

ORIGINAL ARTICLE

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



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, 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, Spain

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

^c MedSavana S.L., Madrid, Spain

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

Received 12 April 2022; accepted 21 May 2022

Available online 17 August 2023

KEYWORDS

Hypothyroidism;
Cancer;
Prevalence;
Big data

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 population of 506,749 patients, 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 the population 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 the diagnosis of hypothyroidism and cancer. However, this association is less evident in hypothyroid patients older than 60 years.

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

* Corresponding author.

E-mail address: juanjose.diez@salud.madrid.org (J.J. Díez).

PALABRAS CLAVE

Hipotiroidismo;
Cáncer;
Prevalencia;
Big data

Prevalencia de cáncer en pacientes con hipotiroidismo: análisis mediante herramientas de *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 (HCE) del Hospital Universitario Puerta de Hierro Majadahonda (Madrid). La información de las HCE 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 mayor frecuencia 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 objetivó en los pacientes hipotiroideos 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 hipotiroideos de más de 60 años presentaban una disminución significativa de la frecuencia de cáncer de próstata, pulmón, colorrectal y hepático.

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 hipotiroideos mayores de 60 años.

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

Introduction

For a number of years, different clinical and epidemiological studies have been carried out to ascertain the relationships between thyroid dysfunction and different types of cancerous diseases,^{1–4} though the results obtained have not been conclusive, particularly with regard to the tumour groups of greater prevalence among the population. A recent systematic review of controlled clinical trials and non-interventional studies found no relationship between subclinical hypothyroidism and breast and prostate cancers, although it did with colorectal cancer.⁵ An increased risk of thyroid cancer was also found in patients with a high serum concentration of thyrotropin (TSH) in excess of 1.64 mU/L. Another systematic review and meta-analysis of 15 non-interventional studies on the association between thyroid dysfunction and different types of cancer⁶ showed that, in comparison to euthyroid subjects, hypothyroidism was associated with a greater risk of thyroid cancer in the first 10 years after diagnosis of thyroid hormone deficiency.

Recently, in the healthcare sector, and particularly among clinicians, there has been an increasing need to use tools of artificial intelligence (AI) to extract valuable information from the enormous quantity of data that are generated in healthcare centres.^{7,8} A good deal of consultation time is devoted to filling out the patient's electronic health record (EHR),⁹ and the majority of the information generated is in the form of free text. Owing to its size and format, this unstructured information is not easily accessible in a traditional manual review.

In this setting, the recent advances in the fields of natural language processing (NLP), *big data* and AI are of particular relevance. This is due to various reasons. Firstly, these tools allow the extraction of real, verifiable and direct information from clinical practice; secondly, access to information is fast, simple and less burdensome than with traditional forms of reviewing clinical data; and, thirdly, they make it possible to obtain huge quantities of data on subjects which, otherwise, it would be impossible to cover. Using techniques from NLP, based on unstructured and hard-to-exploit data, a fully structured database is generated by recognising the clinical entities of interest, thanks to AI models formed by mathematical algorithms. This database can be used for research purposes through Savana Manager, a structured database browser which makes it possible to analyse free texts, interpret the content of EHRs independently of the management system used in hospitals, and to evaluate the principle indicators of a determined clinical process, avoiding selection bias beyond the very existence of the register itself, and requiring no knowledge of programming language.¹⁰ Additionally, the extraction of this clinical information adheres to an external validation methodology which allows the quality of the process to be confirmed.^{10,11} The recent literature shows that this NLP technology for extracting information automatically from EHRs has been used successfully in pulmonology,^{7,8,12} cardiology,¹³ clinical nutrition¹⁴ and gastroenterology.^{15,16}

Given that the number of studies that have examined the association of hypothyroidism with the prevalence of different tumours is limited and with very disparate results, we deemed it appropriate to analyse these possible rela-

tionships using AI *methods*. Thus, the main objective of this study is the description, using *big data* tools, of the frequencies of different types of cancer in patients with hypothyroidism on the basis of data extracted from the EHRs of the Hospital Universitario Puerta de Hierro Majadahonda [Puerta de Hierro Majadahonda University Hospital (HUPHM)]. As a secondary objective, we sought to analyse whether the relationships found varied with the age and sex of subjects, as well as with replacement treatment for hypothyroidism.

