

ORIGINAL

Prevalencia de cáncer en pacientes con hipotiroidismo: análisis mediante herramientas de *big data*



Juan J. Díez^{a,b,*}, Luis Cabrera^{a,b}, Pedro Iglesias^{a,b}, María Benavent^c,
Guillermo Argüello^c, Guillermo López^c, Alejandro Parralejo^c y Javier Leal^d

^a Servicio de Endocrinología y Nutrición, Hospital Universitario Puerta de Hierro Majadahonda, Instituto de Investigación Sanitaria Puerta de Hierro Segovia de Arana, Majadahonda, Madrid, España

^b Departamento de Medicina, Universidad Autónoma de Madrid, Madrid, España

^c MedSavana S.L., Madrid, España

^d Servicio de Informática, Hospital Universitario Puerta de Hierro Majadahonda, Instituto de Investigación Sanitaria Puerta de Hierro Segovia de Arana, Majadahonda, Madrid, España

Recibido el 12 de abril de 2022; aceptado el 21 de mayo de 2022

Disponible en Internet el 26 de agosto de 2022

PALABRAS CLAVE

Hipotiroidismo;
Cáncer;
Prevalencia;
Big data

Resumen

Objetivo: Evaluar la frecuencia de diferentes tipos de cáncer en pacientes con el diagnóstico de hipotiroidismo utilizando metodología de *big data* mediante la plataforma Savana Manager. **Métodos:** Se realizó un estudio observacional, retrospectivo, empleando datos de la historia clínica electrónica del Hospital Universitario Puerta de Hierro Majadahonda (Madrid). La información de las historias clínicas electrónicas se extrajo mediante técnicas de inteligencia artificial y se analizó mediante el software Savana Manager 3.0. Se realizaron búsquedas empleando el término «hipotiroidismo» y los términos correspondientes a los tumores analizados.

Resultados: De un total de 506.749 pacientes estudiados se encontraron 23.570 (4,7%) con el diagnóstico de hipotiroidismo. Los pacientes con este diagnóstico presentaron una frecuencia de cáncer significativamente superior a la hallada en sujetos no hipotiroideos (OR 2,09, intervalo de confianza [IC] al 95%: 2,01-2,17). Esta frecuencia mayor se encontró tanto en mujeres (OR 1,99, IC 95%: 1,90-2,08) como en varones (OR 2,83, IC 95%: 2,63-3,05); sin embargo, no se detectó en los pacientes con hipotiroidismo mayores de 60 años (OR 0,97, IC 95%: 0,92-1,02). Aunque la frecuencia de la mayoría de las neoplasias estudiadas individualmente fue mayor en la población con hipotiroidismo, observamos que los pacientes de más de 60 años presentaban una disminución significativa de la frecuencia de cáncer de próstata, pulmón, colorrectal y hepático.

* Autor para correspondencia.

Correo electrónico: juanjose.diez@salud.madrid.org (J.J. Díez).

Conclusión: Los datos de esta cohorte hospitalaria sugieren que existe una asociación significativa entre el diagnóstico de hipotiroidismo y el cáncer. Sin embargo, esta asociación es menos manifiesta en pacientes mayores de 60 años.

© 2022 SEEN y SED. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

Hypothyroidism;
Cancer;
Prevalence;
Big data

Prevalence of cancer in patients with hypothyroidism: Analysis using big data tools

Abstract

Objective: To evaluate the frequency of different types of cancer in patients diagnosed with hypothyroidism using big data methodology on the Savana Manager platform.

Methods: An observational, retrospective study was carried out using electronic medical record (EMR) data from the Hospital Universitario Puerta de Hierro Majadahonda (Madrid). Information from the EMRs was extracted using artificial intelligence techniques and analysed using the Savana Manager v3.0 software. Searches were performed using the term "hypothyroidism" and the terms corresponding to the tumours analysed.

Results: Of a total of 506,749 patients studied, 23,570 (4.7%) were diagnosed with hypothyroidism. Patients with this diagnosis had a significantly higher frequency of cancer than that found in non-hypothyroid subjects (OR 2.09, 95% confidence interval [CI] 2.01-2.17). This higher frequency was found both in women (OR 1.99, 95% CI 1.90-2.08) and in men (OR 2.83, 95% CI 2.63-3.05). However, this higher frequency of cancer was not observed in hypothyroid patients older than 60 years (OR 0.97, 95% CI 0.92-1.02). Although the frequency of most of the neoplasms studied individually was higher in those with hypothyroidism, we observed that hypothyroid patients over 60 years of age had a significant decrease in the frequency of prostate, lung, colorectal and liver cancer.

Conclusion: Data from this hospital cohort suggest that there is a significant association between a diagnosis of hypothyroidism and cancer. However, this association is less evident in hypothyroid patients older than 60 years.

© 2022 SEEN y SED. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

Desde hace años se han llevado a cabo diversos estudios clínicos y epidemiológicos para averiguar las relaciones entre la disfunción tiroidea y diferentes tipos de enfermedades neoplásicas¹⁻⁴, si bien los resultados obtenidos no han sido concluyentes, especialmente en lo que se refiere a los grupos tumorales de mayor prevalencia en la población. Una reciente revisión sistemática de ensayos clínicos controlados y estudios observacionales no encontró una relación entre el hipotiroidismo subclínico y los cánceres de mama y de próstata, aunque sí con el cáncer colorrectal⁵. Se encontró, además, un incremento del riesgo de cáncer de tiroides en pacientes con elevación de la concentración sérica de tirotrópina (TSH) por encima de 1,64 mU/l. Otra revisión sistemática y metaanálisis de 15 estudios observacionales sobre la asociación entre la disfunción tiroidea y diferentes tipos de cáncer⁶ mostró que, en comparación con sujetos eutiroideos, el hipotiroidismo se asociaba con un mayor riesgo de cáncer de tiroides en los 10 primeros años tras el diagnóstico de la deficiencia hormonal tiroidea.

Recientemente está creciendo en el sector sanitario, y especialmente entre los clínicos, la necesidad de emplear herramientas de inteligencia artificial (IA) para extraer información de valor de la enorme cantidad de datos que se generan en los centros sanitarios^{7,8}. Gran parte del tiempo

de consulta se dedica a completar la historia clínica electrónica (HCE) del paciente⁹, y la mayor parte de la información generada se encuentra en forma de texto libre. Por su tamaño y formato esta información no estructurada es poco accesible en una revisión manual tradicional.

