

## ORIGINAL ARTICLE

## Integration of big data analytics in the investigation of the relationship between acromegaly and cancer

Pedro Iglesias<sup>a,b,\*</sup>, Javier Arias<sup>c</sup>, Guillermo López<sup>c</sup>, Iago Romero<sup>c</sup>, Juan J. Díez<sup>a,b</sup><sup>a</sup> Department of Endocrinology and Nutrition, University Hospital Puerta de Hierro Majadahonda, Instituto de Investigación Sanitaria Puerta de Hierro Segovia de Arana, Majadahonda, Madrid, Spain<sup>b</sup> Department of Medicine, Universidad Autónoma de Madrid, Spain<sup>c</sup> MedSavana S.L., Madrid, Spain

Received 19 February 2024; accepted 2 May 2024

## KEYWORDS

Acromegaly;  
Cancer;  
Prevalence;  
Big data

## Abstract

**Objective:** To evaluate the association between acromegaly and cancer and different types of cancer by using natural language processing systems and big data analytics.**Material and methods:** We conducted an observational, retrospective study utilizing data from the electronic health records (EHRs) of Hospital Universitario Puerta de Hierro Majadahonda, Madrid, Spain. Information from the EHRs was extracted using artificial intelligence techniques and analyzed using Savana Manager 4.0 software.**Results:** Out of a total of 708,047 registered patients (54.7% females), 544 patients (0.08%; 330 women, 60.7%; mean age at diagnosis  $53.0 \pm 15.8$  yr) were diagnosed with acromegaly. The incidence of cancer was higher in patients with acromegaly vs those without this condition (7.7% vs 3.9%,  $p < 0.001$ ; OR, 2.047, 95%CI, 1.493–2.804). Male acromegalic patients had a higher prevalence of cancer vs females (57.1% vs 42.9%,  $p = 0.012$ ). A significantly higher prevalence of colorectal cancer (2.9% vs 1.4%,  $p = 0.006$ ), bladder cancer (1.1% vs 0.3%,  $p = 0.005$ ), and lymphoma (1.1% vs 0.3%,  $p = 0.009$ ) was observed in patients with acromegaly vs those without the condition. Acromegalic men had significantly higher prevalence rates of colorectal cancer (4.7% vs 1.3%,  $p = 0.001$ ), bladder cancer (2.8% vs 0.4%,  $p < 0.001$ ), breast cancer (0.9% vs 0.2%,  $p = 0.042$ ), gastric cancer (0.9% vs 0.1%,  $p = 0.011$ ), lymphoma (1.4% vs 0.3%,  $p = 0.037$ ), and liver cancer (0.9% vs 0.1%,  $p = 0.012$ ) vs non-acromegalic men. On the other hand, acromegalic women showed a higher prevalence of thyroid cancer (1.2% vs 0.4%,  $p = 0.043$ ) vs non-acromegalic women.**Conclusion:** Our study, based on artificial intelligence techniques and analysis of real-world data and information, revealed a significant association between acromegaly and cancer in our hospital population, mainly acromegalic men, with a higher frequency of colorectal cancer, bladder cancer and lymphoma in particular.

© 2024 SEEN y SED. Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

\* Corresponding author.

E-mail address: piglo65@gmail.com (P. Iglesias).

**PALABRAS CLAVE**

Acromegalia;  
Cáncer;  
Prevalencia;  
Big data

**Integración del análisis de Big Data en la investigación de la relación entre acromegalia y cáncer****Resumen**

**Objetivo:** Evaluar la asociación de la acromegalia con el cáncer y los tipos de cáncer mediante el uso de sistemas de procesamiento de lenguaje natural y análisis de Big Data.

**Material y métodos:** Se realizó un estudio observacional retrospectivo a partir de los datos de las historias clínicas electrónicas (HCE) del Hospital Universitario Puerta de Hierro Majadahonda (España). La información de las HCE se extrajo mediante técnicas de inteligencia artificial y se analizó con el software Savana Manager 4.0.

**Resultados:** De un total de 708.047 pacientes registrados (54,7% mujeres), 544 pacientes (0,08%; 330 mujeres [60,7%]; edad media en el momento del diagnóstico  $53,0 \pm 15,8$  años) fueron diagnosticados de acromegalia. La frecuencia de cáncer fue mayor en los pacientes con acromegalia que en los que no la padecían (7,7% vs 3,9%,  $p < 0,001$ ; OR: 2,047; IC 95%: 1,493–2,804). Los pacientes acromegálicos varones presentaron una mayor prevalencia de cáncer en comparación con las mujeres (57,1% vs 42,9%,  $p = 0,012$ ). Se observó una prevalencia significativamente mayor de cáncer colorrectal (2,9% vs 1,4%,  $p = 0,006$ ), cáncer de vejiga (1,1% vs 0,3%,  $p = 0,005$ ) y linfoma (1,1% vs 0,3%,  $p = 0,009$ ) en pacientes con acromegalia en comparación con los que no padecían la enfermedad. Los varones acromegálicos presentaron tasas de prevalencia significativamente superiores de cáncer colorrectal (4,7% vs 1,3%,  $p = 0,001$ ), cáncer de vejiga (2,8% vs 0,4%,  $p < 0,001$ ), cáncer de mama (0,9% vs 0,2%,  $p = 0,042$ ), cáncer gástrico (0,9% vs 0,1%,  $p = 0,011$ ), linfoma (1,4% vs 0,3%,  $p = 0,037$ ) y cáncer de hígado (0,9% vs 0,1%,  $p = 0,012$ ) en comparación con los varones no acromegálicos. Por otra parte, las mujeres acromegálicas mostraron una mayor prevalencia de cáncer de tiroides (1,2% vs 0,4%,  $p = 0,043$ ) en comparación con las mujeres no acromegálicas.