Material and methods

Study design

An observational, retrospective, non-interventional study was conducted using the data found in the EHR of the HUPHM. The study comprised clinical data collected between 1 January 2008 and 31 December 2018, and included a total of 506,749 patients. These data came from documents generated on inpatient wards, in the Accident and Emergency department, and in outpatient consultations. The study was approved by the HUPHM Independent Ethics Committee (Code PI 193/21).

Extraction of information from electronic medical records

Using AI techniques, and specifically NLP, information was extracted from the EHRs and analysed using Savana Manager V.3.0, software that interprets and uses the clinical information found in the health records, converting the data generated in the hospital, including the information contained in the free text, into structured and reusable data for research purposes. Regarding the variables included in the study, the potential number of variables included was limited to the information contained in EHRs during the study period.

Evaluation of the information

The first phase in constructing the database visible through Savana Manager was data acquisition. In compliance with the General Data Protection Regulation, this acquisition was the responsibility of the HUPHM IT Department, which worked with Savana technical staff to pseudonymise the data and transfer them to Savana. The data were integrated into an aggregate database to be processed with EHRead® technology. This technology includes NLP techniques for extracting information from the free text, recognising clinical variables of interest, denials and temporality, among other expressions, and allows a synthetic database of patients to be built. Savana terminology is based on SNOMED CT and includes over 400,000 medical concepts, acronyms and laboratory parameters. Terminological entities detected in patients' medical records were subsequently classified according to the sections in the EHR (e.g. demographics, medical history, medications, diagnoses, etc.). Finally, the HUPHM authors validated the results of the tool and the performance of the technology.

This evaluation was intended to verify the validity of the EHRead® technology in the identification of the records that mentioned *hypothyroidism* and related variables. A set of 91 documents were manually verified that guaranteed the reliability of the manual notation/review and constituted the *gold standard*. Savana's performance was calculated using this *gold standard* evaluation resource created by the experts; i.e. Savana's accuracy in identifying records in which the disease in question and related variables were detected was measured against the *gold standard*. The performance was calculated by the standard metrics of precision (P), recall (R) and the F-score, which is the harmonic mean of the previous two metrics.¹⁷

Precision indicated the reliability of the information retrieved by the system and was calculated as $P = tp / (tp + fp)$. Recall, an indicator of how much information the system retrieves, was calculated as $R = tp / (tp + fn)$. The F-score was calculated as $F = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$. This parameter offered an indicator of the overall performance of information retrieval.

In all cases, true positives (tp) were the sum of correctly identified records, false negatives (fn) were the sum of unidentified records, and false positives (fp) were the sum of incorrectly retrieved records.¹¹

In the data extraction process for the study, the *hypothyroidism* variable took into account all those patients in whom the term hypothyroidism has been recognised in their EHR throughout their clinical course, irrespective of the aetiology of the hormone deficiency (primary or hypothyrotropic), the duration of the disorder (permanent or transitory), the degree of hypofunction (subclinical or overt), the association or not of other clinical syndromes, or the fact that the patient has not received replacement therapy. The linguistic evaluation of the *hypothyroidism* variable has been analysed in the context of the study, yielding a precision, a recall and an F measurement of 0.97, 0.99 and 0.98, respectively. This indicated that hypothyroidism diagnoses were accurately detected in the study population.

The rest of the variables annotated (*breast, prostate, lung, colorectal, bladder, thyroid, lymphoma, renal and Gastric cancer, as well as leukaemia*) showed an F-score greater than 0.75, with the exception of the *hepatic cancer* variable, in which only three had annotations, this sample not being significant, and the *total cancer* variable, in which the degree of agreement among the annotators was 0.56, demonstrating the complexity in the terminological determination of this generic concept (Table 1).

