The correlation between small papillary thyroid cancers and gamma radionuclides Cs-137, Th-232, U-238 and K-40 using spatially-explicit, register-based methods

Haytham Bayadsi [a,*], Paul Van Den Brink [b], Mårten Erlandsson [c], Soffia Gudbjornsdottir [d], Samy Sebraoui [d], Sofi Koorem [a], Pär Nordin [a], Joakim Hennings [a], Oskar Englund [b]

[a] *Department of Surgical and Perioperative Sciences/Surgery, Umeå University, Sweden*
[b] *Department of Natural Sciences, Design and Sustainable Development, Mid-Sweden University, Östersund Campus, Sweden*
[c] *Department of Aquatic Resources, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden*
[d] *Institute of Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden*

ARTICLE INFO

ABSTRACT

A steep increase of small papillary thyroid cancers (sPTCs) has been observed globally. A major risk factor for developing PTC is ionizing radiation. The aim of this study is to investigate the spatial distribution of sPTC in Sweden and the extent to which prevalence is correlated to gamma radiation levels (Caesium-137 (Cs-137), Thorium-232 (Th-232), Uranium-238 (U-238) and Potassium-40 (K-40)) using multiple geospatial and geo-statistical methods. The prevalence of metastatic sPTC was associated with significantly higher levels of Gamma radiation from Th-232, U-238 and K-40. The association is, however, inconsistent and the prevalence is higher in densely populated areas. The results clearly indicate that sPTC has causative factors that are neither evenly distributed among the population, nor geographically, calling for further studies with bigger cohorts. Environmental factors are believed to play a major role in the pathogenesis of the disease.

1. Introduction

Thyroid cancer (TC) is the most common form of cancer in the endocrine system, with papillary thyroid cancer (PTC) being its most common subtype (Dahlberg, 2022). The incidence of PTC has been increasing steeply over recent decades, and in the last years starting to stabilize and even decrease in some areas (Powers et al., 2019; National Health Institute NHI, 2022). The whole reason for the increment, plateau and decrease is still unknown. Many speculate that the small papillary thyroid cancers (sPTCs) ($\leq$ 20 mm in size) and papillary thyroid microcarcinomas (PTCMs) ($\leq$ 10 mm in size) are responsible for most of the increase in incidence rate (Siegel et al., 2019; Thyroid cancer 2021). sPTCs are classified as extremely low risk or low risk, and both have a favourable prognosis (Thyroid cancer 2021; Molinaro et al., 2021). Yet some of these sPTCs have more aggressive properties causing

lateral lymph node metastases, distant metastases (most commonly to the lungs), recurrences or even death. The increase in incidence rate has been partially explained by diagnostic advances, such as the increased use of thyroid gland ultrasound with fine-needle aspiration cytology (Vaccarella et al., 2016) allowing the detection of small non-palpable or visible tumours as well diagnosing the patients at an early stage. However, this alone cannot explain the increasing incidence, as analysis of the data has shown increase in all PTC incidence rates (Lim et al., 2017), even the ones above 20 mm, therefore other risk factors must be considered. One etiologically established culprit is internal and external exposure to ionizing radiation, which is considered a major risk factor for TC in general and PTC specifically (Nikiforov, 2010; Albi et al., 2017).

Various studies have shown an increment in PTC cases in the areas exposed to nuclear reactor or weapons explosions, or populations