**Conclusión:** Nuestro estudio, basado en técnicas de inteligencia artificial y análisis de datos e información del mundo real, reveló una asociación significativa entre acromegalia y cáncer en nuestra población hospitalaria, principalmente en varones acromegálicos, con una mayor frecuencia de cáncer colorrectal, cáncer de vejiga y linfoma en particular.

© 2024 SEEN y SED. Publicado por Elsevier España, S.L.U. All rights are reserved, including those for text y data mining, AI training, y similar technologies.

**Introduction**

Acromegaly is a chronic hormonal disorder characterized by the excessive production of growth hormone (GH) by a GH-secreting pituitary neuroendocrine tumor (PitNET).<sup>1</sup> This tumor is usually benign but can affect both the rest of the healthy pituitary parenchyma, interfering with normal pituitary function, and other surrounding structures such as the optic pathways, the cranial nerves, and hypothalamic structures.

Excessive secretion of both tumor-derived GH and its target hormone—the insulin-like growth factor 1 (IGF-1)—by the liver plays a key role in the development of symptoms primarily related to overgrowth of body tissues and organs, as well as various systemic metabolic, musculoskeletal, cardiovascular, respiratory, neurological, and neoplastic complications.

The relationship between excess GH and cancer is controversial. Until a few years ago, most studies conducted in patients with acromegaly found associations with colorectal and thyroid neoplasms.<sup>2,3</sup> However, recent research has also shown an increased risk of other types of cancer, including gastric, breast, and urinary tract cancers.<sup>4</sup>

Interest in using Artificial Intelligence (AI) tools to draw valuable information from the vast amounts of data generated in health care facilities has grown in recent years.<sup>5–9</sup> Most clinical information collected in electronic health records (EHRs) is in free-text format and, due to its volume, it is difficult to access for traditional manual review. In this context, recent advances made in clinical natural language processing (NLP), big data analytics, and AI programs make it possible to quickly and easily obtain clinical data from many patients with real, verifiable and direct information from clinical practice. Tool Savana Manager can analyze free text, interpret the content of EHRs, regardless of the management system used in hospitals, and evaluate the main indicators of a specific clinical process, avoiding selection biases beyond the very existence of the record.<sup>10,11</sup>

As far as we know, the number of studies that have specifically evaluated the association between cancer and acromegaly using AI methods is limited. Therefore, our aim was to use NLP and big data tools to analyze the prevalence of cancer and the incidence of different types of cancer in patients diagnosed with acromegaly in the database extracted from the EHRs of Hospital Universitario Puerta de Hierro Majadahonda (HUPHM), Madrid, Spain. We also evaluated whether the association found between cancer and

acromegaly was significant vs the general population treated in our hospital and, finally, we studied the effect of sex on the different types of cancer in the acromegalic population.

## Material and methods

### Study design

We conducted an observational, retrospective, non-interventional study using data that from the EHRs of HUPHM. The study included clinical data collected from September 25<sup>th</sup>, 2008, through March 22<sup>nd</sup>, 2023. Patients of any age were studied, including all hospital services and settings both inpatient and outpatient settings. These data came from documents generated in hospitalization, the ER, and outpatient departments. The present study was approved by the HUPHM Clinical Research Ethics Committee (code PI 192/23).

### Study variables

The prevalence of total cancer, as well as different types of cancer (colorectal cancer, breast cancer, lung cancer, bladder cancer, lymphoma, prostate cancer, thyroid cancer, gastric cancer, liver cancer, kidney cancer, and leukemia) was analyzed in both patients diagnosed with acromegaly and those without this diagnosis. These diagnoses were considered when they were recorded in the patient's EHR.

### Extraction and evaluation of information

Data was drawn from the EHRs using AI techniques, specifically NLP. This information was, then, analyzed using Savana Manager version 4.0 software. Savana Manager is a software designed to interpret and utilize clinical information drawn from health records by converting data generated by hospitals, including free text, into structured and reusable data for research purposes. The selection of study variables was limited to the information available in the EHRs during the study period. The pseudonymized data were, then, transferred to Savana and later integrated into a centralized database, where it was processed using EHRead<sup>®</sup> technology.