En este contexto los recientes avances en los campos de procesamiento de lenguaje natural (PLN), *big data* e IA cobran especial relevancia. Ello se debe a varios motivos. Primero, estas herramientas permiten la extracción de información real, contrastable y directa de la práctica clínica; segundo, el acceso a la información es rápido, sencillo y menos engorroso que con las formas tradicionales de revisión de datos clínicos; y, tercero, permiten obtener datos de cantidades ingentes de sujetos que, de otra forma, sería imposible abarcar. Mediante técnicas de PLN se genera, a partir de datos no estructurados y difícilmente explotables, una base de datos estructurada en su totalidad al reconocer las entidades clínicas de interés gracias a modelos de IA formados por algoritmos matemáticos. Dicha base de datos se puede utilizar con fines de investigación a través de Savana Manager, un visor de base de datos estructurada que permite analizar el texto libre, interpretar el contenido de las HCE con independencia del sistema de gestión utilizado en los hospitales y evaluar los principales indicadores de un determinado proceso clínico, evitando los sesgos de selección más allá de la propia existencia del registro y sin

requerir conocimiento de lenguaje de programación¹⁰. Además, la extracción de esta información clínica sigue una metodología de validación externa que permite confirmar la calidad del proceso^{10,11}. La literatura reciente muestra que esta tecnología de PLN para extraer información automáticamente de las HCE se ha empleado ya con éxito en especialidades como neumología^{7,8,12}, cardiología¹³, nutrición clínica¹⁴ o gastroenterología^{15,16}.

Dado que el número de estudios que han examinado la asociación del hipotiroidismo con la prevalencia de diferentes tumores es limitado y con resultados muy dispares, nos pareció oportuno analizar estas posibles relaciones mediante métodos de IA. Por ello, el objetivo principal del presente estudio ha sido la descripción, utilizando herramientas de *big data*, de las frecuencias de diferentes tipos de cáncer en pacientes con hipotiroidismo en la base de datos extraída de las HCE del Hospital Universitario Puerta de Hierro Majadahonda (HUPHM). Como objetivo secundario nos propusimos analizar si las relaciones encontradas variaban con la edad y el sexo de los sujetos, así como con el tratamiento sustitutivo del hipotiroidismo.

Material y métodos

Diseño del estudio

Se realizó un estudio observacional, retrospectivo y no intervencionista utilizando los datos que se encuentran en la HCE del HUPHM. El estudio comprendió los datos clínicos recogidos entre el 1 de enero de 2008 y el 31 de diciembre de 2018, que incluyen un total de 506.749 pacientes. Dichos datos provenían de documentos generados en hospitalización, urgencias y consultas externas. El presente estudio fue aprobado por el Comité Ético de Investigación Clínica del HUPHM (código PI 193/21).

Extracción de la información de la historia clínica electrónica

Mediante técnicas de IA, y en concreto de PLN, se extrajo la información de las HCE y se analizó mediante el software Savana Manager v3.0. Este es un software que permite interpretar y explotar la información clínica que se encuentra en las historias clínicas, convirtiendo los datos generados en el hospital, incluyendo la información contenida en el texto libre, en datos estructurados y reutilizables con fines de investigación. En cuanto a las variables incluidas en estudio, el número potencial de variables que se incluyeron se limitó a la información contenida en las HCE durante el periodo de estudio.

Evaluación de la información

La primera fase para construir la base de datos visible a través de Savana Manager fue la adquisición de datos. En cumplimiento del Reglamento general de protección de datos esta adquisición fue responsabilidad del Servicio de Informática del HUPHM que colaboró con el personal técnico de Savana para pseudonimizar los datos y transferirlos a Savana. Los datos fueron integrados en una base de datos

agregada para ser procesados con la tecnología EHRead®. Esta tecnología incluye técnicas de PLN para extraer la información del texto libre reconociendo las variables clínicas de interés, las negaciones y la temporalidad, entre otras expresiones, y permite construir una base de datos sintética de pacientes. La terminología de Savana está basada en SNOMED CT e incluye más de 400.000 conceptos médicos, acrónimos y parámetros de laboratorio. Las entidades terminológicas detectadas en las historias clínicas de los pacientes se clasificaron posteriormente en función de las secciones de la HCE (por ejemplo datos demográficos, historial médico, medicamentos, diagnósticos, etc.). Por último, los autores del HUPHM validaron los resultados de la herramienta y el rendimiento de la tecnología.

Dicha evaluación pretendía verificar la validez de la tecnología EHRead® en la identificación de los registros que contenían menciones de *hipotiroidismo* y variables relacionadas. Se verificaron manualmente un conjunto de 91 documentos que garantizaron la fiabilidad de la anotación/revisión manual y constituyeron el *gold standard*. El rendimiento de Savana se calculó usando como recurso de evaluación dicho *gold standard* creado por los expertos, es decir, la precisión de Savana en la identificación de registros en los que se detectó la presencia de la enfermedad en estudio y las variables relacionadas se midieron con respecto al *gold standard*. El rendimiento se calculó por las métricas estándar de precisión (P), cobertura[®] y la F-score, que es la media armónica de las 2 anteriores¹⁷.

La precisión indicó la fiabilidad de la información recuperada por el sistema y se calculó como $P = tp / (tp + fp)$. La cobertura, un indicador de la cantidad de información que el sistema recupera, se calculó como $R = tp / (tp + fn)$. La F-score se calculó como $F = 2 \times \text{precisión} \times \text{cobertura} / (\text{precisión} + \text{cobertura})$. Este parámetro ofrecía un indicador del rendimiento global de la recuperación de la información.

En todos los casos los verdaderos positivos (tp) eran la suma de registros identificados correctamente, los falsos negativos (fn) eran la suma de registros no identificados y los falsos positivos (fp) eran la suma de registros recuperados incorrectamente¹¹.

En el proceso de extracción de datos para este estudio la variable *hipotiroidismo* ha tenido en cuenta a todos los pacientes en los que se haya reconocido en su HCE a lo largo de su curso clínico el término hipotiroidismo, con independencia de la etiología de la deficiencia hormonal (primario o hipotirotoxico), la duración del trastorno (permanente o transitorio), el grado de hipofunción (subclínico o franco), la asociación o no con otros síndromes clínicos, o el hecho de que el paciente haya recibido o no tratamiento sustitutivo. La evaluación lingüística de la variable *hipotiroidismo* ha sido analizada en el contexto de este estudio, arrojando una precisión, una cobertura y una medida F de 0,97, 0,99 y 0,98, respectivamente. Esto indicó que los diagnósticos de hipotiroidismo se detectaron con precisión en la población del estudio.