Statistical analysis

The statistical approach to the data included a descriptive analysis of all the variables evaluated, in which the qualitative variables are expressed as absolute frequencies and percentages. To measure the association and compare proportions between qualitative variables, the chi-square test was used. The relative risk of cancer in patients with hypothyroidism in comparison with subjects without hypothyroidism is estimated using the odds ratio (OR). In all cases, differences with a comparison test-associated "P" value less than .05 were considered significant.

Table 1 Performance of Savana expressed by means of the F measurement for the principal variables of the study.

Variables	F measurement
Total Cancer	0.48
Breast	0.81
Prostate	0.86
Lung	0.95
Colorectal	0.76
Liver	0.67
Bladder	0.86
Leukaemia	1.00
Thyroid	1.00
Lymphoma	0.79
Renal	0.94
Gastric	1.00

Results

Patients studied

Of the 506,749 subjects registered in the Savana tool, 53.45% were female and 48.71% were male. In total, 2.17% of the population was classified at some point during their clinical course as male or female, probably due to administrative

errors. The mean age of the subjects was 41 years (95% confidence interval: 40.89–41.02).

Frequency of cancer in subjects with hypothyroidism

The total frequency of malign neoplasms in the cohort of patients with hypothyroidism studied was 13.80%, which was significantly higher than that found in patients without the diagnosis of hypothyroidism (7.11%, OR 2.09 [2.01–2.17]; $P < .001$) (Table 2). This greater frequency was detected in both females (OR 1.99 [1.90–2.08]) and males (OR 2.83 [2.63–3.05]; $P < .001$ in both cases) but not in the group of patients with double assignment of sex (OR 1.14 [0.97–1.34]; $P = .111$). Although the group of young patients (up to 59 years) showed a greater frequency of cancer among hypothyroid patients as opposed to non-hypothyroid patients (OR 2.46 [2.31–2.63]; $P < .001$), this did not occur in the group of patients aged over 60 years, in which the frequencies were very similar (OR 0.97 [0.92–1.02]; $P = .267$) (Table 2).

Table 3 shows the frequency of tumours studied. As can be seen, the frequency of all tumours, with the exception of prostate cancer, was significantly higher in patients with hypothyroidism.

Table 2 Frequency of total cancer in the complete cohort of patients studied and in patients classified by age groups and sex.

	All		Subjects with hypothyroidism			Subjects without hypothyroidism			OR	95% CI	P
	N	n	N	n	%	N	n	%			
All	506,749	37,597	23,570	3253	13.80	483,179	34,344	7.11	2.09	2.01–2.17	< .001
Female	270,858	18,798	19,321	2359	12.21	251,537	16,439	6.53	1.99	1.90–2.08	< .001
Males	246,858	19,149	4747	891	18.77	242,111	18,258	7.54	2.83	2.63–3.05	< .001
Double assignment	11,004	1275	1598	204	12.77	9406	1071	11.39	1.14	0.97–1.34	.111
Age up to 59 years	391,810	13,947	13,250	1058	7.98	378,560	12,889	3.40	2.46	2.31–2.63	< .001
Age: 60 years or older	114,939	23,650	10,320	3080	20.16	104,619	21,570	20.62	0.97	0.92–1.02	.267

CI: confidence interval; N: number of subjects in each group or sub-group; n: number of subjects with cancer in each group or sub-group; OR: odds ratio.

Table 3 Frequency of selected neoplasms in all patients studied with and without hypothyroidism.

