



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE MEDICINA  
UNIDAD DE POSGRADO E INVESTIGACIÓN

INSTITUTO MEXICANO DEL SEGURO SOCIAL  
UNIDAD MÉDICA DE ALTA ESPECIALIDAD  
HOSPITAL DE ESPECIALIDADES "DR. ANTONIO FRAGA MOURET"  
CENTRO MÉDICO NACIONAL "LA RAZA"  
DEPARTAMENTO DE ENDOCRINOLOGÍA

**"Variación interobservador y fuerza de concordancia entre  
diferentes departamentos médicos sobre nódulos tiroideos  
clasificados por ACR TI-RADS"**

**TESIS**

PARA OBTENER EL GRADO DE ESPECIALISTA EN  
ENDOCRINOLOGÍA

**PRESENTA**

DR. MARCOS DANIEL SÁNCHEZ GONZÁLEZ

**ASESOR**

DR. ALEJANDRO SOSA CABALLERO

CIUDAD DE MÉXICO, 2024





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



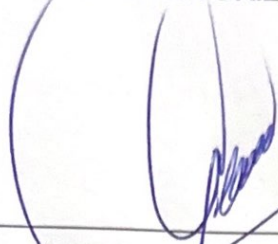
**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

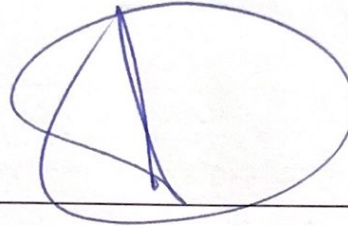
El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

HOJA DE AUTORIZACIÓN DE TESIS



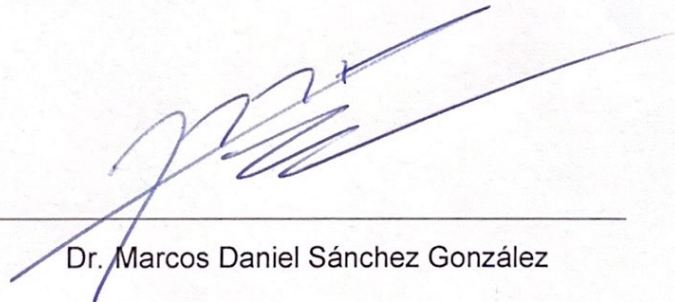
Dr. Andrés Muñoz Solís

Titular del curso universitario de  
Endocrinología



Dr. Alejandro Sosa Caballero

Asesor de tesis



Dr. Marcos Daniel Sánchez González

Médico residente de la especialidad de  
Endocrinología

No. Protocolo: R-2023-3501-057

## ÍNDICE

I.	RESUMEN.....	4
II.	INTRODUCCIÓN.....	6
III.	MATERIAL Y MÉTODOS.....	15
IV.	RESULTADOS.....	18
V.	DISCUSIÓN.....	23
VI.	CONCLUSIÓN .....	27
VII.	REFERENCIAS BIBLIOGRÁFICAS .....	28
VIII.	ANEXOS.....	31

## I. RESUMEN

**Título:** Variación interobservador y fuerza de concordancia entre diferentes departamentos médicos sobre nódulos tiroideos clasificados por ACR TI-RADS.

**Material y métodos:** Se realizó un estudio de concordancia externa donde participaron 2 médicos adscritos por cada uno de los siguientes departamentos del Hospital de Especialidades Centro Médico Nacional la Raza: Cirugía de Cabeza y Cuello, Endocrinología, Medicina Interna y Radiología Intervencionista. Se les mostraron 9 imágenes ultrasonográficas recabadas durante el periodo 2018-2023 de nódulos tiroideos previamente seleccionadas al azar por radiólogos externos a este protocolo, mismas que los distintos grupos clasificaron y valoraron si requirieron biopsia o no utilizando la escala ACR TI-RADS. Los datos revelados después del análisis estadístico fueron medidos con el coeficiente de Kappa de Cohen, donde se valoró la variación interobservador entre los diferentes departamentos a comparación con el estándar de Radiología.

**Resultados:** El departamento de Cirugía contra estándar radiológico obtuvo un kappa de BAAF insatisfactorio con 0.143, Endocrinología contra estándar radiológico obtuvo un kappa de BAAF insatisfactorio con 0.341, Medicina Interna contra estándar radiológico obtuvo un kappa de BAAF insatisfactorio con 0.341, Radiología Intervencionista contra estándar radiológico obtuvo un kappa de BAAF excelente con 0.769.

**Conclusión:** El avance de la tecnología radiológica ultrasonográfica, junto con su menor costo y mayor accesibilidad, ha incrementado la identificación de los nódulos tiroideos. Las escalas para clasificarlos por TIRADS son una herramienta útil, sin embargo con mucha variabilidad interobservador.

**Palabras clave DeCS:** Clasificación, Nódulo tiroideo, Ultrasonografía

## **ABSTRACT**

**Title:** Interobserver variation and strength of agreement between different medical departments of thyroid nodules classified by ACR TI-RADS scale.

**Material and methods:** An external concordance study was carried out where 2 doctors each of the following departments, of La Raza National Medical Center Specialty Hospital, participate: Head and Neck Surgery, Endocrinology, Internal Medicine and Interventional Radiology. We show them 9 ultrasonographic images collected during the period 2018-2023 of thyroid nodules previously and randomly selected by radiologists external to this protocol, where the different groups assessed whether they require a biopsy or not, applying ACR TI-RADS scale. Data concealment was revealed after statistical analysis using the Cohen's Kappa coefficient, where we compared the interobserver kappa of physicians from the same department and the interobserver kappa between different departments compared to the Radiology standard.

**Results:** The Department of Surgery against radiological standard obtained an unsatisfactory BAAF kappa with 0.143, the Endocrinology department against radiological standard obtained an unsatisfactory BAAF kappa with 0.341, the Internal Medicine department against radiological standard obtained an unsatisfactory BAAF kappa with 0.341, the Interventional Radiology department against radiological standard obtained an excellent BAAF kappa with 0.769.

**Conclusion:** The advancement of ultrasonographic radiological technology, together with its lower cost and greater accessibility, has increased the incidence of thyroid nodules. The scales to classify them by TIRADS are a useful tool, however with great interobserver variability.

**Key words MeSH:** Classification, Thyroid Nodule, Ultrasonography

## II. INTRODUCCIÓN

El parénquima tiroideo está compuesto por folículos, mismos que contienen una periferia de células foliculares y parafoliculares. Dentro de estos folículos se encuentra el coloide, que es rico en yodo, tiroglobulina y tirosinas que servirán para la formación de hormonas tiroideas por medio de la tiroperoxidasa. Tanto las células foliculares, adyuvantes productoras de hormonas tiroideas, como parafoliculares productoras de calcitonina, tienen riesgo de proliferación neoplásica.<sup>1,2</sup> Las moléculas producidas en estas células también sirven como marcadores de seguimiento en pacientes con neoplasias tiroideas, calcitonina útil en el seguimiento de cáncer de células parafoliculares y tiroglobulina (TG) junto con anticuerpos anti-tiroglobulina útiles en el seguimiento de cáncer de células foliculares. La incidencia de cáncer tiroideo continua en aumento a nivel mundial, en gran parte por el aumento en el uso de herramientas radiológicas, por lo que lesiones tiroideas en su mayoría son encontradas como incidentalomas.<sup>3</sup>