---

exposed to external beam radiation to the head and neck areas, especially at a young age (Morton et al., 2021; Wakeford, 2004; van Gerwen et al., 2020; Kazakov et al., 1992; Takahashi et al., 2003). Increased PTC risk has also been observed in survivors of the atomic bombs in Nagasaki and Hiroshima (Furukawa et al., 2013; Imaizumi et al., 2015). A study based on autopsy reports by Hayashi et al. (Hayashi et al., 2010) reported an increased prevalence of PTMC in the population of atomic bomb survivors in Japan in 1945, relating PTMC to radiation exposure. One of the most renowned examples is the Chernobyl nuclear reactor accident in 1986, resulting in thousands of new PTC cases in the population of children and adolescents at that time who were exposed to substantial doses of radiation (United-Nations 2013). However, much larger populations in other countries were exposed to the nuclear fallout following the accident, Sweden being one of those (Moberg, 2001). In Sweden, the largest fallout of radioactive substances occurred in parts of the eastern coast: one area measured ten times the normal level directly following the incident (Swedish Radiation Safety Authority SSM, 2017). The highest levels of Caesium 137 (Cs-137) activity exceeded 85 kBq/m^2 and in some spots up to 200 kBq/m^2 (SSM 1986). Even though most areas in Sweden received very low radiation doses as a result of the fallout, an Uppsala University study showed a 5-fold higher overall cancer risk development for all cancer forms including PTC in the population with the highest exposure to the radioactive Cs-137 in Sweden (Alinaghizadeh et al., 2016). Cs-137 is a radioactive isotope of Caesium, produced by nuclear fission for use in medical devices and gauges. It is a biproduct of the nuclear fission process in nuclear reactors and weapons (National Center for Environmental Health NCEH, 2018). The radiation that exists in Sweden today after the Chernobyl accident is mainly due to the 30-year long half-life of Cs-137 (Swedish Radiation Safety Authority SSM, 2017).

Aside from these specific exposure scenarios, the general population is also exposed to background radiation from Thorium-232 (Th-232), Uranium-238 (U-238) and Potassium-40 (K-40). Th-232 is a naturally occurring radioactive metal found in bedrock, soil, and water (National Cancer Institute NCI, 2022). It can be used to make ceramics, gas lantern mantles, welding rods and fuel for generating nuclear power as an alternative to uranium. Th-232 has been linked to an increased risk of developing liver, gallbladder and blood cancer and is classified by the International Agency for Research on Cancer (IARC) as carcinogenic to humans (Agency for Toxic Substances and Disease Registry ATSDR, 2019). To the best of our knowledge, there are no previous studies investigating the association between Th-232 levels and PTC incidence (National Cancer Institute NCI, 2022; Agency for Toxic Substances and Disease Registry ATSDR, 2019).

U-238 is a naturally occurring radioactive element with a wide distribution in soil and higher concentrations in certain rock formations and water (van Gerwen et al., 2020). Uranium has 17 known isotopes but only U-234, U-235, and U-238 are found in the environment due to their long half-lives (250 thousand, 700 million and 4.5 billion years respectively) (Taylor and Taylor, 1997). U-238 is released into the environment through wind and water erosion, as well as mining, milling or other forms of uranium processing (van Gerwen et al., 2020). U-238 can be absorbed into the body through ingestion of contaminated food or drink, or inhalation of uranium-containing dust particles or aerosols (Taylor and Taylor, 1997). All uranium isotopes are radioactive, emitting alpha particles causing radiotoxicity (Taylor and Taylor, 1997). Therefore, chemical toxicity is the limiting factor determining the current US Environmental Protection Agency (EPA) maximum allowable contaminant level of 30 μg/L in drinking water (van Gerwen et al., 2020). Increased thyroid cancer incidence has been documented in volcanic regions where increased urinary uranium concentrations have been found, although the underlying mechanism is still not clearly understood (van Gerwen et al., 2020).

K-40 is an unstable and radioactive naturally occurring isotope of potassium. It has a very long half-life of 1.25 billion years (Connor, 2019) and makes up about 0.012 % of the total amount of potassium

found in nature. The radioactive decay used to calculate the potassium content is the gamma radiation from the decay of K-40 into argon-40. This decay branch represents 10.7 % of the decay (Ruedas, 2017). After Th-232 and U-238, K-40 ranks 3rd as a source of natural radioactivity contributing to earth heat (radiogenic heat) (Connor, 2019; SGU, 2019). K-40 is the largest source of natural radioactivity in animals, including humans, because its traces are found in all potassium (Connor, 2019). A 70 kg human body contains 0.0164 gs of K-40, equalling an annual effective dose of 0.165 mSv, as compared to the estimated total effective dose from natural radionuclides in a human body of 0.3 mSv (Rao, 2012).