	All (n = 506,749)		Subjects with hypothyroidism (n = 23,570)		Subjects without hypothyroidism (n = 483,179)		P
	No.		No.	%	No.	%	
Breast	6348		687	2.91	5,661	1.17	< .001
Prostate	3973		151	0.64	3,822	0.79	.011
Lung	3603		212	0.90	3,391	0.07	< .001
Colorectal	3388		262	1.11	3,126	0.65	< .001
Liver	2055		155	0.66	1,900	0.39	< .001
Bladder	1699		114	0.48	1,585	0.33	< .001
Leukaemia	1677		139	0.59	1,538	0.32	< .001
Thyroid	1246		287	1.22	959	0.20	< .001
Lymphoma	853		108	0.46	749	0.16	< .001
Renal	804		89	0.38	715	0.15	< .001
Gastric	745		77	0.33	668	0.14	< .001

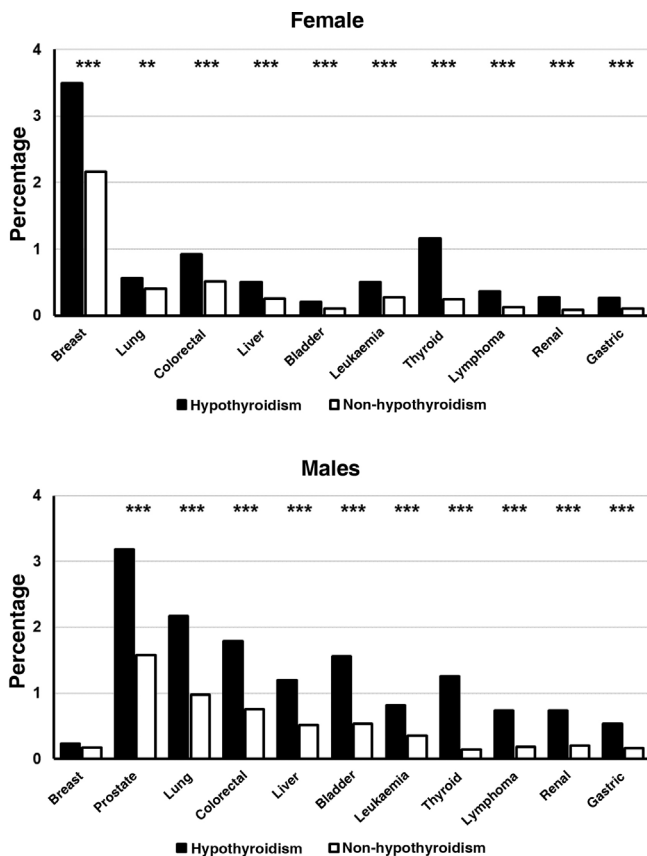


Figure 1 Frequency of neoplasms selected in patients with hypothyroidism (black rectangles) and without hypothyroidism (white rectangles), classified by sex. X axis: types of neoplasms studied. Y axis: percentage of patients in each group. * $P < .05$. ** $P < .01$. *** $P < .001$.

Frequency of cancer by sex and age groups

When the frequencies of the different cancers were analysed by sex groups, it was observed that women with hypothyroidism presented a significantly higher frequency than that found in women without this disease, in all neoplasms analysed ($P < .001$) (Fig. 1). The frequency of the different neoplasms was also significantly higher in male hypothyroid patients compared to those who were not hypothyroid ($P < .001$), the sole exception being breast cancer (Fig. 1).

Young patients with hypothyroidism (up to 59 years) also presented a significantly higher prevalence of breast, lung, colorectal, hepatic, renal, thyroid, and lymphoma cancer, as well as leukaemia, but no significant differences were found between patients with and without hypothyroidism in the prevalence of prostate, bladder and gastric cancer (Fig. 2). In the group of patients aged over 60 years, the diagnosis of hypothyroidism was accompanied by a lower frequency of prostate, lung, colorectal and liver cancer. On the contrary, the prevalence of breast, thyroid, gastric, renal, and lymphoma cancer was significantly higher in subjects with the diagnosis of hypothyroidism. There were no significant differences in the frequencies of bladder cancer and leukaemia (Fig. 2).