This technology utilizes NLP techniques to extract information from free text, identifying relevant clinical variables, negations, associated values, and other expressions. It enables the creation of a synthetic database of patients by processing and organizing the pseudonymized data. The terminology used in Savana based on SNOMED CT includes more than 400,000 medical concepts, acronyms, and laboratory parameters. During the process, the detected terminological entities in the patients' health records were categorized according to different sections within the EHRs, such as demographics, health history, drugs, diagnoses, etc. To guarantee the accuracy and reliability, authors from the HUPHM validated the results of the tool and assessed the performance of the EHRead<sup>®</sup> technology.

A set of 119 documents were manually verified, which guaranteed the reliability of the manual annotation/review and constituted the gold standard. The performance of

Savana was calculated using as the evaluation tools such gold standard created by the experts, i.e., the accuracy of Savana in identifying health records in which the presence of the disease under study and related variables detected was measured in relation to the gold standard. Performance was calculated by the standard metrics of precision ( $P$ ), coverage ( $R$ ) and the  $F$ -score, which is the harmonic mean of the 2 previous metrics.<sup>12</sup>

Accuracy indicated the reliability of the information retrieved by the system and was calculated as  $P = tp / (tp + fp)$ . Coverage, an indicator of the amount of information retrieved by the system, was calculated as  $R = tp / (tp + fn)$ . The  $F$ -score was calculated as  $F = 2 \times \text{precision} \times \text{coverage} / (\text{precision} + \text{coverage})$ . This parameter provided an indicator of overall information retrieval performance. 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.

The linguistic evaluation of the variable "acromegaly" analyzed in the context of this study showed an accuracy, coverage and  $F$ -score of 1, 1, and 1, respectively. This indicated that acromegaly diagnoses were accurately detected in the study population. The variable acromegaly in the context of the present study includes all patients with this diagnosis regardless of growth hormone (GH) and insulin-like growth factor type 1 (IGF-1) levels whether or not the patient had been on somatostatin analogs, dopamine agonists, GH receptor antagonists (pegvisomant), and/or treated with surgery.

The remaining variables observed showed  $F$  scores  $> 0.75$  except for liver cancer (0.66). The  $F$ -score values were  $r$ , 0.76 for colorectal cancer, 0.81 for breast cancer, 0.94 for lung cancer, 0.85 for bladder cancer, 0.78 for lymphoma, 0.85 for prostate cancer, 1 for thyroid cancer, 1 for stomach cancer, 0.94 for kidney cancer, and 1 for leukemia.

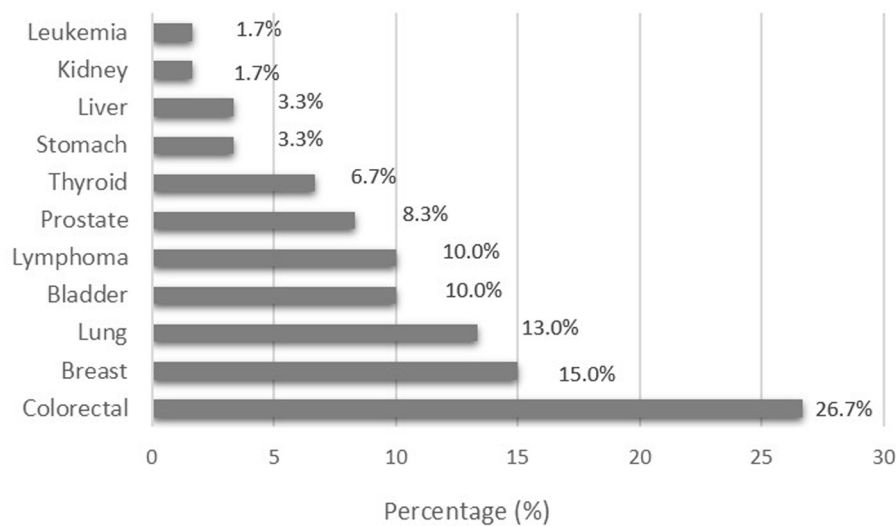
### Statistical analysis

We conducted the descriptive analysis for all the variables evaluated. Qualitative variables were expressed as absolute frequencies and percentages. To measure association and compare proportions between qualitative variables, the Chi-square test and the Fisher's exact test were used, when necessary. The relative risk of total cancer and each type of cancer analyzed in patients diagnosed with acromegaly vs subjects without acromegaly was estimated by odds ratio (OR). In all cases, differences whose  $p$  value associated with the contrast test was  $< 0.05$  were considered significant.

## Results

### Study patients

A total of 708,047 patients (54.7% women; mean age,  $62.2 \pm 19.9$  years) were registered in the Savana Manager version 4.0 tool. Out of the total population studied, 544 patients (330 women, 60.7%) were diagnosed with acromegaly. The group without acromegaly included a total of 707,503 patients (387,182 women, 54.7%). Patients with acromegaly were older than those without the disease



**Figure 1** Percentage distribution of the histological types of cancer seen in the group of patients diagnosed with acromegaly.