El resto de las variables anotadas (*cáncer de mama, próstata, pulmón, colorrectal, vesical, leucemia, tiroides, linfoma, renal y gástrico*) mostraron una F-score mayor de 0,75, con la excepción de la variable *cáncer hepático*, en la que solo había 3 anotaciones, no siendo esta muestra

Tabla 1 Rendimiento de Savana expresado mediante la medida F para las principales variables del estudio

Variabes	Medida F
Cáncer total	0,48
Mama	0,81
Próstata	0,86
Pulmón	0,95
Colorrectal	0,76
Hepático	0,67
Vesical	0,86
Leucemia	1,00
Tiroides	1,00
Linfoma	0,79
Renal	0,94
Gástrico	1,00

significativa, y la variable *cáncer total*, en la que el grado de acuerdo entre los anotadores fue del 0,56, demostrando la complejidad en la determinación terminológica de este concepto genérico (tabla 1).

Análisis estadístico

El abordaje estadístico de los datos incluyó un análisis descriptivo de todas las variables evaluadas, en las que las variables cualitativas se expresan como frecuencias absolutas y porcentajes. Para medir la asociación y comparar proporciones entre variables cualitativas se utilizó la prueba de la Chi-cuadrado. El riesgo relativo de cáncer en pacientes con hipotiroidismo en comparación con sujetos sin hipotiroidismo se estimó por medio de la razón de verosimilitudes (*odds ratio* [OR]). En todos los casos las diferencias cuyo valor de «p» asociado con la prueba de contraste era menor a 0,05 se consideraron significativas.

Resultados

Pacientes estudiados

De los 506.749 sujetos registrados en la herramienta Savana el 53,45% eran mujeres y el 48,71% eran varones. El 2,17% de la población aparecía clasificada en algún momento a lo largo de su curso clínico como hombre o mujer, probablemente debido a errores administrativos. La edad media de los sujetos era de 41 años (intervalo de confianza al 95%: 40,89-41,02).

Frecuencia de cáncer en pacientes con hipotiroidismo

La frecuencia total de neoplasias malignas en la cohorte de pacientes con hipotiroidismo estudiada fue de 13,80%, lo que resultó significativamente superior a la hallada en pacientes sin el diagnóstico de hipotiroidismo (7,11%, OR 2,09 [2,01-2,17]; $p < 0,001$) (tabla 2). Esta mayor frecuencia de cáncer se detectó tanto en mujeres (OR 1,99 [1,90-2,08]) como en hombres (OR 2,83 [2,63-3,05]; $p < 0,001$ en ambos casos), pero no en el grupo de pacientes con doble asignación de

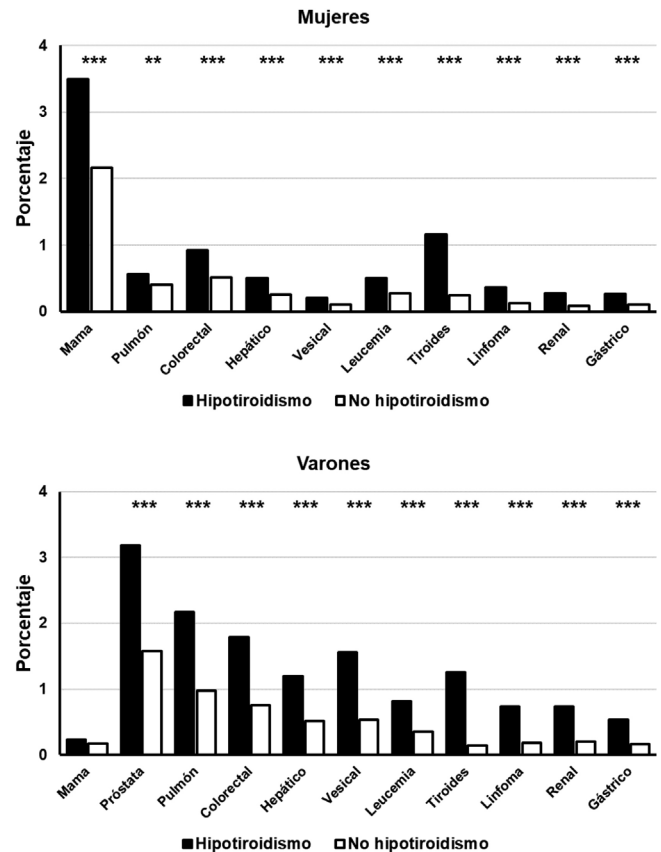


Figura 1 Frecuencia de detección de neoplasias seleccionadas en pacientes con hipotiroidismo (rectángulos negros) y sin hipotiroidismo (rectángulos blancos), clasificados por sexo. Abscisas: tipos de neoplasias estudiadas. Ordinadas: porcentaje de pacientes en cada grupo. * $p < 0,05$. ** $p < 0,01$. *** $p < 0,001$.

sexo (OR 1,14 [0,97-1,34]; $p = 0,111$). Aunque el grupo de pacientes jóvenes (hasta 59 años) mostró una mayor frecuencia de cáncer en hipotiroideos frente a no hipotiroideos (OR 2,46 [2,31-2,63]; $p < 0,001$), esto no ocurrió en el grupo de pacientes mayores de 60 años, en el que las frecuencias fueron muy similares (OR 0,97 [0,92-1,02]; $p = 0,267$) (tabla 2).

La tabla 3 recoge las frecuencias de los tumores estudiados. Como puede apreciarse, la frecuencia de todos los tumores, con excepción del cáncer de próstata, fue significativamente superior en los pacientes con hipotiroidismo.

Frecuencia de cáncer por grupos de sexo y edad

Cuando se analizaron las frecuencias de los diferentes cánceres por grupos de sexo se observó que las mujeres con hipotiroidismo presentaron una frecuencia significativamente superior a la hallada en mujeres sin esta patología en todas las neoplasias analizadas ($p < 0,001$) (fig. 1). La frecuencia de las distintas neoplasias fue también significativamente superior en los varones hipotiroideos con respecto a los no hipotiroideos ($p < 0,001$), con la única excepción del cáncer de mama (fig. 1).

Los pacientes con hipotiroidismo jóvenes (hasta 59 años) presentaron también una prevalencia significativamente

Tabla 2 Frecuencia de cáncer total en la cohorte completa de pacientes estudiados y en pacientes clasificados por grupos de edad y sexo

	Todos		Pacientes con hipotiroidismo			Pacientes sin hipotiroidismo			OR	IC 95%	p
	N	n	N	n	%	N	n	%			
Todos	506.749	37.597	23.570	3.253	13,80	483.179	34.344	7,11	2,09	2,01-2,17	< 0,001
Mujeres	270.858	18.798	19.321	2.359	12,21	251.537	16.439	6,53	1,99	1,90-2,08	< 0,001
Varones	246.858	19.149	4.747	891	18,77	242.111	18.258	7,54	2,83	2,63-3,05	< 0,001
Doble asignación	11.004	1.275	1.598	204	12,77	9.406	1.071	11,39	1,14	0,97-1,34	0,111
Edad hasta 59 años	391.810	13.947	13.250	1.058	7,98	378.560	12.889	3,40	2,46	2,31-2,63	< 0,001
Edad: 60 o más años	114.939	23.650	10.320	3.080	20,16	104.619	21.570	20,62	0,97	0,92-1,02	0,267

IC: intervalo de confianza; N: número de pacientes en cada grupo o subgrupo; n: número de pacientes con cáncer en cada grupo o subgrupo; OR: *odds ratio*.