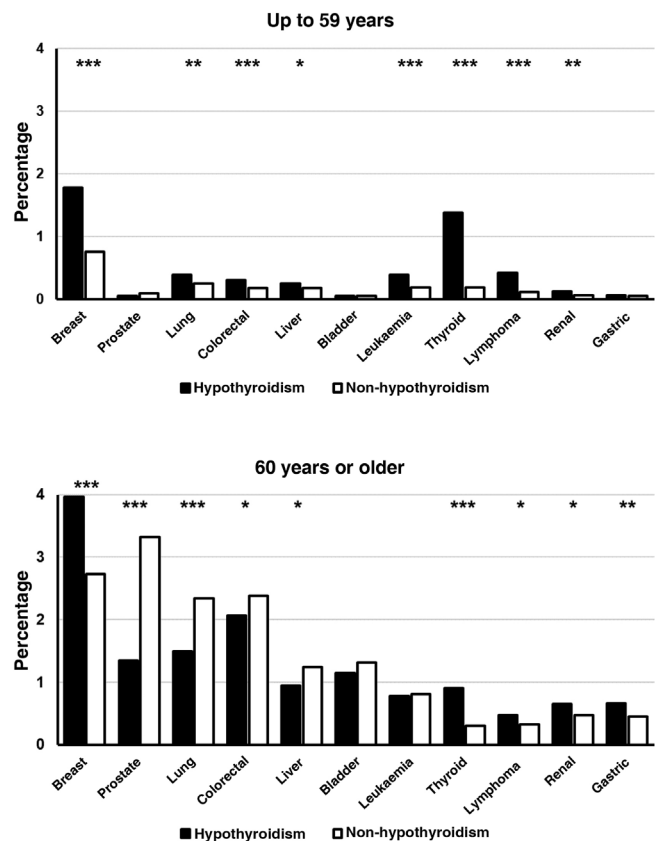


Figure 2 Frequency of neoplasms selected in patients with hypothyroidism (black rectangles) and without hypothyroidism (white rectangles), classified by age groups. X axis: types of neoplasms studied. Y axis: percentage of patients in each group. * $P < .05$. ** $P < .01$. *** $P < .001$.

Frequency of cancer in subjects with hypothyroidism classified by treatment groups

The total frequency of cancer found in the group of 7075 patients with hypothyroidism under replacement therapy was 22.72% (Table 4). This value was significantly higher than that found in the group of 16,495 patients without treatment with levothyroxine (9.97%, OR 2.66 [2.46–2.86]; $P > .001$). When each one of these tumours studied were analysed individually, it was found that, in all cases, replacement therapy for hypothyroidism was associated with a greater frequency of cancer ($P < .001$).

Discussion

To the best of our knowledge, this is the first study carried out in Spain which uses real-life data with *big data* tools to analyse the relationship between hypothyroidism and cancer. The results show that the global frequency of cancer is significantly higher in patients with a diagnosis of hypothyroidism compared to non-hypothyroid subjects. This ratio persisted in both males and females, but is not maintained in subjects aged over 60 years. The analysis of individual cancers shows that all neoplasms (with the exception of prostate cancer) present with greater frequency in

Table 4 Frequency of total cancer and of selected neoplasms in the group of 23,570 hypothyroid patients classified according to initiation of replacement therapy with levothyroxine.

	All (n = 23,570)	Subjects with treatment (n = 7075)		Subjects without treatment (n = 16,495)		P
	No.	No.	%	No.	%	
Total Cancer	3253	1608	22.72	1645	9.97	< .001
Breast	687	248	3.51	439	2.67	< .001
Prostate	151	76	1.07	75	0.45	< .001
Lung	212	129	1.82	83	0.50	< .001
Colorectal	262	110	1.55	152	0.92	< .001
Liver	155	89	1.26	66	0.40	< .001
Bladder	114	59	0.83	55	0.33	< .001
Leukaemia	139	80	1.13	59	0.36	< .001
Thyroid	287	226	3.19	61	0.37	< .001
Lymphoma	108	68	0.96	40	0.24	< .001
Renal	89	46	0.65	43	0.26	< .001
Gastric	77	40	0.57	37	0.22	< .001

hypothyroid patients and, within these, in those undergoing replacement treatment with levothyroxine. Another finding of interest is that in patients aged over 60 years, there were four neoplasms (prostate, lung, colorectal and hepatic) whose frequency was found to be significantly reduced in hypothyroidism, and a further 2 (bladder and leukaemia) with a frequency similar to that of subjects without this disease.