Geographic Information Systems (GIS) is a collective term for software that helps us capture, analyse, and visualize geographical information (Dempsey, 2002; Kulldorff, 1999). Maps allow us to show, understand, interpret, and visualize data in ways that can reveal relationships, patterns, trends, and geographical deviations. GIS gives us the opportunity to explore environmental risk factors (Evers, 2022) and to handle data covering big populations. This enables large scale geospatial assessments, which can be used to compute geostatistical correlations between disease prevalence and many potential environmental risk factors. GIS is primarily used in urban planning and social science and has not been yet used to any great extent in medical epidemiological research (Aneja et al., 2011). This method is innovative and could draw attention to as yet unknown factors influencing thyroid cancer, and may potentially prove useful within other fields of medical research.

Unexplained regional differences in the distribution and increase of sPTC have been listed in Sweden (Bayadsi et al., 2020). Early detection of more aggressive forms of thyroid cancer is essential for proper surgical and oncological treatment of these patients. The aim of this study is to (i) explore spatial patterns in the prevalence of sPTC, using multiple methods for disease mapping, and (ii) investigate whether geological differences in sPTC prevalence in Sweden are correlated to geographical factors in the surrounding environment such as the deposit of Cs-137 fallout following the Chernobyl nuclear powerplant accident, or the levels of the naturally occurring radioactive materials Th-232, U-238 or K-40.

2. Materials and methods

2.1. Patient selection

The study population originates from a nationwide cohort of patients with sPTC, all of whom were registered in the validated and prospectively maintained Scandinavian Quality Register for Thyroid, Parathyroid and Adrenal Surgery (SQRTPA) between 2011 and 2015. The SQRTPA national register was established in 2004 and is the world's first quality register for endocrine surgery. It covers almost 100 % of thyroid surgeries in Sweden. It is validated against the National Patient Register and is one of the few registers to have an internal quality audit that randomly checks the operating centres every year (Nordenström, 2021).

The inclusion criteria comprised a primary diagnosis of sPTC, defined as T1 (tumour ≤ 20 mm). These T1 cases were then divided into subgroups depending on the lymph node status. The Nx group included patients where it was not possible to assess six or more lymph nodes. The N0 group included patients with no lymph node metastases, and the N1 group included patients with regional lymph node metastases indicating a more advanced disease. The inclusion was based on the Tumour–Node–Metastasis classification (TNM) 7th edition (AJCC, 2010). Nx-classified tumours are clinically and prognostically treated as N0 groups and therefore were included in the N0 group. The Chi-squared test was used to compare the difference between genders. The Mann-Whitney U test was used to compare differences in age and tumour size. The statistical analyses were performed using SPSS statistics version 28 (IBM Corporate, Armonk, NY, USA).

Of the 14,827 thyroid surgeries that were performed during the study period, 933 patients were identified according to the inclusion

criteria. Twenty-seven were excluded because they had tumours larger than 20 mm (T-stadium > 1 as defined in the inclusion criteria) at the first surgery or at a complementary surgery. Eighty patients were excluded since we were unable to find their residential addresses and an additional 13 patients were excluded due to double registration. One patient was lost during data collection, leaving 812 patients to be included in the study (Table 1).

2.2. Spatial data

Three sources of natural gamma radiation were studied: Uranium-238 (U-238), Thorium-232 (Th-232) and Potassium-40 (K-40). To enable spatially explicit quantifications, we used national datasets of concentrations of U-238 (ppm eU), Th-232 (ppm eTh), and K-40 (%), respectively, with 200 m resolution (Fig. 1), based on gamma spectrometric measurements from airplanes at low altitude (30–60 m) (AJCC 2016).

Additionally, we studied the deposit of Caesium-137 (Ca-137) fallout after the Chernobyl 1986 incident. For this purpose, we used a national 200 m dataset of estimated Cs-137 deposits on the ground in May 1986 (kBq/m2), based on recurring radiation measurements by airplanes done by the Swedish Radiation Safety Agency (SSM).

