se define la “tarea” asignada a la imagen es posible establecer la efectividad de un determinado método de evaluación de su calidad.

Existe consenso generalizado en que la finalidad de cualquier sistema de imagen en el ámbito médico es la obtención imágenes clínicas para posibilitar un diagnóstico correcto. En términos prácticos, sería facilitar a los expertos (radiólogos) la detección de la presencia/ ausencia de una patología o afección sobre la base de los hallazgos habidos en las imágenes. La única forma de evaluar la efectividad clínica de un sistema de imagen médico es realizar una “experiencia de laboratorio” que reproduzca el contexto clínico<sup>3,4</sup>. Esto en la práctica significa adquirir un número adecuado de imágenes clínicas para tener en cuenta la variabilidad natural de la población de pacientes (diferentes tipos de hallazgos más o menos “imperceptibles” a detectar en las imágenes), y un número apropiado de observadores para tener en cuenta las diferencias individuales de rendimiento en el proceso de toma de decisiones. Sin embargo, los experimentos clínicos requieren mucho tiempo para su desarrollo así como un diseño adecuado y cuidadoso que posibilite alcanzar conclusiones. De hecho, es muy habitual que un estudio del rendimiento clínico dure unos pocos años incluyendo el tiempo para obtener los permisos de las autoridades competentes (típicamente el Comité Ético), el tiempo necesario para la adquisición de las imágenes de los pacientes de acuerdo con unos criterios de inclusión/exclusión, el tiempo para establecer la “verdad” (obtener resultados histológicos y/o hacer un seguimiento) y el tiempo dedicado a la revisión de las imágenes, que requiere de la aplicación de reglas específicas para evitar sesgos que puedan falsear o incluso anular la validez de los resultados finales. Por estas razones, los estudios clínicos necesitan motivaciones importantes y deben estar bien fundamentados. En la actualidad, son los métodos más fiables para evaluar nuevas tecnologías de obtención de imágenes, pero no pueden ser empleados de forma habitual cada vez que un nuevo sistema de imagen es instalado en un hospital siempre que la tecnología asociada haya sido previamente evaluada con estudios clínicos.

En las actividades rutinarias, tales como los controles de calidad que se realizan tanto en las pruebas de aceptación o de estado como en el seguimiento de la estabilidad del equipamiento, la “calidad técnica de la imagen” puede ser evaluada mediante diferentes métodos. El principio general es muy simple: al

contrario que la “calidad clínica”, la “calidad técnica” es evaluada limitando las fuentes de variabilidad tanto como sea posible. Esto puede hacerse reemplazando: (a) la enorme variabilidad clínica con objetos de prueba reproducibles (habitualmente denominados “maniqués”) cuyo contenido es conocido; y (b) la variabilidad natural de los observadores humanos con herramientas informáticas diseñadas para analizar las imágenes de los maniqués y, o bien aplicar umbrales sobre la base de los modelos de observador humano, o bien proporcionar índices cuantitativos de la calidad de imagen. Estos métodos objetivos de evaluación de la calidad son muy prácticos y efectivos si lo que se desea es, por ejemplo, comparar distintos sistemas del mismo tipo. También son efectivos cuando se trata de hacer un seguimiento de la reproducibilidad con el tiempo para poner de relieve posibles variaciones “anormales” que pudieran asociarse con potenciales fallos de funcionamiento. No obstante, estos métodos se deben “manejar con cuidado” cuando las evaluaciones se hagan en términos absolutos, ya que pueden conducir a conclusiones/decisiones que no tengan nada (o muy poco) que ver con el objetivo clínico al estar basados en imágenes de maniqués. De hecho, por definición, los maniqués son objetos simples que incluyen detalles que reproducen de forma muy aproximada las lesiones clínicas de interés y están normalmente embebidos en un fondo uniforme. Los valores “límite” o tolerancias basados en estos maniqués pueden conducir a decisiones en relación con el rechazo o aceptación de un sistema de imagen que no tienen soporte clínico, aún cuando el análisis de la imagen se realice con herramientas informáticas basadas en modelos de observador. En otras palabras, podría ocurrir que el sistema X sea aceptado porque su “calidad técnica” está por encima de un umbral con un maniqué específico, mientras que el sistema Y es rechazado por estar por debajo a pesar de proporcionar una calidad clínica “equivalente” (como resultado, por ejemplo, de un estudio clínico). Asimismo, la condición de “fondo uniforme” común a la mayoría de los maniqués, que es provechosa ya que facilita la medida de “índices de detectabilidad” cuantitativos, está muy alejada de las imágenes clínicas en las que aparece siempre la estructura anatómica con un impacto muy significativo en la detectabilidad de los signos clínicos<sup>5,6</sup>. Finalmente, se asume de forma implícita que las distintas unidades de un mismo modelo de maniqué son perfectamente idénticas cuando los fabricantes no suelen aportar información sobre la reproducibilidad del proceso de fabricación. Por ello, los posibles valores límite (umbrales o tolerancias) deberían fijarse teniendo en cuenta las diferencias entre maniqués debidas al proceso de fabricación.