The relationship between thyroid dysfunction and cancer is complex. Although some studies have shown that both hypothyroidism¹⁸ and hypothyroidism^{19,20} are associated with an increase in mortality from cancer, the data on the incidence of the different tumours have been highly disparate. Thus it is not surprising that the results obtained in our patients contrast with some of those published in recent studies on some of the neoplasms.

Various studies have shown that high TSH^{3,4} and the diagnosis of hypothyroidism^{5,21} are associated with a greater risk of cancer, while other researchers have not found any such association.^{1,2,22} An analysis of the data from the *Women's Health Initiative* study, conducted on 134,122 menopausal women, showed the existence of a significant inverse association between invasive breast cancer and a history of hypothyroidism.²³ A British study,²⁴ which included data on more than 200,000 women, of whom 8.7% were hypothyroid, found no association between hypothyroidism and breast cancer; however, it did record a reduction in risk after more than 10 years from diagnosis of hypothyroidism. On the contrary, and in line with our results, 2 studies on national databases, one in Taiwan²⁵ and another in the Netherlands,²⁶ showed that women with hypothyroidism presented a greater risk of breast cancer than non-hypothyroid women. It was also demonstrated that low levels of thyroid hormone were an independent risk factor for developing breast cancer in peri- and post-menopausal women.²⁶

In a study which included 402 cases of prostate cancer, Mondul et al.²⁷ found that men with higher serum TSH Values presented a lower risk of prostate cancer in comparison to men with low concentrations of this hormone. Males with the diagnosis of hypothyroidism had a lower risk of prostate can-

cer in comparison to euthyroid subjects. These data appear to be consistent with those found in our cohort of subjects aged over 60 years.

A Norwegian prospective study, which included a cohort of over 29,000 subjects, found a non-significant reduction in the risk of lung cancer in patients with hypothyroidism.¹ Two other studies found no relationship between hypothyroidism and lung cancer.²⁸

Mu et al.²⁹ showed that the prevalence of subclinical hypothyroidism was significantly higher in the group of subjects with colorectal cancer in comparison with the controls. Moreover, patients with subclinical hypothyroidism presented advanced colon cancer with greater frequency in comparison to euthyroid subject. Another study,³⁰ conducted on a population database with more than 20,000 patients with colorectal cancer, found a slight increase in the risk of this tumour in patients with untreated hypothyroidism in comparison to euthyroid patients. In contrast to our findings, in said study a negative association was found between colorectal cancer and replacement therapy with levothyroxine.

The study by Reddy et al.³¹ showed that hypothyroidism was more prevalent in patients with hepatocellular carcinoma. A subsequent study of cases and controls³² showed that a history of hypothyroidism is associated with an increased risk of hepatocellular cancer in women. Our data show a higher prevalence of liver cancer in hypothyroid patients. Although we detected no difference between sexes, we did find that the risk of this tumour disappeared in subjects over 60 years of age.

Thyroid dysfunction has also been considered a factor of risk for thyroid cancer in a number of studies.^{6,33,34} A recent meta-analysis of 12 prospective studies³⁵ revealed that, in patients with hypothyroidism, the relative risk of thyroid cancer was 2.72. Consistent with these results, our data show that the frequency of thyroid cancer is high in hypothyroid subjects, as well as in all sub-groups of age and sex studied.