Tabla 3 Frecuencia neoplasias seleccionadas en todos los pacientes estudiados con y sin hipotiroidismo

	Todos (n = 506.749)	Pacientes con hipotiroidismo (n = 23.570)		Pacientes sin hipotiroidismo (n = 483.179)		p
	Núm.	Núm.	%	Núm.	%	
Mama	6.348	687	2,91	5.661	1,17	< 0,001
Próstata	3.973	151	0,64	3.822	0,79	0,011
Pulmón	3.603	212	0,90	3.391	0,07	< 0,001
Colorrectal	3.388	262	1,11	3.126	0,65	< 0,001
Hepático	2.055	155	0,66	1.900	0,39	< 0,001
Vesical	1.699	114	0,48	1.585	0,33	< 0,001
Leucemia	1.677	139	0,59	1.538	0,32	< 0,001
Tiroides	1.246	287	1,22	959	0,20	< 0,001
Linfoma	853	108	0,46	749	0,16	< 0,001
Renal	804	89	0,38	715	0,15	< 0,001
Gástrico	745	77	0,33	668	0,14	< 0,001

mayor de cáncer de mama, pulmón, colorrectal, hepático, renal, leucemia, tiroides y linfoma, pero no se encontraron diferencias significativas entre pacientes con o sin hipotiroidismo en las prevalencias de cáncer de próstata, vesical y gástrico (fig. 2). En el grupo de pacientes de más de 60 años el diagnóstico de hipotiroidismo se acompañó de una menor frecuencia de cáncer de próstata, pulmón, colorrectal y hepático. La prevalencia de cáncer de mama, tiroides, gástrico, renal y linfoma fue, por el contrario, significativamente superior en los sujetos con el diagnóstico de hipotiroidismo. No hubo diferencias significativas en las frecuencias de cáncer vesical y leucemia (fig. 2).

Frecuencia de cáncer en pacientes con hipotiroidismo clasificados por grupos de tratamiento

La frecuencia total de cáncer hallada en el grupo de 7.075 pacientes con hipotiroidismo en tratamiento sustitutivo fue de 22,72% (tabla 4). Este valor resultó significativamente superior al hallado en el grupo de 16.495 pacientes sin tratamiento con levotiroxina (9,97%, OR 2,66 [2,46-2,86]; $p > 0,001$). Cuando se analizaron de forma individual cada uno de los tumores estudiados se encontró que, en todos los

casos, el tratamiento sustitutivo del hipotiroidismo se asoció a una mayor frecuencia de cáncer ($p < 0,001$).

Discusión

En nuestro conocimiento este es el primer estudio realizado en España que utiliza datos de vida real con herramientas de *big data* para analizar las relaciones entre el hipotiroidismo y el cáncer. Los resultados muestran que la frecuencia global de cáncer es significativamente superior en pacientes con el diagnóstico de hipotiroidismo con respecto a sujetos no hipotiroideos. Esta relación persiste tanto en varones como en mujeres, pero no se mantiene en sujetos de edad superior a 60 años. El análisis de cánceres individuales muestra que todas las neoplasias (con excepción de la prostática) se presentan con mayor frecuencia en hipotiroideos y, dentro de ellos en los que siguen tratamiento sustitutivo con levotiroxina. Otro hallazgo de interés es que en los pacientes mayores de 60 años hubo 4 neoplasias (próstata, pulmón, colorrectal y hepática), cuya frecuencia se encontró significativamente reducida en el hipotiroidismo y otras 2 (vesical y leucemia) con frecuencia similar a la de sujetos sin esta patología.

Tabla 4 Frecuencia de cáncer total y de neoplasias seleccionadas en el grupo de 23.570 pacientes hipotiroideos clasificados según la indicación de tratamiento sustitutivo con levotiroxina

	Todos (n = 23.570)	Pacientes con tratamiento (n = 7.075)		Pacientes sin tratamiento (n = 16.495)		p
	Núm.	Núm.	%	Núm.	%	
Cáncer total	3.253	1.608	22,72	1.645	9,97	< 0,001
Mama	687	248	3,51	439	2,67	< 0,001
Próstata	151	76	1,07	75	0,45	< 0,001
Pulmón	212	129	1,82	83	0,50	< 0,001
Colorrectal	262	110	1,55	152	0,92	< 0,001
Hepático	155	89	1,26	66	0,40	< 0,001
Vesical	114	59	0,83	55	0,33	< 0,001
Leucemia	139	80	1,13	59	0,36	< 0,001
Tiroides	287	226	3,19	61	0,37	< 0,001
Linfoma	108	68	0,96	40	0,24	< 0,001
Renal	89	46	0,65	43	0,26	< 0,001
Gástrico	77	40	0,57	37	0,22	< 0,001

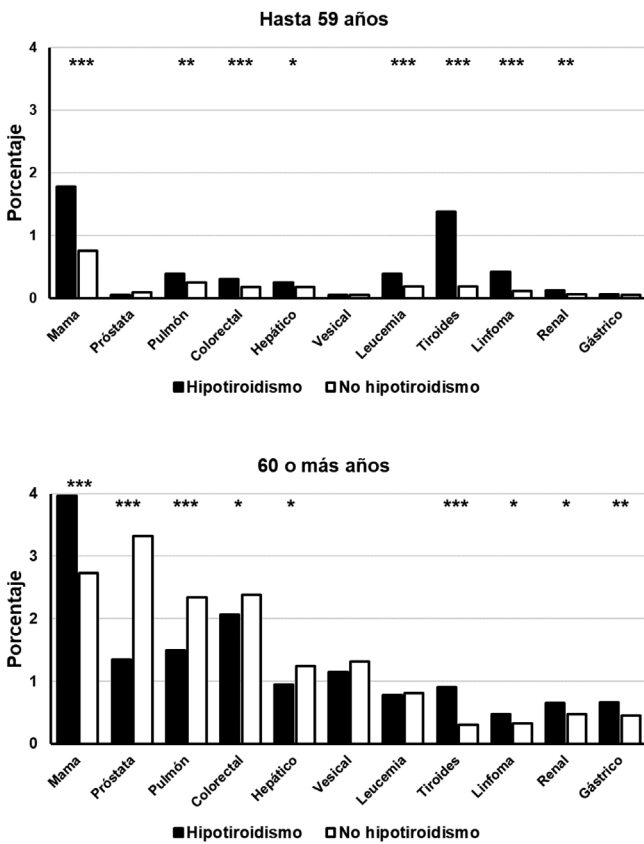


Figura 2 Frecuencia de detección de neoplasias seleccionadas en pacientes con hipotiroidismo (rectángulos negros) y sin hipotiroidismo (rectángulos blancos), clasificados por grupos etarios.