En las últimas dos décadas la incidencia ha aumentado >5% por año e incluso llegando a 14 casos por cada 100,000 personas, esto debido a la elevada sensibilidad ecográfica con respecto a la palpación clásica tiroidea. Tienen una mortalidad global de <2% a los 5 años, sin embargo se ha incrementado 1.1% al año desde 1994 hasta 2013.<sup>4</sup> Existen 4 clases principales de cáncer tiroideo: cáncer papilar de tiroides, cáncer folicular de tiroides, cáncer de células de Hürthle y cáncer poco diferenciado de tiroides. El 96% de estos serán foliculares, de los cuales 99% son cánceres diferenciados (papilares, foliculares, Hürthle, poco diferenciados), 3% cánceres medulares de células parafoliculares y 1% carcinomas anaplásicos de alto crecimiento. Generalmente los carcinomas tiroideos tienen excelente pronóstico, con una supervivencia a los 10 años de >95% de todos los casos, siendo los rangos de mortalidad en mujeres jóvenes los más bajos.<sup>5</sup>

Existen diversos mecanismos de señalización molecular involucrados en la fisiopatología neoplásica, la iniciación del cáncer tiroideo comienza por la vía

miogénica de tirosina cinasa (MAPK) que altera la expresión genética y promueve la proliferación celular, crecimiento celular, angiogénesis y pérdida de la diferenciación celular. La progresión del cáncer tiroideo esta mediado por la cinasa 3 fosfatidilinositol (PI3K) y la cinasa de mamíferos objetivo de Rapamicina (mTOR), estas alteran el promotor humano de telomerasa transcriptasa reversa (TERT) que daña el ADN y crea un arresto del ciclo celular.<sup>5,6</sup>

La mutación más frecuente en cánceres tiroideos no medulares es la BRAF V600E, misma que es exclusiva del cáncer papilar tiroideo. Las mutaciones RAS son más comúnmente encontradas en cáncer folicular tiroideo y las mutaciones que afecten PI3K o el gen supresor de tumores p53 son características de nódulos poco diferenciados y de carcinoma anaplásico tiroideo. Con respecto a neoplasias de células parafoliculares, la mutación del protooncogén RET se cree que es la causa fisiopatológica que explica el cáncer medular de tiroides. Dado que del 1 al 7% de los pacientes que presentan cáncer medular esporádico son portadores de protooncogén RET, se debe estudiar también heredabilidad del gen en busca de neoplasias endocrinas múltiples tipo 2 (NEM2) en familia de primer grado del paciente.<sup>7</sup>

El tratamiento es extremadamente variable, depende del tipo histológico, presencia de metástasis, tamaño del nódulo y edad del paciente. Incluso términos innovadores como teragnóstico utilizados por el Instituto Británico de Radiología nos hablan de herramientas que al mismo tiempo de ser diagnosticas, son terapéuticas y son aplicables en el cáncer tiroideo, como lo es el yodo 131.<sup>8</sup> En literatura previa se describe un tratamiento universal al encontrar nódulos tiroideos con alta probabilidad de malignidad y esta consistía en realizar una tiroidectomía total + aplicación de yodo radioactivo a dosis terapéuticas. Sin embargo, con estudios de morbimortalidad reportados en la actualidad, se ha demostrado que estos tratamientos agresivos no son para todos los tipos de cáncer, las opciones de manejo ahora incluyen vigilancia activa, radioablación, inhibidores de tirosina cinasa y cirugía. En un estudio de más de 5,000 pacientes encontraron que la supervivencia tanto de alto como de bajo riesgo después de una tiroidectomía total



era similar a la de los pacientes con hemitiroidectomía. En la actualidad contamos con distintas herramientas de estratificación de mortalidad posterior a la cirugía como el American Joint Committee on Cancer and Union Internationale Controle Cancer (AJCC-UICC) en su octava edición publicada en el 2017 y el riesgo de recurrencia dinámico de la American Thyroid Association (ATA), mismo que se tiene que realizar en cada consulta de seguimiento.<sup>9</sup>

Los nódulos tiroideos son una entidad clínica frecuente en áreas con deficiencia de yodo, en mujeres y en pacientes con antecedente de radiación de cuello. Estos pueden ser encontrados hasta en el 65% de individuos sanos, la mayoría son asintomáticos, incidentalmente descubiertos y de patología benigna, <5% son malignos y en raras ocasiones causan síntomas opresivos que requieren tratamiento quirúrgico. El riesgo de recurrencia y cáncer persistente varía de 7 a 28% según el riesgo de ATA, <5% en riesgo bajo, 5-20% en riesgo intermedio y >20% en riesgo alto. Los factores de riesgo que se asocian a una mayor probabilidad de malignidad incluyen la historia de radiación en cabeza y cuello, historia familiar de carcinoma tiroideo, obesidad, edad < 20 años o >60 años, sexo masculino, crecimiento rápido, consistencia firme, factores de riesgo ambientales y la presencia de ganglios linfáticos sospechosos a la exploración física y ultrasonido.<sup>10</sup>

Dado el avance de la tecnología radiológica ultrasonográfica junto con su menor costo y mayor accesibilidad, la identificación de estos ha aumentado. El ultrasonido ha sido un importante aliado en el estudio de nódulos tiroideos, nos ayuda a determinar tamaño, composición, ecogenicidad, forma, márgenes y focos ecogénicos. La característica más importante del ultrasonido (USG) es la capacidad que tiene de saber a qué nódulo se debe continuar estudiando y finalmente biopsiar.

Con respecto al tamaño, si bien no nos es útil para predecir malignidad, sabemos que tumores con crecimiento rápido tienen más probabilidad de ser cancerígenos. La ATA define como crecimiento nodular de relevancia a aquellos nódulos que aumentaron un 20% su diámetro con un mínimo de 2mm en 2 diámetros o un

incremento del 50% en su volumen. La composición es otro elemento que estudia el USG, el cáncer tiroideo es raro en nódulos quísticos, aproximadamente del 15 a 26% de cánceres tiroideos tienen componente quístico. De esta manera, entre más componente sólido tenga el nódulo, más probabilidad tiene de ser maligno. Otras características, como composición esponjiforme, nos hablan de benignidad con gran especificidad. Hablamos entonces de un nódulo sólido en aquel cuyo contenido quístico sea <10%, de uno predominantemente sólido en aquel cuyo contenido quístico sea >10 a 50%, y de uno quístico en aquel cuyo contenido quístico sea superior al 90%.<sup>10</sup>

Podemos clasificarlos también por su forma, donde un nódulo más alto que ancho tiene hasta un 90% de especificidad para el diagnóstico de malignidad. Esto es explicado por el patrón de crecimiento, los nódulos malignos crecen en modo centrifugo, mientras que los nódulos benignos crecen en paralelo y tienen forma redonda. Dentro de esta categoría podemos englobar el signo del halo, este halo consiste en una pseudocápsula formada por tejido fibroso, un halo uniforme que rodee por completo al nódulo tiene una alta especificidad de benignidad. Por el contrario, un halo incompleto e irregular se observa en cáncer papilar tiroideo.