Information on other tumours is scarcer. In a cohort of 57,326 subjects discharged from a Danish hospital, a higher risk of renal cancer was found in women with the diagnosis of

myxedema and thyrotoxicosis, as well as an increase in the risk of bladder cancer in women with myxedema and non-functioning goitre. The incidence of haematopoietic cancers was also found to be higher in women with myxedema.³⁶

The disparity of the results found in the different studies could be explained by the different criteria employed in the definition of hypothyroidism (laboratory determinations, review of clinical histories or records, interviews with patients), laboratory methods for quantifying TSH, and thyroid hormones or reference intervals for these parameters. The origin of patients also gives rise to heterogeneity in results since studies conducted in hospitals obtain higher risks than those with a population basis or those which use laboratory determinations.⁶ The different risk estimates between the studies could also reflect different degrees of severity of thyroid dysfunction or comorbidities. On the other hand, the interpretation of the results is complicated by the lack of longitudinal studies, of data on treatments for thyroid dysfunction, on factors of confusion, or owing to the inclusion of prevalent cases of cancer and the possibility of reverse causality; i.e. cases in which thyroid dysfunction may be a consequence of the cancer or the treatments thereof. From the above comments, the biological plausibility of the relation between hypothyroidism and cancer is deduced, but with the nuances of the disparity of results found in the literature. In the same way, we can interpret the usefulness of studies which, like this study, use *big data* for the analysis of an enormous quantity of information which is difficult or impossible to obtain with traditional retrospective or case-control studies.

Although the mechanisms which relate hypothyroidism with cancer are not well known, a number of hypotheses have been suggested which involve the participation of thyroid hormone receptors as oncogenesis inhibitors,³⁷ or the effects of thyroid hormones on the immune system.³⁸ TSH is known to stimulate the growth and differentiation of thyroid follicular cells.³⁹ On the other hand, hypothyroidism is associated with clinical situations such as obesity, diabetes⁴⁰ and cardiovascular disease¹⁸ which, in turn, have been related with an increased risk of cancer.

One of the most surprising findings of this study is the increased prevalence of cancer in patients with hypothyroidism and replacement treatment. This could be due to a greater frequency of overt hypothyroidism and, thus, a greater disease burden in the group of subjects with substitution medication, who could represent the subgroup of patients in which the association of cancer with thyroid hormone deficiency is more evident and, therefore, statistically significant. Another potential explanation for this is a possible proliferative effect induced by thyroid hormones. Indeed, some experimental studies have shown that thyroxine and tri-iodothyronine could be anti-apoptotics and may have a proliferative effect on the cell lines of thyroid, breast and prostate cancer, regulating gene expression,^{41,42} triggering phosphorylation through the MAPK pathway,^{43,44} and stimulating similar effects to oestrogens.⁴⁵ Nonetheless, the computer tool used in this study does not allow these potential mechanisms to be substantiated, but our finding could open doors for future research.

The main strength of our study is given by the extraction of real-life data and the large sample size, which has enabled us to study the frequencies of different neoplasms

without losing statistical power. The use of *big data* techniques, along with and the EHRead technology developed by Savana, has allowed us to evaluate an enormous amount of information, as well as to read, process and organise the free text from the EHRs and convert it into structured data. All diagnoses of more than half a million patients during the study period were included in the analysed data, in a verifiable, bias-free manner, owing to which conclusions can be drawn from actual clinical practice.

The main limitations are those inherent to the use of artificial intelligence tools in large databases, in which detailed individual analysis of the subjects studied is not possible. Hence we have not been able to differentiate between overt and subclinical hypothyroidism. We cannot, therefore, determine the percentage of patients with subclinical hypothyroidism, although we could speculate, but not demonstrate, that it is higher than that of patients with overt hypothyroidism, given that we have found a greater proportion of patients without replacement treatment. Nor was it possible to assess the various aetiologies of hypothyroidism, or assess the development timeline of the different tumours with reference to the date of diagnosis of hypothyroidism. Consequently, we can only infer a diagnostic coincidence, but not speculate on the temporal relationship between thyroid dysfunction and cancer. Although the information was obtained in a verifiable, bias-free manner, we cannot rule out the bias generated by diagnostic errors or omissions in the EHRs.