Population data from December 31st, 2019, were retrieved from the Swedish Tax Agency (Skatteverket) and Statistics Sweden (SCB), with two different resolutions, 1000 m (Fig. 1) and 100 m (Statistics Sweden SCB, 2022).

2.3. Spatial analysis

The analysis was divided into two parts. First, we conducted a geospatial analysis of the distribution of cases across Sweden, to identify areas where sPTC is over- or under-represented. Second, we conducted statistical analyses to determine the extent to which prevalence can be explained by population density and levels of gamma radionuclides from different sources, respectively.

2.3.1. Spatial distribution of cases

Every person who lives in Sweden receives a ten-digit Swedish personal identity number (PIN). The Swedish PIN covers almost 100 % of the Swedish health care system and can be a useful link between medical registers (Ludvigsson et al., 2009). Residential addresses were retrieved through the Swedish Tax Agency (Skatteverket) and Statistics Sweden (SCB) crosslinking pseudonymised data from the SQRTPA. Google's Application Program Interface (API) was then used to convert the addresses of the 812 patients to geographical coordinates.

Several methods were then tested to assess the distribution of cases,

Table 1
Patient characteristics.

	N1	N0 (including Nx)	Total	P-value
Total number of patients	186 (23%)	626 (77%)	812 (100%)	
Sex				
Male	43 (34%)	106 (17%)	149 (18%)	
Female	143 (66%)	520 (83%)	663 (82%)	
Male/female ratio	1:3.3	1:4.9	1:4.5	0.066
Age (years)				
Mean	43.3	48.4	47.3	<0.001
Median	42	49	47	
Range	17–68	16–81	12–84	
Tumour size (mm)				
Mean	10.7	8.2	8.9	<0.001
Median	11	7	8	
Range	1–20	1–20	1–20	

as follows:

2.3.1.1. Aggregated prevalence. The number of cases per 10,000 inhabitants was calculated for multiple geographical areas, of which two are presented here. First, the GIS software QGIS (OSGeo, 2022) was used to count the number of cases in each county and the total population, using an official county polygon dataset. The field calculator was then used to calculate the corresponding prevalence. Second, the same was calculated for each cell in the 1000 m population dataset using the GIS software GRASS GIS (tool: v.vect.stats). As a comparison, the expected number of cases in each cell was calculated based on the population in each cell and the national prevalence. The latter was calculated based on the total population as defined by the same population dataset.

2.3.1.2. Spatial prevalence patterns based on spatial scan statistics. A spatial scan statistics analysis was conducted using SaTScan™ v. 10.1 (Kulldorff, 1997). Spatial clusters were identified where the occurrence of cases are higher and lower than expected, respectively, using the discrete Poisson model. As input data, we used the official 1000 m population dataset described above. Removing all cells with zero population resulted in a total of 104 540 cells. Each observed case was connected to a population cell and was spatially represented by its centroid. The significance of the clusters was tested by running 999 random simulations and comparing the results with the actual geographical distribution of cases. Different cluster sizes were tested with respect to percent of the total population at risk (2–15 %). Otherwise, the default settings of the software were used. The test did not consider the temporal distribution of cases or the age distribution of the population.

2.3.1.3. Spatial prevalence patterns based on heatmaps. Heatmaps can be used to identify "hotspots" within a geographical area, and thus can be potentially useful for indicating areas where sPTC is over- or under-represented. For this purpose, we used the Kernel Density Estimation tool in QGIS with standard quartic kernel shape, a pixel size of 100 m, and a radius of 20,000 m. First, we attempted to indicate hotspots, or over-representation, of sPTC, with the "weight from field" set to the calculated prevalence within a circle that extends to the nearest other sPTC patient. Then, an attempt was made to identify areas where sPTC is under-represented, with the "weight from field" set to the population within a circle that extends to the next nearest sPTC patient multiplied by radius of the circle.