La ecogenicidad es otro elemento esencial en el estudio del nódulo tiroideo, una marcada hipoecogenicidad nos habla hasta de un 92% de riesgo de malignidad. La ecogenicidad del nódulo se compara con la del parénquima tiroideo, y ya que la ecogenicidad tisular varía de persona en persona, otros tejidos de referencia que se pueden usar son los músculos esternocleidomastoideo y el esternohioideo, siendo estos característicos por ser muy hipoecogénicos.<sup>11</sup>

Por último, las calcificaciones se definen como focos ecogénicos, estas pueden o no tener sombra, es importante recalcar que la ausencia de sombra no descarta una calcificación. Algunos reportes mencionan que cualquier calcificación encontrada por estudio ultrasonográfico confiere más riesgo de malignidad. Las calcificaciones pueden estar presentes en nódulos benignos y malignos, se pueden clasificar según su tamaño. Conocemos como microcalcificaciones a aquellas que miden 1mm o menos, con o sin sombra acústica y

macrocalcificaciones a aquellas que miden >1mm. Según los reportes de histopatología, las microcalcificaciones corresponden a cuerpos de Psammoma, características por ser redondas, cristalinas, <1mm y específicas al carcinoma papilar tiroideo. Por el contrario, calcificaciones grandes e irregulares, pueden observarse también en nódulos benignos, incluso calcificaciones periféricas que asemejan a cascara de huevo también pueden ser encontradas en nódulos benignos. Sin embargo, calcificaciones periféricas en cascara de huevo que aparenten ruptura, se asocian con malignidad.<sup>11</sup>

De acuerdo a las características previamente descritas, además de servir para el estudio macroscópico del nódulo, sirven para conocer el valor predictivo positivo de malignidad y por lo tanto saber a cuáles tenemos que extender el estudio a uno microscópico, con biopsia por aspiración de aguja fina (BAAF). La BAAF es la principal herramienta diagnóstica para detectar la probabilidad de cáncer en nódulos tiroideos. La pregunta es ¿Qué nódulos se deben biopsiar?, para identificarlos, se deben de usar sistemas de clasificación ultrasonográfico, llamados Thyroid Imaging Reporting and Data Systems (TIRADS).

Existen varios tipos de TIRADS como Latvia TIRADS (L-TIRADS), European TIRADS (EU-TIRADS), Korean TIRADS (K-TIRADS) y el más utilizado a nivel mundial el American College of Radiology TIRADS (ACR-TIRADS). Siendo este el único instrumento de clasificación, a pesar de tener varias clases de TIRADS, cuenta con varias fallas y poca reproductibilidad, debido a diferentes equipos ultrasonográficos que se usan y la variabilidad inter e intra observador. Esto significa que un mismo nódulo tiroideo puede ser clasificado con diferente grado de TIRADS según el médico que lo vea, ¿Qué da esta variabilidad? Los años de experiencia, la especialidad médica, hospital de formación, tipo de TIRADS usado, calidad de imagen de ultrasonido, país de formación, e incluso un mismo médico puede clasificar con diferente TIRADS el mismo nódulo si se le vuelve a mostrar en un futuro.<sup>12</sup>

Lo que definirá que un nódulo requiera una biopsia o no, es su calificación TIRADS. Los grados van de TIRADS 1 a TIRADS 5 según evalúen composición,

ecogenicidad, forma, márgenes y focos ecogénicos. La composición da 0 puntos si es quístico o esponjiforme, 1 punto si es mixto y 2 puntos si es sólido. La ecogenicidad da 0 puntos si es anecoico, 1 punto si es isoecoico, 2 puntos si es hipoecoico, y 3 puntos si es muy hipoecoico. La forma da 0 puntos si es más ancho que alto y da 3 puntos si es más alto que ancho. Los márgenes dan 0 puntos si son bien delimitados y definidos, 2 puntos si son lobulados o irregulares y 3 puntos si tienen extensión extra tiroidea. Los focos ecogénicos dan 0 puntos si no tiene artefactos, 1 punto si tiene macrocalcificaciones, 2 puntos si tiene calcificaciones periféricas y 3 puntos si tiene microcalcificaciones puntiformes. La sumatoria total de las 5 categorías da la calificación TIRADS, si el nódulo sumo 0 puntos será TIRADS 1, si sumo 2 puntos será TIRADS 2, si sumo 3 puntos será TIRADS 3, si sumo de 4 a 6 puntos será TIRADS 4 y si sumo 7 puntos o más será TIRADS 5.<sup>12</sup>

Estos sistemas han mejorado a través del tiempo y se espera que los carcinomas tiroideos no estén infradiagnosticados ni sobre diagnosticados. Un falso negativo podría dejar a pacientes con carcinomas tiroideos sin BAAF, mientras que un falso positivo implicaría más gasto para el paciente o institución de salud al solicitar BAAF y reporte histopatológico de un nódulo que realmente no necesitaba más estudio.

No solo se necesita la calificación de TIRADS para saber si biopsiar o no un nódulo, se necesita también el tamaño. En general la ATA y la European Thyroid Association (ETA) recomiendan biopsiar aquellos pacientes con alto riesgo, que cuenten con historia familiar de cáncer tiroideo o que hayan estado expuestos a radiación. Nódulos con clasificación TIRADS 3 en adelante requieren BAAF para su estudio, es necesario un tamaño >25mm en TIRADS 3 para BAAF, un tamaño de >15mm en TIRADS 4 para BAAF, un tamaño de >10mm en TIRADS 5 para BAAF.<sup>13</sup>

Una de las metas más importantes es reducir el número de BAAF innecesarias, tanto por las complicaciones del procedimiento como por los costos. A pesar de tener un sistema de clasificación bien estructurado con alta sensibilidad, la

correlación entre TIRADAS y BETHESDA no tienen la mejor especificidad. En un estudio realizado en Latvia por Popova, N.M.; Radzina, M, incluyen a 274 pacientes con 289 nódulos tiroideos a quienes se les evaluó con las escalas TIRADS: EU-TIRADS, K-TIRADS, ACR-TIRADS, L-TIRADS. A todos los nódulos se les realizó BAAF y se dio un reporte de patología de acuerdo a BETHESDA. Concluyen que el utilizar diferentes tipos de TIRADS puede alterar el número real y necesario de biopsias de los nódulos estudiados, reportan una especificidad de 39.9% en EU-TIRADS, 46.6% en K-TIRADS, 41.1% en ACR-TIRADS y 72.7% en L-TIRADS.<sup>14</sup> La principal variable que afectaba las calificaciones de los nódulos era la interpretación de la hipogenicidad. Dentro de su mismo estudio asocian la hipogenicidad con la probabilidad de malignidad, de 81 nódulos isoecoicos, 2 (2.1%) eran malignos, de 151 nódulos hipoecoicos, 18 (12%) eran malignos y de 48 nódulos marcadamente hipoecoicos solo 16 (33%) eran malignos. Esto demuestra que se puede encontrar un nódulo con TIRADS bajo y BETHESDA alto, o por el contrario un nódulo con TIRADS alto que cuando se le realiza BAAF termina siendo benigno. Después de que se introdujo la escala TIRADS no hubo una tendencia significativa en la proporción de nódulos malignos diagnosticados por BAAF. Se ha calculado que el sobre diagnóstico o la identificación de cáncer, que de otro modo sería indolente, representa hasta el 77% de los casos.<sup>15</sup>

Por último la biopsia es reportada con el sistema BETHESDA, esta es una pieza clave en la evaluación del nódulo y es el más aceptado para categorizar el análisis citológico. Es una herramienta sencilla y reproducible para identificar el riesgo de malignidad, se reporta como categoría 1 no diagnóstica o muestra inadecuada con riesgo de malignidad de 1-4%, categoría 2 patología benigna con riesgo de malignidad de 0 a 3%, categoría 3 con riesgo de malignidad de 5 a 15%, categoría 4 con riesgo de malignidad de 15-30%, categoría 5 con riesgo de malignidad de 60 a 75% y categoría 6 con riesgo de malignidad de 97 a 99%. Aun y contando con sistemas como TIRADS y BETHESDA, la correlación de ambos con el diagnóstico final histopatológico no es del todo concordante, pudiendo encontrar TIRADS y BETHESDA altos con reporte histopatológico final de hiperplasia nodular.<sup>16</sup>