The cohort studied was drawn from patients treated at our hospital, owing to which it may not be representative of the general population. Moreover, the frequencies of tumours may be overestimated due to the fact that the HUPHM is a tertiary hospital which receives highly complex patients from outside its catchment area. The double assignation of sex in some patients is a further limitation, although the study of this subgroup showed no significant differences in the frequency of the majority of tumours, owing to which its impact on the final results may be minimal. Although the majority of variables presented a high *F score*, some of them show a value below 0.75.

Despite its limitations, we believe that our analysis can be considered a pilot study with clinical interest, given that it includes a considerable number of patients with the clinical problems studied, and that hypothyroidism is a hormonal disorder of high prevalence in our setting. Our results could serve as a basis for prospective studies to analyse the incidence of different types of cancer in hypothyroid patients in a targeted manner.

By way of conclusion, in the cohort of hospital patients studied using artificial intelligence tools, we have found a significant association between the diagnoses of hypothyroidism total cancer, as well as between hypothyroidism and different neoplasms studied individually. These associations are maintained in patients classified by sex, but not in different age groups, since in patients over 60 years of age, our data point to the absence of a relationship between hypothyroidism and cancer in general, and the reduction in the frequency of prostate, lung, colorectal and hepatic cancer. Given that the reliability of results obtained with *big data* is conditional on the proper coding of diagnoses and the proper clinical judgement being recorded in EHRs, spe-

cific studies will need to be carried out in order to confirm the results obtained.

Funding

This article has received no funding from any public or private entity.

Conflicts of interest

María Benavent, Guillermo Argüello, Guillermo López and Alejandro Parraljo are employees Savana and they declare that they have no other conflicts of interest. Juan J. Díez, Luis Cabrera, Pedro Iglesias and Javier Leal declare that they have no conflicts of interest relating to this study.

References

- Hellevik AI, Asvold BO, Bjørø T, Romundstad PR, Nilsen TI, Vatten LJ. Thyroid function and cancer risk: a prospective population study. *Cancer Epidemiol Biomarkers Prev*. 2009;18:570–4, <http://dx.doi.org/10.1158/1055-9965.EPI-08-0911>.
- Khan SR, Chaker L, Ruiter R, Aerts JG, Hofman A, Dehghan 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>.
- 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;145:3257–66, <http://dx.doi.org/10.1002/ijc.32283>.
- 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;147:1895–903, <http://dx.doi.org/10.1002/ijc.32988>.
- 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>.
- 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>.
- 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.
- 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 Invest Allergol Clin Immunol*. 2021;11, <http://dx.doi.org/10.18176/jiaci.0765>, 0.
- Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeters 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>.
- Hernández-Medrano I, Belda C, Urena A, Salcedo I, Espinosa-Anke LT, Saggion H. Savana: re-using electronic health records with artificial intelligence. *IJIMAI*. 2017;4:8–12.
- 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>.
- Izquierdo JL, Almonacid C, González Y, Del Rio-Bermudez 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>.
- González-Juanatey C, Anguita-Sá Nchez 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>.
- 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.
- Montoto C, Gisbert JP, Guerra I, Plaza R, Pajares Villarroja 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>.
- Gomollón F, Gisbert JP, Guerra I, Plaza R, Pajares Villarroja 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>.
- Baeza-Yates RA, Ribeiro-Neto B. *Modern information retrieval*. Boston: Addison-Wesley Longman Publishing Co, Inc.; 1999.
- 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>.
- Tseng FY, Lin WY, Li CI, 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>.
- Krashin E, Silverman B, Steinberg DM, Yekutieli D, Givon 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>.
- 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>.
- 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>.
- 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>.
- 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>.
- 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, <http://dx.doi.org/10.47244/cssn.Vol10.Iss1.174> [accessed 10 May 2022] Available from: <<http://revistas.esPOCH.edu.ec/index.php/cssn/article/view/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. Mellemegaard 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 Invest*. 2011;5:79–89, <http://dx.doi.org/10.1515/HMBCI.2010.04>.
 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>.