2.3.1.4. Spatial prevalence patterns based on k-nearest-neighbour (KNN) analysis. Spatial patterns in prevalence were also assessed using a k-nearest-neighbour (KNN) approach (Uddin and al., 2022; Uddin et al., 2019). First, a national grid with a resolution of 100 m was created and the population and number of patients were calculated for each cell. Second, a circle around each cell was expanded until it contained 10,000 and 50,000 people, respectively. Third, based on the number of cases in the circle, the prevalence of sPTC among the 10,000 and 50,000 nearest neighbours to each cell was calculated, respectively. The KNN analyses were performed using the software Equipop (Östh, 2014).

2.3.2. Statistical analysis

2.3.2.1. Spatial distribution of sPTC patients within the population and across Sweden. To determine the extent to which sPTC is evenly distributed within the population, we conducted three statistical analyses:

To test the degree to which prevalence is connected to population density, we first divided all population cells into deciles (based on population values). We then calculated the number of patients residing in population cells belonging to the respective deciles. Finally, we conducted 1000 simulations, where 812 cases were randomly assigned
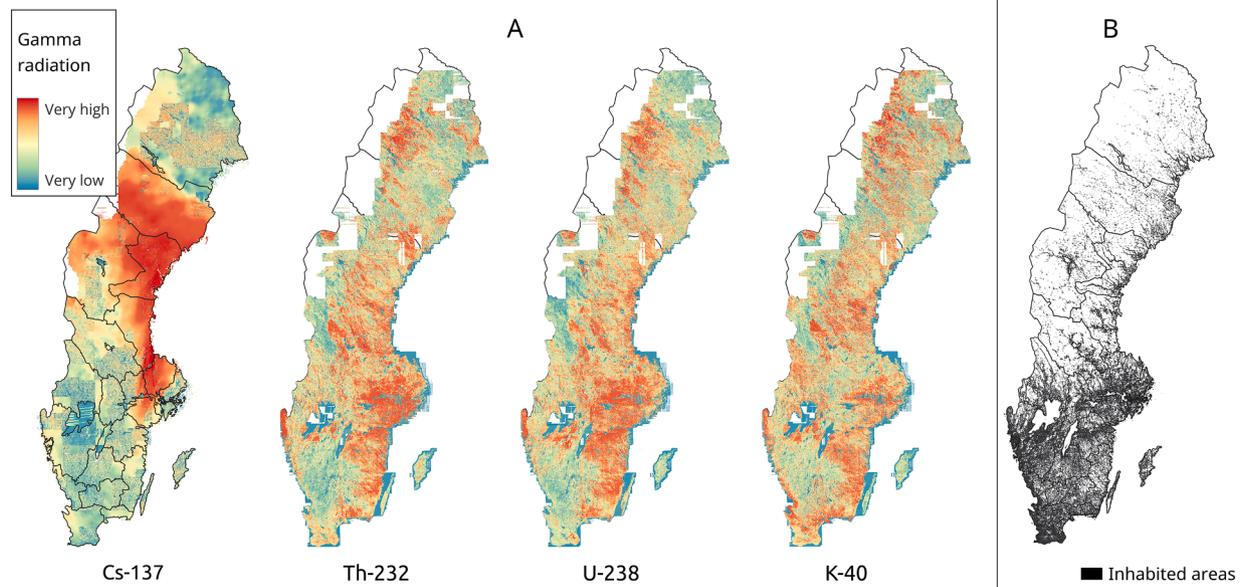
Fig. 1. (A) Radiation levels displayed as percentiles for the respective datasets. Note that the maps are not comparable in terms of absolute radiation levels; (B) The 1000 m population dataset.

to population cells based on the rule that the probability of a case is directly proportionate to the population density. That is, the probability of assigning a case to a cell with population value 1000 is 100 times greater than the probability of assigning a case to a cell with population value 10. We then counted the number of simulated cases in population cells belonging to the previously identified deciles in each of the 1000 simulations, and calculated the resulting median, 2nd quartile, and 3rd quartile.