Respecto a lo anterior, investigadores, matemáticos y médicos han creado sistemas más eficientes que los ya mencionados. Incluso la notable llegada de la inteligencia artificial (AI) ha llevado a un increíble progreso en la identificación radiológica de nódulos tiroideos. Esta inteligencia usa algoritmos matemáticos para recrear habilidades cognitivas humanas con el fin de diagnosticar neoplasias y predecir respuestas a tratamiento.<sup>17</sup> Otra herramienta innovadora de diagnóstico son los marcadores moleculares, cuando la citología no es definitiva y cae en categorías BETHESDA 3 o 4 estos son de ayuda, los marcadores incluyen BRAF, RET, NTRK, ALK, MSI y dMMR.<sup>18</sup> Estos ayudan a la toma de decisiones a pesar del contexto clínico del paciente, sin embargo dentro de nuestro sistema económico institucional, estas nuevas herramientas están lejos de llegar para explotar su uso. La fuerza diagnóstica en nuestro medio continúa siendo el estudio radiológico con USG-TIRADS y BAAF-BETHESDA, por lo que es primordial conocer las carencias y fortalezas de nuestro sistema para identificación de cualquier cáncer tiroideo incluyendo los de alto grado.<sup>19,20</sup>

Los estudios de concordancia son estudios donde lo que se desea estimar es la probabilidad de desacuerdo entre los observadores al estudiar un fenómeno. El desacuerdo tiene un fuerte componente de variabilidad, por lo que su cálculo se encuentra basado en la estimación. El problema radica en encontrar el número exacto en el que la variabilidad pierda azar, ya que por el simple hecho de azar, 2 investigadores pueden coincidir en respuestas no porque ambos piensen igual, sino porque el resultado fue igual por mera coincidencia.<sup>21</sup> Este estudio trata de estimar el grado de desacuerdo y el poder de concordancia que tienen sus respuestas al evaluar nódulos tiroideos. Descartaremos el azar al enfrentarlos a diferentes nódulos con diferentes clasificaciones, esto nos permitirá conocer que tanto variaría el manejo de cada nódulo según el médico que lo observe.<sup>22</sup>

Ni el más perfecto de los diseños sería capaz de paliar el daño causado por un sistema de medida que cause azar, por lo que una importante fuente de error de medición es producto de la variabilidad entre observadores, cuya magnitud es posible de estimar a través de los llamados estudios de concordancia. Dos

aspectos distintos forman parte típicamente del estudio de concordancia: El sesgo entre observadores, que es la tendencia de un observador a dar consistentemente valores mayores o menores que otro observador y la concordancia entre observadores, que mide hasta punto los observadores coinciden en su medición. Las fuentes de error en la concordancia de ACR-TIRADS son la variabilidad de los observadores, la variabilidad por el instrumento de medida y la variabilidad en momentos diferentes del tiempo.<sup>23</sup> Por lo tanto es posible que dos o más observadores estén de acuerdo, sólo por efecto del azar. Bajo esta premisa, se han diseñado modelos estadísticos que estiman el grado de acuerdo existente después de retirar el efecto del azar.

El objetivo principal fue demostrar la variación interobservador externa, entre médicos de diferentes departamentos a comparación del estándar de Radiología, dentro del Hospital de Especialidades del Centro Médico Nacional La Raza involucrados en el manejo de nódulos tiroideos, utilizando la escala ACR TI-RADS para valorar necesidad de biopsia. Los objetivos secundarios fueron demostrar la variación interobservador externa, entre médicos de diferentes y mismo departamento a comparación del estándar de Radiología utilizando la escala ACR TI-RADS para clasificarlos.

La trascendencia de la investigación radica en advertir la concordancia de conocimientos entre diferentes departamentos del Hospital de Especialidades del Centro Médico Nacional La Raza. Múltiples patologías que deben ser atendidas por diferentes servicios médicos crean la necesidad de construir manejos conjuntos para alcanzar el mejor desenlace en el menor tiempo posible. Mientras algunos servicios diagnostican, otros tratan, sin embargo, la base patológica de la enfermedad debe ser entendida por ambas partes. El conocer la variabilidad interobservador no solo nos permite notar las áreas de oportunidad para capacitación médica, sino que también nos permite crear soluciones a problemas de comunicación y nos da herramientas para mejorar algoritmos internos de tratamiento que simplificarán el trabajo y reducirán así costos institucionales.

### III. MATERIAL Y MÉTODOS

El presente estudio cuenta con las siguientes características de diseño, analítico, transversal, observacional y prospectivo. Con respecto al grupo de estudio, los siguientes serán los criterios de selección. Inclusión: médicos de base con al menos 5 años de antigüedad adscritos a los departamentos de radiología intervencionista, cirugía de cabeza y cuello, medicina interna y endocrinología dentro del Hospital de Especialidades Centro Médico Nacional La Raza. No inclusión: médicos adscritos asesores de este protocolo de estudio, médicos adscritos con menos de 5 años de antigüedad en su respectivo servicio, médicos adscritos a distintos departamentos a los ya descritos. Eliminación: médicos adscritos previamente seleccionados quienes cambien de sede hospitalaria durante el transcurso del estudio, médicos adscritos que decidan abandonar el estudio.

Con respecto a las imágenes seleccionadas, los siguientes serán criterios de selección. Inclusión: imágenes de nódulos tiroideos realizadas en el Hospital de Especialidades Centro Médico Nacional La Raza, se seleccionarán por médicos radiólogos externos a este protocolo, 9 imágenes al azar de pacientes entre el periodo de 2018-2023 con diferentes clasificaciones TIRADS (I-V). Se incluirán imágenes de reportes de USG realizadas por médicos adscritos al departamento de radiología observadas por el mismo equipo ultrasonográfico y mismo transductor. No inclusión: no se incluirán imágenes de pacientes menores de 18 años, ni imágenes de hospitales externos a esta unidad médica, tampoco se tomarán en cuenta USG realizados por médicos residentes del departamento de radiología (año 1 a 4). Eliminación: imágenes previamente seleccionadas las cuales sean eliminadas de la plataforma hospitalaria HISWEB. Con una N final de: 70 observaciones (70 imágenes), mismas que serán mostradas a los 8 médicos de base adscritos a los distintos departamentos ya descritos. Con un total de 8.75 (9) imágenes por médico.



A cada uno de los 8 médicos de los diferentes servicios, (2) Cirugía de Cabeza y Cuello, (2) Endocrinología, (2) Medicina Interna y (2) Radiología Intervencionista, dentro del Hospital de Especialidades Centro Médico Nacional la Raza, se les mostraron 9 imágenes de nódulos tiroideos previamente seleccionadas al azar por radiólogos adscritos externos a este protocolo. La hoja de recolección de datos incluía la última versión del ACR-TIRADS donde los médicos pudieron asesorarse con la escala al evaluar los nódulos tiroideos. Se comparó cada médico de cada servicio con la BAAF y TIRADS control, así como también se compraron los médicos de cada servicio para conocer que tanto variaban sus respuestas entre ellos. Por lo tanto se comparó cirujano de cabeza y cuello 1 contra BAAF/TIRADS control, cirujano de cabeza y cuello 2 contra BAAF/TIRADS control, cirujano de cabeza y cuello 1 contra cirujano de cabeza y cuello 2. Endocrinólogo 1 contra BAAF/TIRADS control, Endocrinólogo 2 contra BAAF/TIRADS control, Endocrinólogo 1 contra Endocrinólogo 2. Médico internista 1 contra BAAF/TIRADS control, Médico internista 2 contra BAAF/TIRADS control, Médico internista 1 contra Médico internista 2. Radiólogo intervencionista 1 contra BAAF/TIRADS control, Radiólogo intervencionista 2 contra BAAF/TIRADS control, Radiólogo intervencionista 1 contra Radiólogo intervencionista 2.