Otro aspecto que diferencia la “calidad técnica” de la “calidad clínica” es el post-procesado. De hecho, la “calidad técnica” es normalmente evaluada sobre imá-

genes sin procesar, relacionadas casi exclusivamente con la física de la formación de la imagen. Sin embargo, el papel del post-procesado en las imágenes digitales es enormemente relevante, al ser las imágenes post-procesadas las únicas examinadas por los radiólogos en la práctica clínica.

Volviendo a la proposición planteada en la controversia, más que argumentar en contra remarcaría una vez más que, en el caso específico de evaluación de la calidad de imagen, la efectividad de un método de medida concreto está inextricablemente vinculada a la razón por la que tal método es aplicado, es decir, al objetivo final de la medida. Aunque los estudios clínicos sean más efectivos que los estudios con maniqués para valorar el rendimiento clínico, no tiene ningún sentido planificar un estudio clínico cada vez que haya que aceptar un nuevo equipo: no es en absoluto viable por ser demasiado costoso y lento. Por el contrario, las pruebas de aceptación y, de forma general, los controles de calidad, son parte de una actividad habitual y, como tal, requieren métodos rigurosos y reproducibles y decisiones rápidas. En este sentido, resulta más efectivo determinar la “calidad técnica” a partir de imágenes de maniqués, bien extrayendo parámetros cuantitativos o bien aplicando modelos de observador humano, que faciliten la obtención de resultados objetivos en un tiempo adecuado<sup>7</sup>. La “calidad técnica” proporciona también buenos resultados cuando se trata de hacer comparaciones objetivas entre distintos sistemas de imagen<sup>8,9</sup>. Sin embargo, sus limitaciones, así como las inherentes de los objetos de prueba, deberían ser reconocidas con el fin de aplicar correctamente el método y tomar decisiones razonables. Sería un error “sobrestimar” los potenciales de la “calidad técnica” para predecir el “impacto clínico” de un determinado sistema de imagen. La “calidad técnica” es, sin duda, sensible para detectar diferencias entre distintos sistemas de imagen y puede considerarse como el primer paso para “clasificar” sus características funcionales. No obstante, se debe tener siempre presente que las conclusiones derivadas de las imágenes obtenidas con maniqués y los métodos de evaluación asociados no pueden transponerse al entorno clínico a pesar de ser razonable asumir la existencia de una cierta correlación entre la “calidad técnica” y la “clínica”; muy probablemente, las diferencias en la “calidad técnica” quedarían diluidas al comparar la “calidad clínica”. El concepto generalmente asumido por el que una nueva tecnología solo puede ser introducida en el ámbito clínico si proporciona al menos la misma calidad de imagen que las tecnologías existentes y probadas, puede aplicarse teniendo siempre presente las limitaciones de la “calidad técnica”. Los métodos de evaluación han de ser adecuados y los índices de calidad han de ser “equitativos” para evitar favorecer una determinada tecnología sobre el resto.

## A favor de la proposición

**Wouter J.H. Veldkamp**  
**Hospital Universitario de Leiden, LUMC**  
**(Leiden, Holanda)**

El diagnóstico por imagen es un elemento esencial de los sistemas sanitarios modernos que deben garantizar un nivel de calidad de la imagen que posibilite a los radiólogos identificar una patología con precisión. La evaluación de la calidad de imagen en el caso de las tecnologías de imagen médicas se basa generalmente en medidas técnicas (por ejemplo, funciones de transferencia) realizadas por físicos. La evaluación del funcionamiento clínico se basa por lo general en estudios clínicos (con pacientes reales o imágenes de maniqués antropomórficos) realizados por radiólogos.