We then conducted a chi-square goodness-of-fit test where the observed number of cases in each population density decile was compared with the expected number of cases, estimated as the product of population in each decile and the national prevalence.

Finally, we conducted a chi-square goodness-of-fit test where the observed number of cases in each county was compared with the expected number of cases, estimated as the product of population in each county and the national prevalence.

2.3.2.2. Comparison of sPTC patients and the entire Swedish population in terms of gamma radiation levels at the location of residence. To identify differences between the patients and the total population, in terms of gamma radiation levels at the location of residence, we first assigned radiation values for Cs-137, U-238, Th-232, and K-40, for each patient (using GRASS GIS: w.what.rast), using point data for patients and the different radiation raster datasets. The output was saved as a csv file for further statistical analysis using a stand-alone Python script, as described below. The same procedure was then repeated using the 100 m population vector dataset instead of patient point data.

The csv data for the cases consisted of a table with one row per patient and multiple columns for patient information, including whether sPTC was metastatic or not, and radiation levels for the assessed radiation sources. Population data, however, consisted of a frequency table with one row for each 100 m population square, and columns for total population, radiation levels, and the number of cases subdivided into metastatic and non-metastatic disease. To facilitate a statistical comparison between the sample and the population, the population table was converted to a raw dataset with one individual per line, using Python.

To illustrate the distribution of radiation levels for the patients and the entire population, respectively, box plots were created using the Pandas boxplot function (whiskers showing the 10th and 90th

percentiles). Two-sample t-tests were then conducted for the sample and population datasets to assess how their respective means differ. The ratio of variance (patients: population) was calculated and considered acceptable (0.59 for Cs-137 and near 1.0 for other radiation sources). The t-tests were then computed using the scipy.stats.ttest_ind function in Python.

2.3.2.3. Connections between radiation levels and spatial prevalence patterns based on KNN analysis. For each 100 m population cell in Sweden (see 2.3.1.4), the average radiation value was extracted from the four radiation maps (Fig. 1). We then used statistical modelling to explore how the prevalence among the 10,000 and 50,000 nearest neighbours to each cell relate to the four radiation variables. Because of count data and non-linear patterns, Poisson distributed generalized additive models (GAM) were used via the R package mgcv (Wood, 2017). The model was formulated as follow:

Prevalence among k nearest neighbours $\sim$ *offset(log(pop))* $+ f_1(Cs\text{-}137)$ $+ f_2(Th\text{-}232) + f_3(U\text{-}238) + f_4(K\text{-}40)$ *- where* f_i *are thin plate splines.*

3. Results

3.1. Spatial prevalence

The 812 cases of sPTC are not evenly distributed according to population density. sPTC is over-represented in cells (1000 m) that are densely populated and under-represented in sparsely populated cells (Fig. 2). The goodness-of-fit test confirmed that there is a significant difference between observed and expected number of cases in deciles of population density (chi2 = 21.2; CV0.01 = 21.7; p = 0.0117).

At the county level, there are notable differences in prevalence between counties: from 0.277 (Dalarna) to 1.153 (Jämtland) cases per 10,000 inhabitants (Fig. 3A; cf. the national prevalence of ca. 0.81). The goodness-of-fit test confirmed significant differences between observed and expected number of cases at county level (χ^2 = 124; CV0.01 = 38; p = 3.5 × 10–17). sPTC cases are thus not evenly distributed between counties.