( $64.2 \pm 16.0$  vs  $60.0 \pm 19.9$  years;  $p < 0.001$ ). The female-to-male ratio was higher in the group of patients with acromegaly 1.5 vs 1.2 ( $p = 0.004$ ). The mean age at diagnosis of acromegaly was  $53.0 \pm 15.8$ . No significant differences were found in relation to age at diagnosis of acromegaly between males and females ( $51.8 \pm 15.7$  vs  $53.8 \pm 15.8$  years, ns).

### Incidence of cancer

The diagnosis of cancer was more frequent in the acromegalic group vs the group without the disease (OR, 2.046; 95%CI, 1.493–2.805). In the group of acromegalic patients, cancer was detected in 42 patients (7.7%), a significantly higher percentage than the one found in the group of patients without acromegaly (27,786 patients, 3.9%;  $p < 0.001$ ). In acromegalic patients, cancer was more frequently observed in males ( $n = 24$ , 57.1%) vs females ( $n = 18$ , 42.9%),  $p = 0.012$ .

In the group of acromegalic patients, 60 cases of cancer were reported. The most common histological types were colorectal cancer ( $n = 16$ ), breast cancer ( $n = 9$ ), lung cancer ( $n = 8$ ), bladder cancer ( $n = 6$ ), and lymphoma ( $n = 6$ ). The percentage distribution of the histological types of cancer seen in the group of patients diagnosed with acromegaly is shown in Fig. 1.

The overall incidence of all types of cancer analyzed in the cohort of patients studied with and without acromegaly is illustrated in Table 1. When comparing the incidence of the different types of cancer analyzed in the overall cohort of patients with and without a diagnosis of acromegaly, we found a significantly higher prevalence only for colorectal cancer (2.9% vs 1.4%,  $p = 0.006$ ; OR 2.029, 95%CI, 1.233–3.334), bladder cancer (1.1% vs 0.3%,  $p = 0.005$ ; OR 3.938, 95%CI, 1.759–8.815), and lymphoma (1.1% vs 0.3%,  $p = 0.009$ ; OR 3.466, 95%CI, 1.548–7.758) in patients with a diagnosis of acromegaly vs patients without this diagnosis.

When the same groups were compared by sex, acromegalic men had a significantly higher prevalence of colorectal cancer (4.7% vs 1.3%,  $p = 0.001$ ; OR, 3.653, 95%CI,

1.935–6.898), bladder cancer (2.8% vs 0.4%,  $p < 0.001$ ; OR, 7.877, 95%CI, 3.491–17.771), breast cancer (0.9% vs 0.2%,  $p = 0.042$ ; OR, 6.280, 95%CI, 1.556–25.346), gastric cancer (0.9% vs 0.1%,  $p = 0.011$ ; OR, 13.085, 95%CI, 3.232–52.987), lymphoma (1.4% vs 0.3%,  $p = 0.037$ ; OR, 4.191, 95%CI, 1.339–13.121), and liver cancer (0.9% vs 0.1%,  $p = 0.012$ ; OR, 12.388, 95%CI, 3.068–50.141) vs non-acromegalic men, while acromegalic women only had a higher prevalence of thyroid cancer (1.2% vs 0.4%,  $p = 0.043$ ; OR, 3.096, 95%CI, 1.154–8.310) compared to non-acromegalic women (Table 1).

### Discussion

The results of this hospital-based study indicate that there is a significant association between the diagnosis of acromegaly and cancer. Although the most common cancers in our acromegalic population were colorectal cancer, breast cancer, lung cancer, bladder cancer, and lymphoma, only colorectal cancer, bladder cancer, and lymphoma they all showed a statistically significant association vs the non-acromegalic population. Moreover, the presence of cancer was observed more commonly in men vs women. Acromegalic men had a significantly higher prevalence of various types of cancer than patients without a diagnosis of acromegaly, whereas acromegalic women only had a higher prevalence of thyroid cancer vs non-acromegalic women. These results confirm the association between cancer and acromegaly showing a predilection for cancer in the male acromegalic population.

The relationship between acromegaly and cancer is not completely clear.<sup>13</sup> Some studies have suggested a possible association between acromegaly and an increased risk of certain types of cancer, such as colon, breast, and prostate cancer.<sup>14–19</sup> However, the results of these studies vary, and there is no definitive consensus in the scientific community regarding this association.<sup>17,20</sup>

Our study, using artificial intelligence and big data techniques in a hospital setting, suggests a higher prevalence of cancer associated with acromegaly, which is consistent with

**Table 1** Prevalence of the different types of cancer analyzed in the cohort of patients studied with and without a diagnosis of acromegaly (Chi-square test).