En los servicios de radiología existen numerosas situaciones que requieren de la evaluación de la imagen, como son el caso de las pruebas para aprobar el uso clínico de un sistema o de las pruebas para el seguimiento de su estabilidad técnica. Estas pruebas son abordadas realizando fundamentalmente medidas objetivas de las características físicas. Este tipo de medidas muestran aspectos físicos de la calidad de la imagen tales como el contraste, la definición y el ruido. La ventaja de las medidas físicas radica en que están bien descritas, son reproducibles y son relativamente fáciles y rápidas de hacer. De forma alternativa, cuando se pretende optimizar el uso de un sistema de imagen para una finalidad clínica específica, se realizan estudios clínicos y con maniqués antropomórficos para analizar el rendimiento subjetivo (calificación) y objetivo (detección de la lesión) de los observadores. La realización de estudios clínicos con observadores humanos consume mucho tiempo y es costosa, ya que exige un número elevado de observaciones y un número significativo de observadores. Además, el número de posibles casos a investigar puede llegar a ser extremadamente grande.

La evaluación física de la calidad de imagen se basa con frecuencia en la utilización de maniqués homogéneos fáciles de adquirir y las medidas físicas se realizan típicamente sobre imágenes pre-procesadas. Existen varias medidas físicas a tener en cuenta. La eficiencia cuántica de detección (DQE) es una medida del efecto combinado del ruido y del rendimiento de contraste de un determinado sistema y es función de la frecuencia espacial. El ruido puede expresarse por la relación señal-ruido (RSR), la relación contraste ruido (RCR) o por el espectro de potencia de ruido (NPS). La capacidad de un sistema de imagen de transmitir el contraste asociado a un objeto en función de la frecuencia espacial es descrita tradicionalmente a través de la función de transferencia de modulación (MTF). La combinación de las funciones NPS y MTF determina la anteriormente mencionada DQE. Estas medidas

físicas objetivas describen la eficiencia de los sistemas o sus características funcionales técnicas, pero resulta complicado trasladar sus resultados al entorno clínico, que es más complejo que el descrito por esas medidas.

Los factores fundamentales de la “calidad técnica de imagen” son el contraste, la resolución y el ruido. Una prueba física semiobjetiva a menudo utilizada para determinar el efecto combinado de estos factores es la “detectabilidad a bajo contraste” (LCD); también denominado “detectabilidad contraste-detalle”. Esta prueba está más directamente relacionada con la tarea de detección clínica que las medidas físicas anteriores. La detectabilidad a bajo contraste se determina frecuentemente con observadores humanos que puntúan la visibilidad de objetos de bajo contraste sobre imágenes de maniqués que se fabrican con materiales homogéneos (por ejemplo, PMMA) en los que se insertan las señales de bajo contraste (véase la figura final). Con estos materiales se pretende simular la atenuación, filtración y radiación dispersa asociadas a un paciente. Los objetos de bajo contraste suelen distribuirse de acuerdo con un patrón fijo que el observador puede conocer de antemano (reduce la objetividad) lo que podría introducir un cierto sesgo en el método. Otros inconvenientes de este método son la variabilidad inter e intra-observador y su coste en tiempo. Para evaluar de forma más objetiva y eficiente la detectabilidad en imágenes médicas está muy extendido el uso de modelos de observador informáticos como alternativa a los observadores humanos (para más detalles véanse modelos de observador más abajo). Por ejemplo, para demostrar de forma más evidente el impacto de los parámetros radiográficos y de las nuevas tecnologías en la calidad de imágenes de tomografía computarizada (TC) hemos desarrollado un método automático y objetivo para investigar la detectabilidad a bajo contraste utilizando un modelo de observador<sup>10</sup> (véase la figura final).

Se debe tener cuidado con trasladar de forma directa al ámbito clínico los resultados de la detectabilidad a bajo contraste relativos a la combinación tamaño de objeto y contraste que sería visible en las imágenes clínicas. A diferencia de las lesiones reales, los objetos en los estudios LCD tienen con frecuencia forma cilíndrica con bordes muy definidos, a lo que hay que añadir que su apariencia en las imágenes finales homogéneas puede ser distinta a la de las lesiones en las imágenes de pacientes en las que el fondo anatómico es relevante<sup>11</sup>. Las medidas anteriores están por tanto destinadas a caracterizar el funcionamiento técnico del receptor de la imagen en combinación con el tubo de rayos X.