Abscisas: tipos de neoplasias estudiadas. Ordenadas: porcentaje de pacientes en cada grupo. *p < 0,05. **p < 0,01. ***p < 0,001.

La relación entre la disfunción tiroidea y el cáncer es compleja. Aunque algunos estudios han puesto de manifiesto que tanto el hipertiroidismo¹⁸ como el hipotiroidismo^{19,20} se asocian a un incremento de la mortalidad por cáncer, los

datos sobre la incidencia de los diferentes tumores han sido muy dispares. No es de extrañar, por tanto, que los resultados obtenidos en nuestros pacientes contrasten con algunos de los publicados en los estudios recientes en algunas de las neoplasias.

Varios estudios han mostrado que la elevación de TSH^{3,4} o el diagnóstico de hipotiroidismo^{5,21} se asocian con un riesgo de cáncer de mama, mientras que otros investigadores no han encontrado tal asociación^{1,2,22}. Un análisis de datos del estudio *Women’s Health Initiative*, llevado a cabo en 134.122 mujeres posmenopáusicas, mostró la existencia de una asociación inversa significativa entre el cáncer de mama invasivo y la historia de hipotiroidismo²³. Un estudio británico²⁴, que incluyó datos de más de 200.000 mujeres, de las que el 8,7% eran hipotiroideas, no encontró asociación entre el hipotiroidismo y el cáncer de mama; sin embargo, registró una reducción del riesgo después de más de 10 años del diagnóstico de hipotiroidismo. Por el contrario, y en consonancia con nuestros resultados, 2 estudios de bases de datos nacionales, uno en Taiwán²⁵ y otro en Holanda²⁶, mostraron que las mujeres con hipotiroidismo presentaban un mayor riesgo de cáncer de mama que las no hipotiroideas. Además, pudo demostrarse que los niveles bajos de hormona tiroidea eran un factor de riesgo independiente para el desarrollo de cáncer de mama en mujeres peri y posmenopáusicas²⁶.

En un estudio que incluyó 402 casos de cáncer de próstata, Mondul et al.²⁷ hallaron que los hombres con mayores valores de TSH sérica presentaban un riesgo disminuido de cáncer de próstata en comparación con los hombres con concentraciones más reducidas de esta hormona. Los hombres con el diagnóstico de hipotiroidismo tuvieron un riesgo menor de cáncer de próstata en comparación con los eutiroides. Estos datos parecen ser concordantes con los hallados en nuestra cohorte de sujetos mayores de 60 años.

Un estudio prospectivo noruego, que incluyó una cohorte de más de 29.000 sujetos, encontró una reducción no significativa del riesgo de cáncer de pulmón en pacientes con hipotiroidismo¹. Otros 2 estudios no encontraron relación entre el hipotiroidismo y el cáncer de pulmón²⁸.

Mu et al.²⁹ mostraron que la prevalencia de hipotiroidismo subclínico fue significativamente superior en el grupo de sujetos con cáncer colorrectal en comparación con los controles. Además, los pacientes con hipotiroidismo subclínico presentaron con mayor frecuencia cáncer de colon avanzado en comparación con los sujetos eutiroideos. Otro estudio³⁰, realizado en una base de datos poblacional con más de 20.000 pacientes con cáncer colorrectal, encontró un leve aumento del riesgo de este tumor en pacientes con hipotiroidismo no tratado en comparación con sujetos eutiroideos. En contraste con nuestros resultados, en este estudio se encontró una asociación negativa entre el cáncer colorrectal y la sustitución con levotiroxina.

El estudio de Reddy et al.³¹ mostró que el hipotiroidismo era más prevalente en pacientes con carcinoma hepatocelular. Un estudio posterior de casos y controles³² mostró que la historia de hipotiroidismo se asociaba con un riesgo elevado de cáncer hepatocelular en las mujeres. Nuestros datos muestran una mayor prevalencia de cáncer hepático en pacientes hipotiroideos. Aunque no detectamos diferencias entre ambos sexos, sí encontramos que el riesgo de este tumor desaparecía en los sujetos mayores de 60 años.

La disfunción tiroidea se ha considerado también como un factor de riesgo de cáncer de tiroides en diversos estudios^{6,33,34}. Un reciente metaanálisis de 12 estudios prospectivos³⁵ puso de manifiesto que en los pacientes con hipotiroidismo el riesgo relativo de cáncer de tiroides fue de 2,72. En consonancia con estos resultados nuestros datos muestran que la frecuencia de cáncer de tiroides está elevada en los sujetos hipotiroideos, así como en todos los subgrupos de edad y sexo estudiados.

La información sobre otros tumores es más escasa. En una cohorte de 57.326 sujetos dados de alta de un hospital danés se encontró un riesgo elevado de cáncer renal en mujeres con el diagnóstico de mixedema y tirotoxicosis, así como un aumento del riesgo de cáncer vesical en mujeres con mixedema y bocio no funcionante. La incidencia de cánceres hematopoyéticos también se encontró elevada en mujeres con mixedema³⁶.

La disparidad de los resultados hallados en los distintos estudios podría explicarse por los diferentes criterios empleados en la definición de hipotiroidismo (determinaciones analíticas, revisión de historias clínicas o de registros, entrevistas con los pacientes), métodos de laboratorio para cuantificar TSH y hormonas tiroideas o intervalos de referencia para estos parámetros. El origen de los pacientes también da lugar a heterogeneidad de resultados, ya que los estudios realizados en hospitales obtienen riesgos más elevados que los de base poblacional o los que utilizan determinaciones analíticas⁶. Las diferentes estimaciones de riesgo entre los estudios también pueden reflejar diferentes grados de gravedad de la disfunción tiroidea o comorbilidades. Por otra parte, la interpretación de los resultados se ve dificultada por la falta de estudios longitudinales, de datos sobre tratamientos para la disfunción tiroidea, sobre factores de confusión o por la inclusión de casos de cáncer prevalentes y la posibilidad de una causalidad inversa, es decir, casos en los que la disfunción tiroidea puede ser consiguiente al cáncer o sus tratamientos. De los comentarios anteriores se deduce la plausibilidad biológica de la relación entre el hipotiroidismo y el cáncer, si bien con los matices de la disparidad de resultados hallados en la

literatura. Del mismo modo, cabe interpretar la utilidad de los estudios que, como el presente, utilizan herramientas de *big data* para el análisis de una enorme cantidad de información que es difícil o imposible obtener con estudios tradicionales retrospectivos o de casos y controles.