### **Análisis estadístico**

El análisis estadístico de las proporciones de acuerdo observados sobre la proporción de acuerdos esperados por azar se realizó mediante el programa IBM SPSS en su última versión (v27.0), aplicando el coeficiente de variación de kappa como instrumento de medida estadístico, se aplicaron como variables: BAAF control (BAAFC), BAAF cirujano de cabeza y cuello 1 (BAAFCX1), BAAF cirujano de cabeza y cuello 2 (BAAFCX2), BAAF endocrinólogo 1 (BAAFEN1), BAAF endocrinólogo 2 (BAAFEN2), BAAF médico internista 1 (BAAFMI1), BAAF médico internista 2 (BAAFMI2), BAAF radiólogo intervencionista 1 (BAAFRX1), BAAF radiólogo intervencionista 2 (BAAFRX2), TIRADS control (TIRADSC), TIRADS cirujano de cabeza y cuello 1 (TIRADSCX1), TIRADS cirujano de cabeza y cuello

2 (TIRADSCX2), TIRADS endocrinólogo 1 (TIRADSEN1), TIRADS endocrinólogo 2 (TIRADSEN2), TIRADS médico internista 1 (TIRADSMI1), TIRADS médico internista 2 (TIRADSMI2), TIRADS radiólogo intervencionista 1 (TIRADSRX1), TIRADS radiólogo intervencionista 2 (TIRADSRX2).

La variable de control siempre se colocó en columna dentro de la tabla cruzada, mientras que la respuesta de cada médico se colocó en filas dentro de la tabla cruzada. Siendo todas de tipo numérico, con anchura de 1, decimales de 0, con medida nominal y rol de entrada.

Con respecto a las BAAF, se consideró el numero 1 como indicación de realizar BAAF y 0 como no tener indicación de realizar BAAF. Con respecto a la calificación ACR-TIRADS, se utilizó la escala numérica del 1 al 5, según el puntaje que cada médico le diera al nódulo tiroideo.

Dentro de los anexos se encuentran las tablas de medidas simétricas, con el valor de coeficiente de kappa, error estándar asintótico y significación aproximada de cada tabla cruzada que calculó la proporción de acuerdos estimados sobre los acuerdos por azar.

#### IV. RESULTADOS

Se consideró como una excelente fuerza de concordancia si  $0.76 \leq \text{kappa} \leq 1$ , una satisfactoria fuerza de concordancia si  $0.41 \leq \text{kappa} \leq 0.75$  y una insatisfactoria fuerza de concordancia si  $-1 \leq \text{kappa} \leq 0.40$ .

Dentro del departamento de Cirugía de cabeza y cuello, la concordancia de BAAF entre BAAFC y BAAFCX1 fue insatisfactoria con kappa de 0.143.

**Tabla cruzada BAAFCX1\*BAAFC**

Recuento

		BAAFC		Total
		0	1	
BAAFCX1	0	2	1	3
	1	3	3	6
Total		5	4	9

La concordancia por BAAF entre BAAFC y BAAFCX2 fue insatisfactoria con kappa de 0.143.

**Tabla cruzada BAAFCX2\*BAAFC**

Recuento

		BAAFC		Total
		0	1	
BAAFCX2	0	2	1	3
	1	3	3	6
Total		5	4	9

La concordancia por BAAF entre BAAFCX1 y BAAFCX2 fue satisfactoria con kappa de 0.500.

**Tabla cruzada BAAFCX2\*BAAFCX1**

Recuento		BAAFCX1		Total
		0	1	
BAAFCX2	0	2	1	3
	1	1	5	6
Total		3	6	9

La concordancia de TIRADS entre TIRADSC y TIRADSCX1 fue insatisfactoria con kappa de 0.182. La concordancia de TIRADS entre TIRADSC y TIRADSCX2 fue insatisfactoria con kappa de 0.274. La concordancia de TIRADS entre TIRADSCX1 y TIRADSCX2 fue insatisfactoria con kappa de 0.045.

Dentro del departamento de Endocrinología, la concordancia de BAAF entre BAAFC y BAAFEN1 fue insatisfactoria con kappa de 0.143.

**Tabla cruzada BAAFEN1\*BAAFC**

Recuento		BAAFC		Total
		0	1	
BAAFEN1	0	2	1	3
	1	3	3	6
Total		5	4	9

La concordancia por BAAF entre BAAFC y BAAFEN2 fue insatisfactoria con kappa de 0.341.

**Tabla cruzada BAAFEN2\*BAAFC**

Recuento		BAAFC		Total
		0	1	
BAAFEN2	0	3	1	4
	1	2	3	5
Total		5	4	9

La concordancia por BAAF entre BAAFEN1 y BAAFEN2 fue excelente con kappa de 0.769.

**Tabla cruzada BAAFEN2\*BAAFEN1**

Recuento

		BAAFEN1		Total
		0	1	
BAAFEN2	0	3	1	4
	1	0	5	5
Total		3	6	9

La concordancia por TIRADS entre TIRADSC y TIRADSEN1 fue insatisfactoria con kappa de 0.156. La concordancia por TIRADS entre TIRADSC y TIRADSEN2 fue insatisfactoria con kappa de -0.143. La concordancia por TIRADS entre TIRADSEN1 y TIRADSEN2 fue insatisfactoria con kappa de 0.143.

Dentro del departamento de Medicina interna, la concordancia de BAAF entre BAAFC y BAAFMI1 fue insatisfactoria con kappa de -0.154.

**Tabla cruzada BAAFMI1\*BAAFC**

Recuento

		BAAFC		Total
		0	1	
BAAFMI1	0	3	3	6
	1	2	1	3
Total		5	4	9

La concordancia por BAAF entre BAAFC y BAAFMI2 fue insatisfactoria con kappa de 0.341.

**Tabla cruzada BAAFMI2\*BAAFC**

Recuento

		BAAFC		Total
		0	1	
BAAFMI2	0	3	1	4
	1	2	3	5
Total		5	4	9

La concordancia por BAAF entre BAAFMI1 y BAAFMI2 fue satisfactoria con kappa de 0.571.

**Tabla cruzada BAAFMI2\*BAAFMI1**

Recuento

		BAAFMI1		Total
		0	1	
BAAFMI2	0	4	0	4
	1	2	3	5
Total		6	3	9

La concordancia por TIRADS entre TIRADSC y TIRADSMI1 fue insatisfactoria con kappa de 0.156. La concordancia por TIRADS entre TIRADSC y TIRADSMI2 fue insatisfactoria con kappa de 0.156. La concordancia por TIRADS entre TIRADSMI1 y TIRADSMI2 fue insatisfactoria con kappa de 0.308.

Dentro del departamento de Radiología intervencionista, la concordancia de BAAF entre BAAFC y BAAFRX1 fue excelente con kappa de 0.769.

**Tabla cruzada BAAFRX1\*BAAFC**

Recuento

		BAAFC		Total
		0	1	
BAAFRX1	0	5	1	6
	1	0	3	3
Total		5	4	9

La concordancia por BAAF entre BAAFC y BAAFRX2 fue perfecta con kappa de 0.571.

**Tabla cruzada BAAFRX2\*BAAFC**

Recuento

		BAAFC		Total
		0	1	
BAAFRX2	0	5	0	5
	1	0	4	4
Total		5	4	9

La concordancia por BAAF entre BAAFRX1 y BAAFRX2 fue excelente con kappa de 0.769.

**Tabla cruzada BAAFRX2\*BAAFRX1**

Recuento

		BAAFRX1		Total
		0	1	
BAAFRX2	0	5	0	5
	1	1	3	4
Total		6	3	9

La concordancia por TIRADS entre TIRADSC y TIRADSRX1 fue insatisfactoria con kappa de 0.286. La concordancia por TIRADS entre TIRADSC y TIRADSRX2 fue insatisfactoria con kappa de 0.016. La concordancia por TIRADS entre TIRADSRX1 y TIRADSRX2 fue satisfactoria con kappa de 0.410.