Como se ha apuntado antes, los estudios con observadores humanos pueden llegar a ser largos y costosos y el número de casos posibles a investigar puede ser grande. Como alternativa a los observadores humanos están los modelos informáticos de observador que predicen el rendimiento visual humano en imágenes

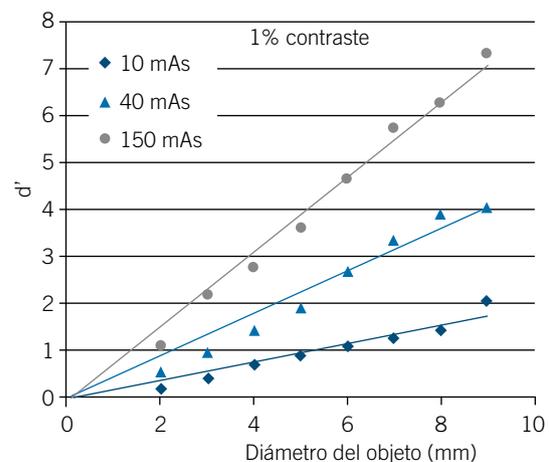
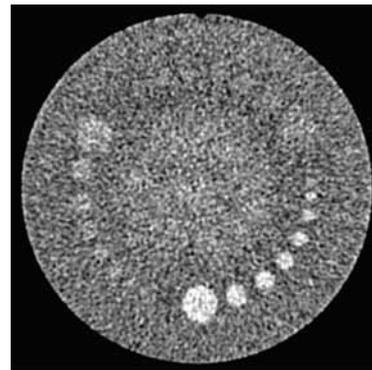
diagnósticas. Las posibles tareas a llevar a cabo por el modelo de observador se dividen de forma genérica en tareas de clasificación y tareas de estimación. Una muestra de tarea de clasificación en las aplicaciones médicas sería la detección de una lesión (por ejemplo se puede estimar la probabilidad de que un observador humano detecte un objeto) y una tarea de estimación podría ser la determinación de la función cardiaca<sup>12</sup>. Los modelos de observador pueden utilizarse para evaluar y optimizar los sistemas siempre que se asuma que el sistema mejor para el modelo de observador lo es también para el observador humano. Estos modelos parecen ser muy útiles para investigar las numerosas y diferentes condiciones de la imagen médica diagnóstica<sup>12</sup>. Más concretamente, está ampliamente aceptado que ciertos modelos de observador (por ejemplo, los modelos de Hotelling canalizado y NonPreWhitening incluyendo la respuesta del sistema visual humano (NPWE)) están bien correlacionados con el observador humano.

Es un principio generalmente aceptado que la calidad de imagen se define y mide de forma más apropiada en relación con la tarea a la que está destinada<sup>12</sup>. Por tanto la mejor manera, en teoría, de evaluar la calidad de la imagen médica debería ser medir el rendimiento clínico realizando estudios con observadores humanos. Sin embargo, estos estudios largos y costosos muy a menudo solo detectan grandes diferencias de calidad debido, entre otras razones, a la variabilidad inter e intra observador, así como a la variabilidad asociada a la relación casos-muestra<sup>12</sup>. Más aún, los aspectos de variabilidad obstaculizan la comparación de resultados de estudios con observadores humanos obtenidos en diferentes hospitales o con distintos sistemas.

Las pruebas de constancia (reproducibilidad) de los controles de calidad requieren métodos que no sean trabajosos ni caros. Deben ser además métodos sensibles para detectar cambios sutiles del sistema de formación de imagen y deben ser reproducibles. Para alcanzar estos objetivos es conveniente definir requisitos basados en medidas físico-técnicas (como las discutidas más arriba). Aunque los correspondientes resultados no puedan ser trasladados directamente a la clínica, para tener una buena "calidad clínica" es esencial que el equipo funcione correctamente. *No obstante, durante las pruebas de estado –justo después de la aceptación técnica del sistema– o justo después de pruebas de mantenimiento importantes, se debería llevar a cabo algún tipo de evaluación clínica adicional para comprobar la "calidad clínica"*.