|                          | Acromegalic patients<br>(n = 544; 330 women) | Non-acromegalic patients<br>(n = 707,503; 387,182 women) | Total (n = 708,047;<br>387,512 women) |
|--------------------------|--|--|---------------------------------------|
| <i>Colorectal cancer</i> |  |  |                                       |
| Females                  | 6 (1.8%)                                     | 5901 (1.5%)  | 5907 (1.5%)                           |
| Males                    | 10 (4.7%)**                                  | 4245 (1.3%)  | 4255 (1.3%)                           |
| Overall                  | 16 (2.9%)**                                  | 10,146 (1.4%)  | 10,162 (1.4%)                         |
| <i>Thyroid cancer</i>    |  |  |                                       |
| Females                  | 4 (1.2%)*                                    | 1527 (0.4%)  | 1531 (0.4%)                           |
| Males                    | 0 (0%)                                       | 394 (0.1%)   | 394 (0.1%)                            |
| Overall                  | 4 (0.7%)                                     | 1921 (0.3%)  | 1925 (0.3%)                           |
| <i>Bladder cancer</i>    |  |  |                                       |
| Females                  | 0 (0%)                                       | 828 (0.2%)   | 828 (0.2%)                            |
| Males                    | 6 (2.8%)***                                  | 1170 (0.4%)  | 1176 (0.4%)                           |
| Overall                  | 6 (1.1%)**                                   | 1998 (0.3%)  | 2004 (0.3%)                           |
| <i>Breast cancer</i>     |  |  |                                       |
| Females                  | 7 (2.1%)                                     | 5667 (1.5%)  | 5674 (1.5%)                           |
| Males                    | 2 (0.9%)*                                    | 481 (0.2%)   | 483 (0.1%)                            |
| Overall                  | 9 (1.4%)                                     | 6148 (0.9%)  | 6157 (0.9%)                           |
| <i>Lung cancer</i>       |  |  |                                       |
| Females                  | 4 (1.2%)                                     | 2902 (0.7%)  | 2906 (0.7%)                           |
| Males                    | 4 (1.9%)                                     | 2304 (0.7%)  | 2308 (0.7%)                           |
| Overall                  | 8 (1.5%)                                     | 5206 (0.7%)  | 5214 (0.7%)                           |
| <i>Gastric cancer</i>    |  |  |                                       |
| Females                  | 0 (0%)                                       | 456 (0.1%)   | 456 (0.1%)                            |
| Males                    | 2 (0.9%)*                                    | 231 (0.1%)   | 233 (0.1%)                            |
| Overall                  | 2 (0.4%)                                     | 687 (0.1%)   | 689 (0.1%)                            |
| <i>Lymphoma</i>          |  |  |                                       |
| Females                  | 3 (0.9%)                                     | 1185 (0.3%)  | 1188 (0.3%)                           |
| Males                    | 3 (1.4%)*                                    | 1084 (0.3%)  | 1087 (0.3%)                           |
| Overall                  | 6 (1.1%)**                                   | 2269 (0.3%)  | 2275 (0.3%)                           |
| <i>Leukemia</i>          |  |  |                                       |
| Females                  | 1 (0.3%)                                     | 948 (0.2%)   | 949 (0.2%)                            |
| Males                    | 0 (0%)                                       | 543 (0.2%)   | 543 (0.2%)                            |
| Overall                  | 1 (0.2%)                                     | 1491 (0.2%)  | 1492 (0.2%)                           |
| <i>Liver cancer</i>      |  |  |                                       |
| Females                  | 0 (0%)                                       | 583 (0.1%)   | 583 (0.1%)                            |
| Males                    | 2 (0.9%)*                                    | 244 (0.1%)   | 246 (0.1%)                            |
| Overall                  | 2 (0.4%)                                     | 827 (0.1%)   | 829 (0.1%)                            |
| <i>Kidney cancer</i>     |  |  |                                       |
| Females                  | 0 (0%)                                       | 335 (0.1%)   | 335 (0.1%)                            |
| Males                    | 1 (0.5%)                                     | 422 (0.2%)   | 423 (0.2%)                            |
| Overall                  | 1 (0.2%)                                     | 757 (0.1%)   | 758 (0.1%)                            |
| <i>Prostate cancer</i>   |  |  |                                       |
| Males                    | 5 (2.3%)                                     | 4214 (1.3%)  | 4219 (1.3%)                           |

\*  $p < 0.05$ .\*\*  $p < 0.01$ .\*\*\*  $p < 0.001$  (acro- vs non-acromegalic patients).

other population-based studies and meta-analyses, which have shown a slightly increased overall risk of cancer in patients with acromegaly vs the general population. The prevalence of cancer in our acromegalic population (7.7%)

was similar to the one found in other studies (6.9%).<sup>18</sup> This prevalence was 1.97 times higher ( $p < 0.001$ ) than the one seen in the non-acromegalic population seen in our hospital. This high prevalence was similar to that found in a U.S.

database study of 1175 acromegalic patients, which showed a 2.6-fold higher prevalence of malignancies vs matched controls.<sup>15</sup>

Acromegaly is a clinical condition associated with elevated levels of GH and IGF-1. Numerous lines of evidence suggest that there is a relationship between the GH/IGF-1 axis and the development and progression of cancer.<sup>13</sup> For example, animals and humans born with a GH receptor (GHR) deficiency have a significantly lower risk of cancer.<sup>4</sup> In contrast, elevated levels of both GH and GHR expression have been found in a variety of human cancers.<sup>21,22</sup> Moreover, several studies have shown that both IGF-1 and its transporter protein IGF-BP3 are closely associated with cancer growth and progression.<sup>23</sup> Elevated circulating IGF-1 levels, even within the reference range, have reportedly been associated with an increased incidence of several cancers in the general population, including colorectal, breast, prostate, and lung.<sup>22</sup>