Aunque los mecanismos que relacionan el hipotiroidismo con el cáncer no se conocen bien, se han sugerido algunas hipótesis que implican la participación de los receptores de hormonas tiroideas como inhibidores de la oncogénesis³⁷, o los efectos de las hormonas tiroideas sobre el sistema inmune³⁸. Es conocido que la TSH estimula el crecimiento y la diferenciación de las células foliculares tiroideas³⁹. Por otra parte, el hipotiroidismo se asocia a situaciones clínicas como la obesidad, la diabetes⁴⁰ y la enfermedad cardiovascular¹⁸ que, a su vez, se han relacionado con un incremento del riesgo de cáncer.

Uno de los hallazgos más sorprendentes del presente estudio es la mayor prevalencia de cáncer en los pacientes con hipotiroidismo y tratamiento sustitutivo. Esto podría deberse a una mayor frecuencia de hipotiroidismo franco y, por tanto, mayor carga de enfermedad en el grupo de sujetos con medicación sustitutiva que podrían representar el subgrupo de pacientes en los que la asociación de cáncer con la deficiencia hormonal tiroidea es más manifiesta y, por tanto, estadísticamente significativa. Otra potencial explicación vendría dada por un posible efecto proliferativo inducido por las hormonas tiroideas. En efecto, algunos estudios experimentales han mostrado que la tiroxina y la triyodotironina podrían ser antiapoptóticos y tener un efecto proliferativo en las líneas celulares de cáncer de tiroides, mama y próstata regulando la expresión génica^{41,42}, provocando fosforilación a través de la vía MAPK^{43,44} y estimulando efectos similares a los estrógenos⁴⁵. La herramienta informática utilizada en este estudio, no obstante, no permite contrastar estos potenciales mecanismos, pero nuestro hallazgo puede abrir puertas a futuras investigaciones.

La principal fortaleza de nuestro estudio viene dada por la extracción de datos de la vida real y el gran tamaño muestral, que nos ha permitido estudiar las frecuencias de las diferentes neoplasias sin perder potencia estadística. El uso de técnicas de *big data* y la tecnología EHRead desarrolladas por Savana nos ha permitido evaluar una ingente cantidad de información, así como leer, procesar y ordenar el texto libre de las HCE y convertirlo en datos estructurados. En los datos analizados se incluyeron todos los diagnósticos de más de medio millón de pacientes durante el periodo de estudio, de forma verificable y sin sesgo, por lo que pueden extraerse conclusiones derivadas de la práctica clínica real.

Las limitaciones principales son las inherentes al uso de herramientas de inteligencia artificial en grandes bases de datos, en las que no es posible el análisis individual pormenorizado de los sujetos estudiados. Por ello, no hemos podido diferenciar entre hipotiroidismo franco y subclínico. No podemos, por tanto, conocer el porcentaje de pacientes con hipotiroidismo subclínico, aunque podríamos especular, pero no demostrar, que es superior al de pacientes con hipotiroidismo franco, dado que hemos encontrado una mayor proporción de pacientes sin tratamiento sustitutivo. Tampoco fue posible valorar las diferentes etiologías del hipotiroidismo, ni evaluar la cronología del desarrollo de los diferentes tumores en referencia a la fecha del diagnóstico de hipotiroidismo. En consecuencia, solo podemos

inferir una coincidencia diagnóstica, pero no especular sobre la relación temporal entre la disfunción tiroidea y el cáncer. Aunque la información se obtuvo de forma verificable y sin sesgo, no podemos descartar el sesgo generado por errores u omisiones de diagnósticos en las HCE.

La cohorte estudiada procede de los pacientes atendidos en nuestro hospital, por lo que puede no ser representativa de la población general. Además, las frecuencias de los distintos tumores pueden estar sobrevaloradas por el hecho de que el HUPHM es un hospital de tercer nivel que recibe pacientes de alta complejidad de fuera de su área de influencia. La doble asignación de sexo en algunos pacientes es otra limitación, aunque el estudio de este subgrupo no mostró diferencias significativas en la frecuencia de la mayoría de los tumores, por lo que es posible que su repercusión en los resultados finales sea mínima. Aunque la mayoría de las variables presentó un *F score* elevado, algunas de ellas mostraron un valor inferior a 0,75.

A pesar de sus limitaciones creemos que nuestro análisis puede considerarse un estudio piloto con interés clínico, dado que incluye un considerable número de pacientes con los problemas clínicos estudiados y que el hipotiroidismo es un trastorno hormonal de alta prevalencia en nuestro medio. Nuestros resultados podrían servir de base para estudios prospectivos que analicen de forma dirigida la incidencia de diferentes tipos de cáncer en pacientes hipotiroideos.

En conclusión, en la cohorte de pacientes hospitalarios estudiados mediante herramientas de inteligencia artificial hemos encontrado una asociación significativa entre los diagnósticos de hipotiroidismo y cáncer total, así como entre hipotiroidismo y diversas neoplasias estudiadas individualmente. Estas asociaciones se mantienen en pacientes clasificados por sexo, pero no en los diferentes grupos etarios, ya que en los pacientes de más de 60 años nuestros datos muestran una ausencia de relación entre hipotiroidismo y cáncer en general, y una disminución de la frecuencia de cáncer de próstata, pulmón, colorrectal y hepático. Dado que la fiabilidad de los resultados obtenidos con herramientas de *big data* está condicionada por la adecuada codificación de los diagnósticos y el adecuado juicio clínico registrado en las HCE, es deseable realizar estudios específicos para confirmar los resultados obtenidos.

Financiación

Este artículo no ha recibido financiación de ninguna entidad pública ni privada.

Conflicto de intereses

María Benavent, Guillermo Argüello, Guillermo López y Alejandro Parraljo son empleados de Savana y declaran no tener ningún conflicto de intereses adicional. Juan J. Díez, Luis Cabrera, Pedro Iglesias y Javier Leal declaran no tener ningún conflicto de intereses en relación con este estudio.