En la actualidad existe mucho interés en el uso de maniqués antropomórficos que contengan estructuras anatómicas y objetos que simulen lesiones como alternativa a las evaluaciones clínicas con pacientes<sup>13</sup>. En este caso, los maniqués antropomórficos sustituyen al paciente mientras que los algoritmos informáticos de evaluación automática de la calidad de imagen sustituyen a los radiólogos. Los modelos de observador pueden utilizarse

para evaluar, por ejemplo, la detectabilidad de los objetos a partir de los datos asociados a las imágenes digitales del maniqué. La impresión 3D<sup>14</sup> es una novedosa tecnología que posibilita la fabricación de todo tipo de maniqués antropomórficos realistas así como las correspondientes lesiones. Esta línea de trabajo, unida a las medidas físicas objetivas asociadas con los maniqués estándar, posee en mi opinión, un potencial importante para evaluar la "calidad clínica de imagen". Al contrario de lo que ocurre con las medidas físicas, cuyo uso está esencialmente limitado a las imágenes preprocesadas (crudas), es posible en este caso tener en cuenta el procesamiento de la imagen de objetos que reproducen la anatomía humana. Los análisis pueden ser realizados eficientemente y ser repetidos con distintos sistemas de imagen, diferentes *software* de post-procesado o métodos de reconstrucción de la imagen. Además, en los estudios con maniqués antropomórficos se pueden adquirir múltiples imágenes modificando a voluntad los parámetros de adquisición, cosa que sería prohibitiva con pacientes reales.



Arriba, plano reconstruido de una imagen de TC del módulo de bajo contraste del maniqué Catphan 600 (con una calidad de imagen muy alta con el propósito de ilustrar). Abajo, se muestran los resultados (índice de detectabilidad  $d'$  vs. diámetro) de un método diseñado para investigar la detectabilidad de objetos de bajo contraste utilizando un modelo de observador<sup>10</sup>.

## Bibliografía

1. Wagner R, Weaver K. An assortment of image quality indices --Can they be resolved? En: Carson PL, Hendee WR, Zarnstorff WC, editors. *Application of Optical Instrumentation in Medicine I*. SPIE Proceedings 1972;35:83-94.
2. Burgess A. The Rose model, revisited. *J Opt Soc Am A* 1999;16:633-46.
3. Weinstein S, Obuchowski N, Lieber M. Clinical evaluation of diagnostic tests. *AJR* 2005;184:14-9.
4. Obuchowski N. How many observers are needed in clinical studies of medical imaging? *AJR* 2004;182:867-9.
5. Kotre C. The effect of background structure on the detection of low contrast objects in mammography. *BJR* 1998;71:1162-7.
6. Brettel D, Berry E, Smith M. The effect of experience on detectability in local area anatomical noise. *BJR* 2007;80:186-93.
7. Martin C, Sharp P, Sutton D. Measurement of image quality in diagnostic radiology. *Appl Radiat Isot* 1989;50:21-38.
8. Kehler M, Lyttkens K, Andersson B, Hochbergs P, Lindberg C, Medin J et al. Phantom study of chest radiography with storage phosphor, selenium, and film-screen systems. *Acta Radiol* 1996;37:332-6.
9. Hughes T, Celler A. A multivendor phantom study comparing the image quality produced from three state-of-the-art SPECT-CT systems. *Nucl Med Commun* 2012;33:663-70.
10. Hernandez-Girón I, Geleijns J, Calzado A, Veldkamp W. Automated assessment of low contrast sensitivity for CT systems using a model observer. *Med Phys* 2011;38 Suppl 1:S25.
11. Tapiovaara M. Relationships between physical measurements and user evaluation of image quality in medical radiology – a review. STUK-A219. Helsinki: Radiation and Nuclear Safety Authority 2006.
12. Barrett H, Yao J, Rolland J, Myers K. Model observers for assessment of image quality. *Proc Natl Acad Sci USA* 1993;90:9758-65.
13. Siewerdsen J, Fessler J, Myers K. Assessment of Image Quality in the New CT. AAPM 54<sup>th</sup> Annual meeting, July 29 – August 2, Charlotte, NC.
14. [http://en.wikipedia.org/wiki/3D\\_printing](http://en.wikipedia.org/wiki/3D_printing)