Our study showed not only a higher prevalence of cancer in acromegalic men vs acromegalic women, but also significant sex differences in cancer types. Acromegalic men showed a significantly higher prevalence of colorectal cancer, bladder cancer, breast cancer, gastric cancer, lymphoma, and liver cancer vs non-acromegalic men, while acromegalic women had a higher prevalence of thyroid cancer vs non-acromegalic women. Some sex differences in the cancer incidence rate in acromegalic patients have been reported too.<sup>23,24</sup> A recent nationwide population-based study conducted in Sweden among 1296 patients with acromegaly showed a standardized incidence ratio (SIR) for cancer of 1.3 (95% CI: 1.1–1.5;  $p < 0.001$ ) in acromegalic males ( $n = 621$ ) and 1.3 (95% CI: 1.0–1.5;  $p = 0.044$ ) in acromegalic females ( $n = 675$ ) vs males and females from the overall population, respectively.<sup>19</sup> Although the interpretation of these results is still to be elucidated, these sex differences may be related to different risk factors, hormonal factors, and age at diagnosis of acromegaly. Males with active acromegaly have higher IGF-1 concentrations vs females at diagnosis, which seem to persist after treatment regardless of age and type of treatment. In addition, females tend to be older than males at the time of diagnosis of acromegaly. The coexistence of higher circulating IGF-1 levels throughout time could have a positive impact on cancer growth.

Most data suggest that the incidence of colorectal cancer is higher in people with acromegaly. The risk is even higher in people with acromegaly who have other risk factors for colorectal cancer, such as older age, family history of colorectal cancer, or obesity.<sup>19</sup> Although it is not known exactly why people with acromegaly have a higher risk for colorectal cancer it is believed that increased GH and IGF-1 levels may stimulate cell proliferation, angiogenesis, mutation risk, inhibition of tumor suppressor genes, and apoptosis, thus promoting a tumor microenvironment and abnormal cell growth in the colon or rectum, which may increase the risk for developing cancer. In our acromegalic population, the prevalence of colorectal cancer was 2.01 times higher vs the non-acromegalic population ( $p < 0.001$ ). This rate was 3.61 times higher in acromegalic men ( $p < 0.001$ ) and 1.2 times higher in acromegalic women (ns) vs non-acromegalic men and women, respectively.

Urothelial carcinoma, also called transitional cell cancer, develops in the urothelial cells lining the renal pelvis, ureters, bladder, and urethra. An association between urothelial carcinoma and acromegaly has recently been reported.<sup>19</sup> In the above-discussed Swedish population-based study, renal and urothelial cancer, evaluated together, showed a SIR of 4.0 (95% CI, 2.3–6.5;  $p < 0.001$ ).<sup>19</sup> Most bladder cancers are known to be urothelial carcinomas. In our study, bladder cancer was also significantly associated with acromegalia (OR, 3.938, 95%CI, 1.759–8.815;  $p = 0.005$ ). There is evidence that IGF-1 may play an important role in the development of bladder cancer.<sup>25–27</sup> In addition, it has been reported that IGF-1 may be involved in the development and progression of bladder cancer, as increased local IGF-1 expression has been observed in bladder cancer biopsies vs normal urothelial tissue. A close relationship between IGF-1 expression levels and disease recurrence has also been reported, as well as between IGF-1 receptor (IGF-1R) expression levels and tumor grade, tumor differentiation, and disease recurrence. In our study, as in colorectal cancer, the association with bladder cancer was observed only in acromegalic men. Both types of cancer are known to be more common in men vs women from the overall population. It is possible that the presence of higher IGF-1 levels in acromegalic males plays a positive role in the development of bladder cancer. However, further studies are needed to confirm this association and to establish an etiopathogenic relationship between bladder cancer and acromegaly in men.

In recent years, several hematologic neoplasms have been reported in patients with acromegaly, including polycythemia vera, essential thrombocytosis, acute and chronic myeloid leukemia, multiple myeloma, and non-Hodgkin's lymphoma.<sup>28</sup> The coexistence of both diseases may be coincidental or indicate the existence of a common pathogenic mechanism. Given the rarity of both diseases, it is difficult to establish a causal relationship between elevated IGF-1 levels associated with acromegaly and the development of hematologic malignancies.<sup>28</sup> In our acromegalic population we did not find an increased incidence of leukemia; however, lymphoma was significantly associated with acromegaly (OR, 3.466, 95%CI, 1.548–7.758;  $p = 0.009$ ), which is similar to what has been reported for colorectal cancer and bladder cancer, an association significant only in acromegalic men. More studies conducted in a larger number of patients are needed before these findings can be confirmed and a cause–effect relationship between these 2 entities can be established.

Some studies have reported an increased incidence of thyroid, breast, and prostate cancers in patients with acromegaly. In the recently published Swedish population-based study among 1296 acromegalic patients, the risk of thyroid, breast, and prostate cancer was not different from that of the overall population.<sup>19</sup> In our acromegalic population, we found a significant association with thyroid cancer in females only, and with breast cancer in males only, while no association was ever found with prostate cancer. The results of the association study may be affected by the limited number of cases observed.