## V. DISCUSIÓN

El análisis estadístico nos permitió conocer la variación entre observaciones de BAAF y TIRADS entre los diferentes servicios médicos. Logramos comparar las observaciones entre médicos del mismo servicio, así como también con el control de radiología. La gran mayoría de las concordancias fueron insatisfactorias, el único servicio que logró concordancias excelentes fue Radiología intervencionista. De igual forma, el único servicio que logro una concordancia perfecta de 100% fue Radiología intervencionista. El resto de los servicios, Cirugía de cabeza y cuello, Medicina interna y Endocrinología no lograron tener alguna concordancia satisfactoria, excelente o perfecta en comparación con el estándar.

De los 9 nódulos tiroideos que se les mostraron, solo 4 requerían biopsias, 3 nódulos tenían categoría TIRADS 5, 1 nódulo tenía categoría TIRADS 4, 2 nódulos tenían categoría TIRADS 3, 2 nódulos tenían categoría TIRADS 2 y 1 nódulo tenía categoría TIRADS 1.

Cirugía de cabeza y cuello logró una concordancia satisfactoria de 0.500 entre ellos (BAAFCX1 y BAAFCX2), sin embargo cada cirujano por separado en comparación con la BAAFC estándar, tuvieron concordancias insatisfactorias con kappa de 0.143. Por lo tanto se concluye que ambos cirujanos coincidieron en sus observaciones de BAAF, pensaron igual, no obstante a comparación del estándar, tuvieron respuestas no favorables. Este servicio realizo en promedio 25% más biopsias de las que el estándar marcaba. Con respecto a la calificación TIRADS otorgada a cada nódulo, todas las respuestas fueron insatisfactorias, TIRADSC y TIRADSCX1 kappa 0.182, TIRADSC y TIRADSCX2 kappa 0.274, TIRADSCX1 y TIRADSCX2 kappa 0.045. El servicio de Cirugía de cabeza y cuello no calificó satisfactoriamente ningún nódulo tiroideo, las calificaciones que hayan coincidido a comparación del estándar, según el coeficiente de kappa, fue por simple azar.

El servicio de Endocrinología, así como el de Cirugía de cabeza y cuello, lograron en BAAF una concordancia excelente con kappa de 0.769 entre ellos (BAAFEN1 y



BAAFEN2). Significa que a experiencia de los endocrinólogos, pensaron igual en cuales biopsiar y cuales no biopsiar. Sin embargo al compararlos con el estándar de radiología, todas las observaciones fueron insatisfactorias, BAAFC y BAAFEN1 con kappa de 0.143, BAAFC y BAAFEN2 con kappa de 0.341. Este servicio realizó en promedio 50% más biopsias de las que se debieron haber tomado según el estándar. Con respecto a la calificación TIRADS otorgada a cada nódulo, todas las respuestas fueron insatisfactorias, TIRADSC y TIRADSEN1 kappa 0.156, TIRADSC y TIRADSEN2 kappa -0.143, TIRADSEN1 y TIRADSEN2 kappa 0.143. Las calificaciones que hayan coincidido a comparación del estándar, según el coeficiente de kappa, fueron también por efecto de azar.

Medicina interna, así como Endocrinología y Cirugía de cabeza y cuello, lograron en BAAF una concordancia satisfactoria con kappa de 0.571 entre ellos (BAAFMI1 y BAAFMI2). Por separado, cada uno de los médicos internistas tuvo respuestas insatisfactorias a comparación del control, BAAFC y BAAFMI1 con kappa de -0.154, BAAFC y BAAFMI2 con kappa de 0.341. Este servicio realizó un 25% menos biopsias de las que el control marcaba realizar. Con respecto a la calificación TIRADS otorgada a cada nódulo, todas las respuestas fueron no concordantes, por lo que ni entre ellos, ni en comparación con el control, las respuestas coincidieron. TIRADSC y TIRADSMI1 con kappa 0.156, TIRADSC y TIRADSMI2 con kappa 0.156, TIRADSMI1 y TIRADSMI2 con kappa 0.308.

El último servicio analizado fue Radiología intervencionista, que como se mencionó previamente, fue el único servicio que obtuvo concordancias excelentes e incluso una perfecta. Entre los 2 radiólogos, lograron una concordancia de BAAF excelente BAAFRX1 y BAAFRX2 con kappa de 0.769, que significa que sus respuestas fueron concordantes entre ellas. Al momento de comparar sus respuestas contra el control, también se logró una excelente concordancia con el radiólogo 1, BAAFC y BAAFRX1 con kappa de 0.769 y una concordancia perfecta del radiólogo 2 contra control, BAAFC y BAAFRX2 con kappa de 1.000. Estas pruebas indican que no solo los radiólogos pensaron igual entre ellos, sino que también sus respuestas coincidieron con el control de BAAF. A pesar de tener

estas nulas o pocas variaciones interobservadores en cuanto al análisis de BAAF, el análisis de TIRADS fue como en los servicios previos. El radiólogo 1 y radiólogo 2 coincidieron en sus respuestas de TIRADSRX1 y TIRADSRX2 con kappa satisfactorio de 0.410, sin embargo al comparar contra control, TIRADSC y TIRADSRX1 con kappa 0.286, TIRADSC y TIRADSRX2 con kappa de 0.016. Este servicio no tuvo variación interobservador en BAAF pero si al momento de clasificar nódulos por TIRADS, lo que significa que biopsiaron los nódulos que se tenían que biopsiar, sin embargo no coincidieron entre ellos ni contra control con la clasificación otorgada a cada nódulo por ACR-TIRADS.

Al confrontar los hallazgos de este estudio con los de la literatura actual, se observa que existen diferentes sensibilidades y especificidades descritas sobre los diferentes modelos de TIRADS, sin embargo la literatura no es clara en cuanto al uso universal de un TIRADS en particular.<sup>16</sup> Incluso, como se comentó antes, cada TIRADS reporta una especificidad diferente, 39.9% en EU-TIRADS, 46.6% en K-TIRADS, 41.1% en ACR-TIRADS y 72.7% en L-TIRADS.<sup>14</sup> El factor determinante en la variación de resultados es la subjetividad del médico que lo utiliza, no importando que se use la misma escala de TIRADS, que se use el mismo equipo ultrasonográfico, e incluso siendo de una misma especialidad médica, las observaciones varían entre ellas. Por lo tanto el médico con más experiencia en evaluar nódulos tiroideos deberá ser el encargado de determinar que nódulo continua con vigilancia activa y que nódulo se debe biopsiar.

Este estudio difiere con las especificidades reportadas en literatura Mexicana, el único artículo que habla de variabilidad interobservador por TIRADS en México es el publicado por Murrieta E., Vázquez J., et.al. Donde reportan coeficientes de kappa de 0.955 en TIRADS y 0.892 para decisión de BAAF.<sup>21</sup> Cabe resaltar que el estudio solo se hizo entre radiólogos, por lo que las observaciones fueron más homogéneas, a diferencia de este estudio que involucro distintos departamentos médicos y las observaciones fueron más heterogéneas.

Con respecto a objetivos secundarios, se revisó también la concordancia entre los diferentes servicios según la calificación de ACR-TIRADS; fue insatisfactoria o no

concordante las observaciones entre el servicio de Cirugía y el control radiológico, fue insatisfactoria o no concordante las observaciones entre el servicio de Medicina interna y el control radiológico, fue insatisfactoria o no concordante las observaciones entre el servicio de Medicina Interna y el control radiológico, e incluso fue insatisfactoria o no concordante las observaciones entre el servicio de Radiología Intervencionista y el control radiológico.