Bibliografía

1. Hellevik AI, Asvold BO, Bjørø T, Romundstad PR, Nilsen TI, Vatten LJ. Thyroid function and cancer risk: A prospective population study. *Cancer*

- 2009;18:570–4, <http://dx.doi.org/10.1158/1055-9965.EPI-08-0911>.
2. Khan SR, Chaker L, Ruitter R, Aerts JG, Hofman A, DeGhan A, et al. Thyroid function and cancer risk: The Rotterdam Study. *J Clin Endocrinol Metab.* 2016;101:5030–6, <http://dx.doi.org/10.1210/jc.2016-2104>.
3. Kim EY, Chang Y, Lee KH, Yun JS, Park YL, Park CH, et al. Serum concentration of thyroid hormones in abnormal and euthyroid ranges and breast cancer risk: A cohort study. *Int J Cancer.* 2019;145(12):3257–66, <http://dx.doi.org/10.1002/ijc.32283>.
4. Yuan S, Kar S, Vithayathil M, Carter P, Mason AM, Burgess S, et al. Causal associations of thyroid function and dysfunction with overall, breast and thyroid cancer: A two-sample Mendelian randomization study. *Int J Cancer.* 2020;147(7):1895–903, <http://dx.doi.org/10.1002/ijc.32988>.
5. Gómez-Izquierdo J, Filion KB, Boivin JF, Azoulay L, Pollak M, Yu OHY. Subclinical hypothyroidism and the risk of cancer incidence and cancer mortality: A systematic review. *BMC Endocr Disord.* 2020;20:83, <http://dx.doi.org/10.1186/s12902-020-00566-9>.
6. Tran TV, Kitahara CM, de Vathaire F, Boutron-Ruault MC, Journy N. Thyroid dysfunction and cancer incidence: A systematic review and meta-analysis. *Endocr Relat Cancer.* 2020;27:245–59, <http://dx.doi.org/10.1530/ERC-19-0417>.
7. Izquierdo JL, Morena D, González Y, Paredero JM, Pérez B, Graziani D, et al. Clinical management of COPD in a real-world setting. A big data analysis. *Arch Bronconeumol (Engl Ed).* 2021;57:94–100, <http://dx.doi.org/10.1016/j.arbres.2019.12.025>. English, Spanish.
8. Izquierdo JL, Almonacid C, Campos C, Morena D, Benavent M, González-de-Olano D, et al. Systemic corticosteroids in patients with bronchial asthma: A real-life study. *J Investig Allergol Clin Immunol.* 2021;11, <http://dx.doi.org/10.18176/jiaci.0765>, 0.
9. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. *Ann Intern Med.* 2016;165:753–60, <http://dx.doi.org/10.7326/M16-0961>.
10. Hernández-Medrano I, Belda C, Urena A, Salcedo I, Espinosa-Anke LT, Saggion H. Savana: Re-using electronic health records with artificial intelligence. *IJMAI.* 2017;4:8–12.
11. Canales L, Menke S, Marchesseau S, D'Agostino A, Del Rio-Bermudez C, Taberna M, et al. Assessing the performance of clinical natural language processing systems: Development of an evaluation methodology. *JMIR Med Inform.* 2021;9:e20492, <http://dx.doi.org/10.2196/20492>.
12. Izquierdo JL, Almonacid C, González Y, del Río-Bermúdez C, Ancochea J, Cárdenas R, et al. The impact of COVID-19 on patients with asthma. *Eur Respir J.* 2021;57:2003142, <http://dx.doi.org/10.1183/13993003.03142-2020>.
13. González-Juanatey C, Anguita-Sánchez M, Barrios V, Núñez-Gil I, Gómez-Doblas JJ, García-Moll X, et al., SAVANA Research Group. Assessment of medical management in coronary type 2 diabetic patients with previous percutaneous coronary intervention in Spain: A retrospective analysis of electronic health records using Natural Language Processing. *PLoS One.* 2022;17:e0263277, <http://dx.doi.org/10.1371/journal.pone.0263277>.
14. Ballesteros Pomar MD, Pintor de la Maza B, Barajas Galindo D, Cano Rodríguez I. Searching for disease-related malnutrition using big data tools. *Endocrinol Diabetes Nutr (Engl Ed).* 2020;67:224–7, <http://dx.doi.org/10.1016/j.endinu.2019.11.009>. English, Spanish.
15. Montoto C, Gisbert JP, Guerra I, Plaza R, Pajares Villarroya R, Moreno Almazán L, et al., PREMONITION-CD Study Group. Evaluation of natural language processing for the identification of crohn disease-related variables in Spanish electronic health records: A validation study for the