The results of our study show a significant association between gastric cancer and acromegaly, especially in

men. This association has been documented in a recent meta-analysis, which found that acromegaly was associated with an increased risk of gastric cancer (standardized incidence ratio SIR, 3.09; 95%CI, 1.47–6.50), even higher than the risk associated with colorectal and anal cancer (SIR, 1.95; 95%CI, 1.32–2.87).<sup>29</sup> Active detection of colorectal lesions and resection of premalignant lesions in patients with acromegaly could explain this high prevalence of colonic malignant lesions. Our findings, together with those of this meta-analysis, suggest the need to consider a specific screening program for gastric neoplasms in patients with acromegaly on an individual basis. It would also be interesting to have specific data on the number of colonoscopies performed per patient, the availability of fecal occult blood test results, the type of gastric cancer reported, information on *Helicobacter pylori* infection, and presence of anemia as a red flag for gastroscopy. All these data could be interesting for future research focused on the study of acromegaly-related GI neoplasms.

AI is a methodology of computer systems that uses algorithms to relentlessly process data, automatically learn and understand its meaning, generate computer models, and identify the best predictive features that are present in the training data. With the rapid advancement of computer technology, the use of AI in studying and treating patients with pituitary adenomas has grown exponentially, including facial imaging, radiomics, pathological studies, and analyzing EHRs with both textual information and medical Images.<sup>30</sup> As far as we know, this is the first study ever conducted to use free text from EHRs as a source of data for NPL extraction to study the incidence of cancer in patients with acromegaly. This real-world study provides clinical information suggesting the association of acromegaly with cancer in general and with colorectal, urothelial (bladder), and hematologic (lymphoma) cancers in particular, especially in men. It also confirms the association with thyroid cancer in women. Finally, our study did not show a statistically significant association with breast and prostate cancers.

The main strength of our research lies in the extraction of real-world data and the achievement of a substantial and unbiased sample size. By implementing big data techniques and using Savana EHRead technology, we could draw a large amount of information, effectively read, process, and organize unstructured text from EHRs, and ultimately transform it into structured data. This allowed us to effectively understand, process, organize the unstructured textual content of EHRs, and ultimately transform it into organized data. Dataset we analyzed included diagnoses from more than 700,000 patients across the study period. This ensured the reliability and objectivity of the conclusions drawn from real-world clinical practice.

The main limitations of our study were the impossibility of analyzing differences based on acromegaly disease progression, histopathological characteristics and tumor size, circulating levels of GH and IGF-1, use of drugs, such as somatostatin analogs, dopaminergic agonists, GH receptor antagonists and pituitary surgery-induced changes with or without adjuvant radiotherapy. The database included the date of the first report in which the variable appeared; however, we cannot be completely sure that this date corresponds to the exact date of diagnosis of both acromegaly and cancer. Consequently, we cannot establish a definitive

temporal relationship between the period of exposure to elevated IGF-1 levels and the diagnosis of cancer. Furthermore, we must recognize as a limitation of the study a possible bias given the difference in age and sex distribution of the patients in the 2 study samples.

Another limitation is that Savana only extracts information from EHRs without generalizing inferences. Therefore, the system may not be able to detect subtle patterns or draw conclusions that are not explicitly documented in the EHR. In addition, the accuracy of the results obtained by the system depends on the accuracy of the diagnoses recorded in the patient's health record. Finally, the cohort analyzed was studied in our own hospital, which may not be representative of the overall population. In addition, it is important to keep in mind that our hospital is a tertiary referral center for pituitary pathology, as evidenced by the high number of patients with acromegaly studied. It is possible that a significant number of patients were referred to us for surgery and subsequently monitored in these patients' referring hospitals. This situation may lead to misclassification of cases in the cancer-free group due to lack of adequate follow-up and may induce bias in the analysis of the long-term relationship between acromegaly and cancer.

In conclusion, this NLP and big data analysis-based study shows that acromegaly is generally associated with cancer and this association is more common in men vs women. Colorectal cancer, bladder cancer and lymphoma are the main associated neoplasms. Other associated cancers include thyroid cancer in women and stomach, liver and breast cancer in men. The concordance of our results with most previously published studies provides valuable insights into the potential application of AI methods in the analysis of real-world data and information.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

1. Fleseriu M, Langlois F, Lim DST, Varlamov EV, Melmed S. Acromegaly: pathogenesis, diagnosis, and management. *Lancet Diabetes Endocrinol*. 2022;10:804–26.
2. Tirosh A, Shimon I. Complications of acromegaly: thyroid and colon. *Pituitary*. 2017;20:70–5.
3. Ruchala M, Wolinski K. Health-related complications of acromegaly-risk of malignant neoplasms. *Front Endocrinol*. 2019;10:268.
4. Guevara-Aguirre J, Peña G, Acosta W, Pazmiño G, Saavedra J, Soto L, et al. Cancer in growth hormone excess and growth hormone deficit. *Endocr Relat Cancer*. 2023;30:e220402.
5. Iglesias P, Benavent M, López G, Arias J, Romero I, Díez JJ. Hyperthyroidism and cardiovascular disease: an association study using big data analytics. *Endocrine*. 2024;83:405–13.
6. Díez JJ, Cabrera L, Iglesias P, Benavent M, López G, Argüello G, et al. Thyroid carcinoma in elderly people: characterization using big data tools. *Endocrinol Diabetes Nutr*. 2023;70:179–88.
7. 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 Nov 11;0, <http://dx.doi.org/10.18176/jiaci.0765>.