Este estudio es el único que se ha realizado siguiendo la metodología previamente descrita, no existe ningún otro reportado en México, Latinoamérica ni Sociedades internacionales endocrinológicas o radiológicas que evalué la variación interobservador entre diferentes departamentos médicos. Esta puede ser la razón principal por la que en guías internacionales reporten especificidades tan altas, estas escalas solo se han puesto a prueba y evaluado entre radiólogos, pero no entre más especialidades médicas que estén involucradas en el manejo de nódulos tiroideos.<sup>18,19</sup>

## VI. CONCLUSIÓN

Dado el avance de la tecnología radiológica ultrasonográfica, junto con su menor costo y mayor accesibilidad, la identificación de los nódulos tiroideos ha aumentado. El USG entonces ha sido un importante aliado en el estudio de nódulos tiroideos. La característica más importante, es la capacidad de saber a qué nódulo se debe continuar estudiando y finalmente biopsiar.

En el caso de este estudio, los radiólogos intervencionistas fueron los que menos variación tuvieron entre ellos y más se acercaron al estándar control.

Fue insatisfactoria o no concordante las observaciones entre el servicio de Cirugía y el control radiológico, fue insatisfactoria o no concordante las observaciones entre el servicio de Medicina interna y el control radiológico, fue satisfactoria o concordante las observaciones entre el servicio de Radiología intervencionista y el control radiológico y por último no se logró comprobar la hipótesis de que las observaciones entre el servicio de Endocrinología y el control radiológico serían concordantes. El único servicio que logró concordancia contra el control radiológico fue Radiología intervencionista.

La correcta evaluación del nódulo tiroideo contribuye a evitar saturación de servicios, disminuir gastos y riesgos de los pacientes. Como se observó en el análisis estadístico, a pesar de que entre médicos del mismo servicio pensarán igual y tuvieran observaciones con poca variabilidad, no necesariamente significó que eran las respuestas correctas al compararlas con el estándar radiológico. A falta de evidencia, se requieren estudios multicéntricos que comprueben lo que se demostró en este estudio, así como más ensayos que demuestren la variabilidad que existe al usar estas escalas. Los estudios donde se basan para dar sensibilidades y especificidades de los TIRADS son realizados en centros especializados con ambientes controlados, se requiere que cada hospital y servicio reconozca las áreas de oportunidad en capacitación para el uso de la herramienta TIRADS y así concordar con el estándar.

## VII. REFERENCIAS BIBLIOGRÁFICAS

1. Rago, T.; Vitti, P. Risk, Stratification of Thyroid Nodules: From Ultrasound Features to TIRADS. *Cancers* 2022, 14, 717. <https://doi.org/10.3390/cancers14030717>.
2. Wells S, Asa S, et.al, Revised American Thyroid Association Guidelines for the Management of Medullary Thyroid Carcinoma, The American Thyroid Association Guidelines Task Force on Medullary Thyroid Carcinoma. *Thyroid*, 2015, 25:6(567-610). DOI: 10.1089/thy.2014.0335.
3. Xu J, Bergren R, et.al, Thyroglobulin Antibody Resolution after Total Thyroidectomy for Cancer, *J Surg Res*. 2015 October; 198(2): 366–370. doi: 10.1016/j.jss.2015.03.094.
4. Cabanillas M, McFadden D, Durante C, Thyroid cáncer, *Lancet* 2016; 388: 2783–95. [http://dx.doi.org/10.1016/S0140-6736\(16\)30172-6](http://dx.doi.org/10.1016/S0140-6736(16)30172-6).
5. Livia Lamartina, Giorgio Grani, Cosimo Durante, Isabelle Borget, Sebastiano Filetti, Follow-up of differentiated thyroid cancer – what should (and what should not) be done, *Nature*, 2018. <https://doi.org/10.1038/s41574-018-0068-3>.
6. Haigh PI, Urbach DR, Rotstein LE. Extent of thyroidectomy is not a major determinant of survival in low- or high-risk papillary thyroid cancer. *Ann Surg Oncol* 2005; 12:81–89.
7. Fallahi P, Ferrari S, et.al, Molecular targets of tyrosine kinase inhibitors in thyroid cancer, *Seminars in Cancer Biology* 79 (2022) 180–196. <https://doi.org/10.1016/j.semcancer.2020.11.013>.
8. Choudhury PS, Gupta M. Differentiated thyroid cancer theranostics: radioiodine and beyond. *Br J Radiol* 2018; 91: 20180136. <https://doi.org/10.1259/bjr.20180136>.
9. Robert I. Haddad, MD1; Lindsay Bischoff, MD2; Douglas Ball, Thyroid Carcinoma, Version 2.2022, Thyroid Carcinoma, Version 2.2022, *J Natl Compr Canc Netw* 2022;20(8):925–951. doi: 10.6004/jnccn.2022.0040.

10. Popova, N.M.; Radzina, M.; Prieditis, P.; Liepa, M.; Rauda, M.; Stepanovs, K. Impact of the Hypoechoogenicity Criteria on Thyroid Nodule Malignancy Risk Stratification Performance by Different TIRADS Systems. *Cancers* 2021, 13, 5581. <https://doi.org/10.3390/cancers13215581>.
11. Pappa T, Ahmadi S, et.al, Thyroid Nodule Shape Independently Predicts Risk of Malignancy. *The Journal of Clinical Endocrinology & Metabolism*, 2022, 107, 1865–1870. <https://doi.org/10.1210/clinem/dgac246>.
12. Schlumberger M, and Leboulleux S, Current practice in patients with differentiated thyroid cancer, Springer Nature Limited, *Nature Reviews Endocrinology*, 2020. <https://doi.org/10.1038/s41574-020-00448-z>.
13. Lamartina L, Grani G, Follow-up of differentiated thyroid cancer – what should (and what should not) be done, Springer Nature Limited, *Nature Reviews Endocrinology*, 2018. <https://doi.org/10.1038/s41574-018-0068-3>.
14. Hawkins S, Jamieson S, the global epidemic of thyroid cancer overdiagnosis illustrated using 18 months of consecutive nodule biopsy correlating clinical priority, ACR-TIRADS and Bethesda scoring. 2021 Jun;65(3):309-316. doi: 10.1111/1754-9485.1316.
15. Mora I, Muñoz JL, et.al, Rendimiento del sistema Bethesda en el diagnóstico citopatológico del nódulo tiroideo, *CIRESP* 2018;96(6):363–368. DOI: 10.1016/j.ciresp.2018.02.017.
16. Tessler FN, Middleton WD, Grant EG. Thyroid Imaging Reporting and Data System (TI-RADS): A User's Guide. *Radiology*. 2018 Apr;287(1):29-36. Erratum in: *Radiology*. 2018 Jun;287(3):1082. doi: 10.1148/radiol.2017171240.
17. Bini, F.; Pica, A.; Azzimonti, L.; Giusti, A.; Ruinelli, L.; Marinozzi, F.; Trimboli, P. Artificial Intelligence in Thyroid Field—A Comprehensive Review. *Cancers* 2021, 13, 4740. <https://doi.org/10.3390/cancers13194740>.
18. Atil Y, Bustamante M, Genomic Profiling of Thyroid Nodules: Current Role for ThyroSeqNext-Generation Sequencing on Clinical Decision-Making. *Mol Imaging Radionucl Ther* 2017; 26:24-35. DOI:10.4274/2017.26.