- PREMONITION-CD Project. *JMIR Med Inform.* 2022;10:e30345, <http://dx.doi.org/10.2196/30345>.
16. Gomollón F, Gisbert JP, Guerra I, Plaza R, Pajares Villarroya R, Moreno Almazán L, et al., Premonition-CD Study Group. Clinical characteristics and prognostic factors for Crohn's disease relapses using natural language processing and machine learning: A pilot study. *Eur J Gastroenterol Hepatol.* 2022;34:389–97, <http://dx.doi.org/10.1097/MEG.0000000000002317>.
 17. Baeza-Yates RA, Ribeiro-Neto B. *Modern information retrieval.* Boston: Addison-Wesley Longman Publishing Co, Inc.; 1999.
 18. Brandt F, Thvilum M, Almind D, Christensen K, Green A, Hegedüs L, et al. Graves' disease and toxic nodular goiter are both associated with increased mortality but differ with respect to the cause of death: A Danish population-based register study. *Thyroid.* 2013;23:408–13, <http://dx.doi.org/10.1089/thy.2012.0500>.
 19. Tseng FY, Lin WY, Li Cl, Li TC, Lin CC, Huang KC. Subclinical hypothyroidism is associated with increased risk for cancer mortality in adult Taiwanese—a 10 years population-based cohort. *PLoS One.* 2015;10:e0122955, <http://dx.doi.org/10.1371/journal.pone.0122955>.
 20. Krashin E, Silverman B, Steinberg DM, Yekutieli D, Giveon S, Fabian O, et al. Pre-diagnosis thyroid hormone dysfunction is associated with cancer mortality. *Endocr Relat Cancer.* 2021;28:705–13, <http://dx.doi.org/10.1530/ERC-21-0187>.
 21. Chen S, Wu F, Hai R, You Q, Xie L, Shu L, et al. Thyroid disease is associated with an increased risk of breast cancer: a systematic review and meta-analysis. *Gland Surg.* 2021;10:336–46, <http://dx.doi.org/10.21037/gs-20-878>.
 22. Wang B, Lu Z, Huang Y, Li R, Lin T. Does hypothyroidism increase the risk of breast cancer: Evidence from a meta-analysis. *BMC Cancer.* 2020;20:733, <http://dx.doi.org/10.1186/s12885-020-07230-4>.
 23. Weng CH, Okawa ER, Roberts MB, Park SK, Umbricht CB, Manson JE, et al. Breast cancer risk in postmenopausal women with medical history of thyroid disorder in the women's health initiative. *Thyroid.* 2020;30:519–30, <http://dx.doi.org/10.1089/thy.2019.0426>.
 24. Tran TV, Maringe C, Benitez Majano S, Rachet B, Boutron-Ruault MC, Journy N. Thyroid dysfunction and breast cancer risk among women in the UK Biobank cohort. *Cancer Med.* 2021;10:4604–14, <http://dx.doi.org/10.1002/cam4.3978>.
 25. Huang CH, Wei JC, Chien TC, Kuo CW, Lin SH, Su YC, et al. Risk of breast cancer in females with hypothyroidism: A nationwide, population-based cohort study. *Endocr Pract.* 2021;27:298–305, <http://dx.doi.org/10.1016/j.eprac.2020.09.007>.
 26. Kuijpers JL, Nyklíček I, Louwman MW, Weetman TA, Pop VJ, Coebergh JW. Hypothyroidism might be related to breast cancer in post-menopausal women. *Thyroid.* 2005;15:1253–9, <http://dx.doi.org/10.1089/thy.2005.15.1253>.
 27. Mondul AM, Weinstein SJ, Bosworth T, Remaley AT, Virtamo J, Albanes D. Circulating thyroxine, thyroid-stimulating hormone, and hypothyroid status and the risk of prostate cancer. *PLoS One.* 2012;7:e47730, <http://dx.doi.org/10.1371/journal.pone.0047730>. Epub 2012 Oct 30.
 28. Chan YX, Knuiman MW, Divitini ML, Brown SJ, Walsh J, Yeap BB. Lower TSH and higher free thyroxine predict incidence of prostate but not breast, colorectal or lung cancer. *Eur J Endocrinol.* 2017;177:297–308, <http://dx.doi.org/10.1530/EJE-17-0197>.
 29. Mu G, Mu X, Xing H, Xu R, Sun G, Dong C, et al. Subclinical hypothyroidism as an independent risk factor for colorectal neoplasm. *Clin Res Hepatol Gastroenterol.* 2015;39:261–6, <http://dx.doi.org/10.1016/j.clinre.2014.08.002>.
 30. Boursi B, Haynes K, Mamtani R, Yang YX. Thyroid dysfunction, thyroid hormone replacement and colorectal cancer risk. *J Natl Cancer Inst.* 2015;107:djv084, <http://dx.doi.org/10.1093/jnci/djv084>.
 31. Reddy A, Dash C, Leerapun A, Mettler TA, Stadheim LM, Lazaridis KN, et al. Hypothyroidism: A possible risk factor for liver cancer in patients with no known underlying cause of liver disease. *Clin Gastroenterol Hepatol.* 2007;5:118–23, <http://dx.doi.org/10.1016/j.cgh.2006.07.011>.
 32. Hassan MM, Kaseb A, Li D, Patt YZ, Vauthey JN, Thomas MB, et al. Association between hypothyroidism and hepatocellular carcinoma: A case-control study in the United States. *Hepatology.* 2009;49:1563–70, <http://dx.doi.org/10.1002/hep.22793>.
 33. Kitahara CM, Farkas DK, Jørgensen JOL, Cronin-Fenton D, Sørensen HT. Benign thyroid diseases and risk of thyroid cancer: A nationwide cohort study. *J Clin Endocrinol Metab.* 2018;103:2216–24, <http://dx.doi.org/10.1210/jc.2017-02599>.
 34. Montúfar Silva MR, Cevallos Montalvo JP, Sánchez Andino BM, Villavicencio Soledispa JI, Espinoza López SE. Relación entre hipotiroidismo y cáncer de tiroides en el Hospital Pablo Arturo Suárez de la ciudad de Quito de enero 2014 a diciembre 2017. *CSSN.* 2019;10:21–6 [Consultado 10 May 2022] Disponible en: <<http://revistas.esPOCH.edu.ec/index.php/cssn/article/view/174>> doi: <https://doi.org/10.47244/cssn.Vol10.Iss1.174>.
 35. Liang L, Zheng XC, Hu MJ, Zhang Q, Wang SY, Huang F. Association of benign thyroid diseases with thyroid cancer risk: A meta-analysis of prospective observational studies. *J Endocrinol Invest.* 2019;42:673–85, <http://dx.doi.org/10.1007/s40618-018-0968-z>.
 36. Mellempgaard A, From G, Jørgensen T, Johansen C, Olsen JH, Perrild H. Cancer risk in individuals with benign thyroid disorders. *Thyroid.* 1998;8:751–4, <http://dx.doi.org/10.1089/thy.1998.8.751>.
 37. Ruiz-Llorente L, Martínez-Iglesias O, García-Silva S, Tenbaum S, Regadera J, Aranda A. The thyroid hormone receptors as tumor suppressors. *Horm Mol Biol Clin Investig.* 2011;5:79–89, <http://dx.doi.org/10.1515/HMBCI.201004>.
 38. De Vito P, Incerpi S, Pedersen JZ, Luly P, Davis FB, Davis PJ. Thyroid hormones as modulators of immune activities at the cellular level. *Thyroid.* 2011;21:879–90, <http://dx.doi.org/10.1089/thy.2010.0429>.
 39. Hard GC. Recent developments in the investigation of thyroid regulation and thyroid carcinogenesis. *Environ Health Perspect.* 1998;106:427–36, <http://dx.doi.org/10.1289/ehp.106-1533202>.
 40. Díez JJ, Iglesias P. An analysis of the relative risk for hypothyroidism in patients with type 2 diabetes. *Diabet Med.* 2012;29:1510–4, <http://dx.doi.org/10.1111/j.1464-5491.2012.03687.x>.
 41. Tsui KH, Hsieh WC, Lin MH, Chang PL, Juang HH. Triiodothyronine modulates cell proliferation of human prostatic carcinoma cells by downregulation of the B-cell translocation gene 2. *Prostate.* 2008;68:610–9, <http://dx.doi.org/10.1002/pros.20725>.
 42. Pinto M, Soares P, Ribatti D. Thyroid hormone as a regulator of tumor induced angiogenesis. *Cancer Lett.* 2011;301:119–26, <http://dx.doi.org/10.1016/j.canlet.2010.11.011>.
 43. Moeller LC, Führer D. Thyroid hormone, thyroid hormone receptors, and cancer: A clinical perspective. *Endocr Relat Cancer.* 2013;20:R19–29, <http://dx.doi.org/10.1530/ERC-12-0219>.
 44. Hercbergs A, Mousa SA, Leinung M, Lin HY, Davis PJ. Thyroid hormone in the clinic and breast cancer. *Horm Cancer.* 2018;9:139–43, <http://dx.doi.org/10.1007/s12672-018-0326-9>.
 45. Dinda S, Sanchez A, Moudgil V. Estrogen-like effects of thyroid hormone on the regulation of tumor suppressor proteins, p53 and retinoblastoma, in breast cancer cells. *Oncogene.* 2002;21:761–8, <http://dx.doi.org/10.1038/sj.onc.1205136>.