8. 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.
9. Morena D, Fernández J, Campos C, Castillo M, López G, Benavent M, et al. Clinical profile of patients with idiopathic pulmonary fibrosis in real life. *J Clin Med*. 2023;12:1669.
10. Medrano IH, Guijarro JT, Belda C, Ureña A, Salcedo I, Espinosa-Anke L, et al. Savana: re-using electronic health records with artificial intelligence. *Int J Interact Multimed Artif Intell*. 2018;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.
12. Baeza-Yates RA, Ribeiro-Neto B. Modern information retrieval. New York: ACM Press; 1999.
13. Boguszewski CL, Boguszewski MC, da S. Growth hormone's links to cancer. *Endocr Rev*. 2019;40:558–74.
14. Dal J, Leisner MZ, Hermansen K, Farkas DK, Bengtson M, Kistorp C, et al. Cancer incidence in patients with acromegaly: a cohort study and meta-analysis of the literature. *J Clin Endocrinol Metab*. 2018;103:2182–8.
15. Fleseriu M, Barkan A, Del Pilar Schneider M, Darhi Y, de Pierrefeu A, Ribeiro-Oliveira A, et al. Prevalence of comorbidities and concomitant medication use in acromegaly: analysis of real-world data from the United States. *Pituitary*. 2022;25:296–307.
16. Jenkins PJ. Acromegaly and cancer. *Horm Res*. 2004;62 Suppl. 1:108–15.
17. Terzolo M, Reimondo G, Berchialla P, Ferrante E, Malchiodi E, De Marinis L, et al. Acromegaly is associated with increased cancer risk: a survey in Italy. *Endocr Relat Cancer*. 2017;24:495–504.
18. Ucan B, Kizilgul M, Karci AC, Duger H, Erkam Sencar M, Imga NN, et al. The prevalence of cancer and its relation to disease activity in patients with acromegaly: two centers' experience. *Endocr Pract*. 2021;27:51–5.
19. Esposito D, Ragnarsson O, Johannsson G, Olsson DS. Incidence of benign and malignant tumors in patients with acromegaly is increased: a nationwide population-based study. *J Clin Endocrinol Metab*. 2021;106:3487–96.
20. Maione L, Brue T, Beckers A, Delemer B, Petrossians P, Borson-Chazot F, et al. Changes in the management and comorbidities of acromegaly over three decades: the French Acromegaly Registry. *Eur J Endocrinol*. 2017;176:645–55.
21. Perry JK, Liu D-X, Wu Z-S, Zhu T, Lobie PE. Growth hormone and cancer: an update on progress. *Curr Opin Endocrinol Diabetes Obes*. 2013;20:307–13.
22. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. *Lancet Lond Engl*. 2004;363:1346–53.
23. Durmuş ET, Atmaca A, Çolak R, Durmuş B. Cancer prevalence and cancer screening in patients with acromegaly: a single center experience. *Endocrine*. 2022;77:363–71.
24. Gullu BE, Celik O, Gazioglu N, Kadioglu P. Thyroid cancer is the most common cancer associated with acromegaly. *Pituitary*. 2010;13:242–8.
25. Eich M-L, Tregnago AC, Faraj SF, Palsgrove DN, Fujita K, Bezerra SM, et al. Insulin-like growth factor-1 receptor expression in upper tract urothelial carcinoma. *Virchows Arch Int J Pathol*. 2019;474:21–7.
26. Mourmouras N, Philippou A, Christopoulos P, Kostoglou K, Gri-vaki C, Konstantinidis C, et al. Differential expression of IGF-I transcripts in bladder cancer. *Anticancer Res*. 2018;38:3453–9.
27. Xie Q-X, Lin X-C, Zhang M-F, Han C-X, Guo Y-H. Expression of IGF-I and IGF-IR in bladder cancer. *Ai Zheng Aizheng Chin J Cancer*. 2004;23:707–9.
28. Gupta P, Dutta P. Co-occurrence of acromegaly and hematological disorders: a myth or common pathogenic mechanism. *Integr Med Int*. 2017;4:94–100.
29. Xiao Z, Xiao P, Wang Y, Fang C, Li Y. Risk of cancer in acromegaly patients: an updated meta-analysis and systematic review. *PLOS ONE*. 2023;18:e0285335.
30. Dai C, Sun B, Wang R, Kang J. The application of artificial intelligence and machine learning in pituitary adenomas. *Front Oncol*. 2021;11:784819.