19. Haugen B, Alexander E, 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer the American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*, Volume 26, Number 1, 2016. DOI: 10.1089/thy.2015.0020.
20. Bible K, Kebebew E, et.al, 2021 American Thyroid Association Guidelines for Management of Patients with Anaplastic Thyroid Cancer American Thyroid Association Anaplastic Thyroid Cancer Guidelines Task Force. *Thyroid*, 2021, 31:3(337-386). DOI: 10.1089/thy.2020.0944.
21. Murrieta E, Vázquez A, Variabilidad interobservador para la categorización de pacientes con nódulos tiroideos de acuerdo con el TIRADS 2017. *Anales de Radiología México*. 2021; 20:275-287. DOI: 10.24875/ARM.200001801.
22. Édgar Cortés-Reyes TF, Jorge Andrés Rubio-Romero, Métodos Estadísticos De Evaluación De La Concordancia Y La Reproducibilidad De Pruebas Diagnósticas, *Revista Colombiana de Obstetricia y Ginecología* Vol. 61 No. 3, 2010 (247-255).
23. C.M. González Vásquez, J.A. Muñoz Durán, S. Isaza Zapata, J.F. González Londoño, V. García Gómez, Concordancia del TIRADS-ACR, *Radiología*, Volumen 63, Issue 6, 2021, Pages 469-475.
24. Víctor M. Verónica A. Muestreo y Tamaño de Muestra, guía práctica para personal de salud que realiza investigación, capítulo 3, página 85, 2003, México.
25. 64<sup>a</sup> Asamblea General, Fortaleza, Brasil, octubre 2013, DECLARACIÓN DE HELSINKI DE LA AMM – PRINCIPIOS ÉTICOS PARA LAS INVESTIGACIONES MÉDICAS EN SERES HUMANOS.
26. Secretaría de Salud, Diario Oficial de la Federación. Ley General de Salud, Secretaría de Salud, 1984, Última Reforma DOF 24-03-2023.
27. Giovanella L, Avram A, et.al, EANM practice guideline/SNMMI procedure standard for RAIU and thyroid scintigraphy. *Eur J Nucl Med Mol Imaging*, 2019. DOI: 10.1007/s00259-019-04472-8.

## VIII. ANEXOS

### ESCALA ACR TI-RADS

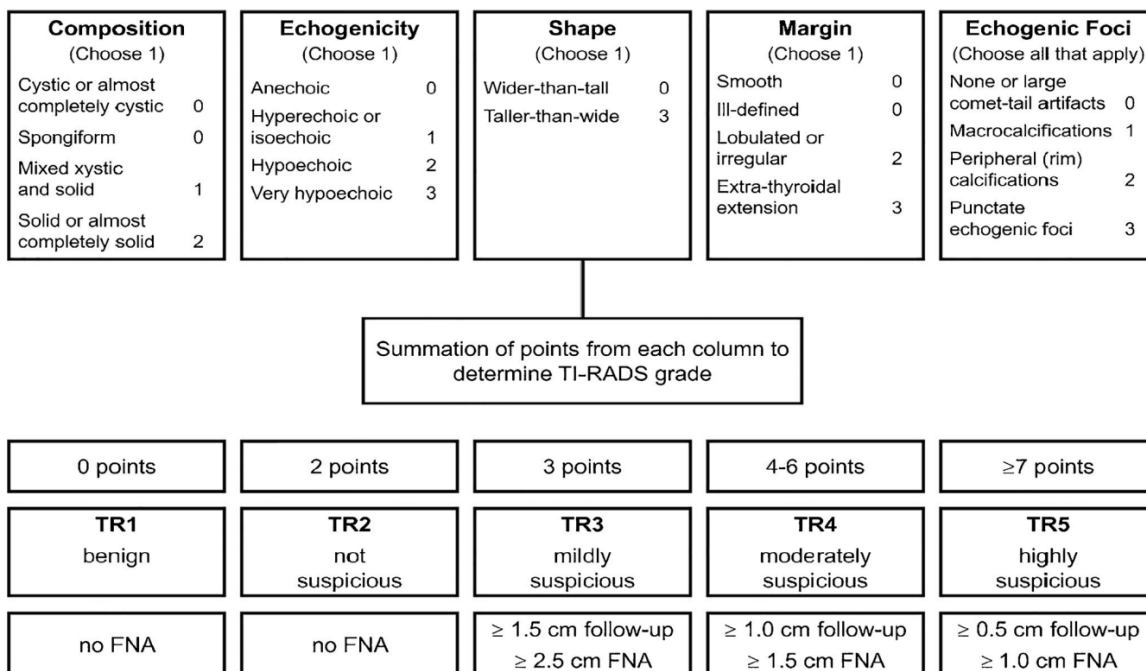


Fig. 1 American College of Radiology (ACR) TI-RADS classification of thyroid nodules and management recommendations

-Imagen obtenida de: Giovannella L, Avram A, et.al, EANM practice guideline/SNMMI procedure standard for RAIU and thyroid scintigraphy. Eur J Nucl Med Mol Imaging, 2019. DOI: 10.1007/s00259-019-04472-8.<sup>27</sup>

## MEDIDAS SIMÉTRICAS DEL SERVICIO DE CIRUGÍA DE CABEZA Y CUELLO

### BAAFC VS BAAFCX1

#### Medidas simétricas

	Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo Kappa	.143	.295	.474	.635
N de casos válidos	9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.



## BAAFC VS BAAFCX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.143	.295	.474	.635
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## BAAFCX1 VS BAAFCX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.500	.306	1.500	.134
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSC VS TIRADSCX1

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.182	.178	1.188	.235
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSC VC TIRADSCX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.274	.199	1.643	.100
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSCX1 VC TIRADSCX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.045	.134	.306	.759
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## MEDIDAS SIMÉTRICAS DEL SERVICIO DE ENDOCRINOLOGÍA

### BAAFC VS BAAFEN1

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.143	.295	.474	.635
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

### BAAFC VS BAAFEN2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.341	.304	1.050	.294
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

### BAAFEN1 VS BAAFEN2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.769	.212	2.372	.018
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSC VS TIRADSEN1

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.156	.169	.966	.334
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSC VS TIRADSEN2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	-.143	.108	-.843	.399
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSEN1 VS TIRADSEN2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.143	.208	.843	.399
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## MEDIDAS SIMÉTRICAS DEL SERVICIO DE MEDICINA INTERNA

### BAAFC VS BAAFMI1

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	-.154	.315	-.474	.635
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## BAAFC VS BAAFMI2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.341	.304	1.050	.294
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## BAAFMI1 VS BAAFMI2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.571	.241	1.897	.058
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSC VS TIRADSMI1

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.156	.166	.949	.343
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSC VS TIRADSMI2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.156	.186	.932	.351
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## TIRADSMI1 VS TIRADSMI2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.308	.203	1.921	.055
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

## MEDIDAS SIMÉTRICAS DEL SERVICIO DE RADIOLOGÍA INTERVENCIONISTA

### BAAFC VS BAAFRX1

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.769	.212	2.372	.018
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

### BAAFC VS BAAFRX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	1.000	.000	3.000	.003
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

### BAAFRX1 VS BAAFRX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.769	.212	2.372	.018
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

### TIRADSC VS TIRADSRX1

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.286	.223	1.657	.098
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

### TIRADSC VS TIRADSRX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.016	.201	.098	.922
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.

### TIRADSRX1 VS TIRADSRX2

### Medidas simétricas

		Valor	Error estándar asintótico <sup>a</sup>	T aproximada <sup>b</sup>	Significación aproximada
Medida de acuerdo	Kappa	.410	.234	2.337	.019
N de casos válidos		9			

a. No se presupone la hipótesis nula.

b. Utilización del error estándar asintótico que presupone la hipótesis nula.