The results from the spatial scan statistics provided similar results for 2, 5, and 15 % for population-at-risk. The size of the clusters correlates with the threshold and 5 % was considered a reasonable compromise for both higher-than-expected and lower-than-expected prevalence. The
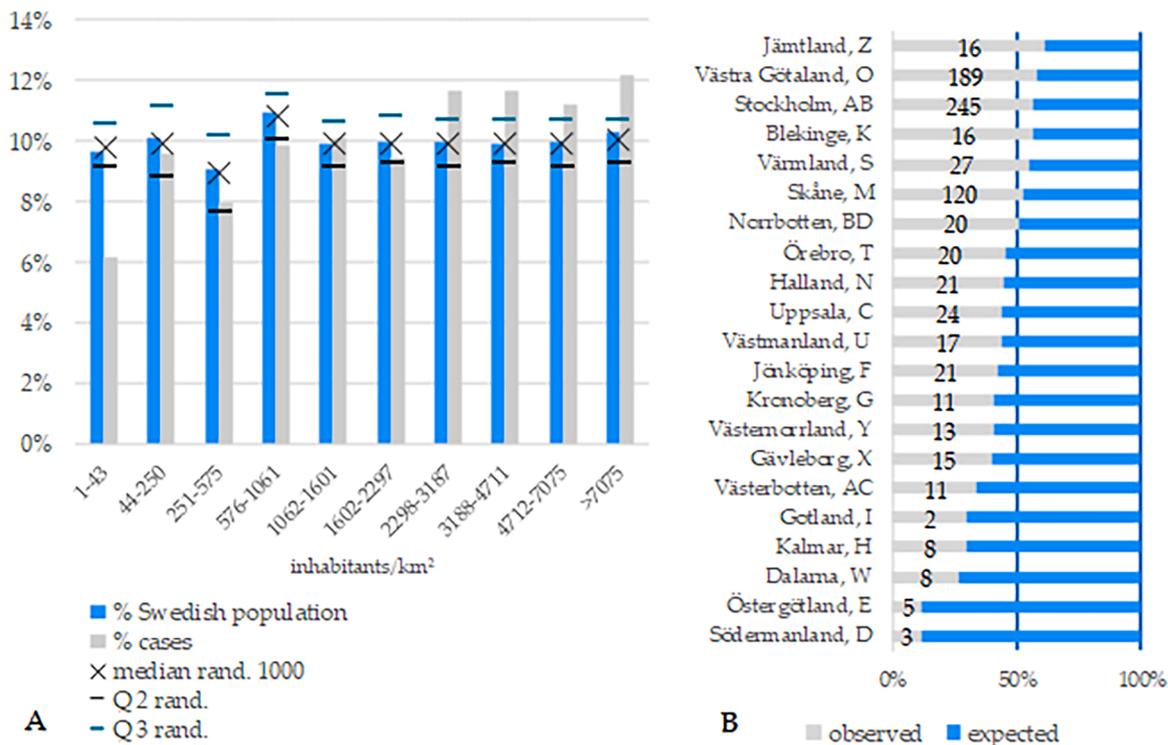
Fig. 2. (A) Share of (i) total population, (ii) sPTC cases, and (iii) simulated sPTC cases, where the probability of a case is directly proportionate to the population, divided into deciles of total population per cell (x-axis). "Median rand" refers to the median share of total simulated cases per cell in the respective population density intervals, and "Q2 rand" and "Q3 rand" refer to the corresponding second and third quartile, respectively. (B) Expected and observed cases in each county, ordered from more than expected (>50%) to less than expected.

analysis identified two clusters with statistically significant ($p <0.05$) higher-than-expected prevalence, in Gothenburg and southern Stockholm, and three additional insignificant clusters (Fig. 3B). Apart from the (insignificant) cluster in county Y, all clusters are located in counties that have higher prevalence than the national average, (Fig. 3A and B). Three clusters of statistically significant lower-than-expected prevalence were identified: south-west of Stockholm, east of lake Vättern, and north-west of Stockholm in county W/X. An additional seven insignificant clusters were identified (Fig. 3C). With the exception of the (insignificant) clusters located in county O and S, all clusters are located within counties that have lower prevalence than the national average, (Fig. 3A and C).

There are notable similarities between the results from the spatial scan statistics analysis and the heatmap analysis. Most areas with higher-than expected prevalence from the heatmap analysis are located within the corresponding clusters from the spatial scan statistics analysis (Fig. 3B). The only cluster from the spatial scan statistics analysis that is not confirmed by the heatmap analysis is (the significant) area c (Fig 3B). Similar patterns can be seen for areas with lower-than-expected prevalence (Fig. 3C).

The analysis of prevalence among the nearest neighbours to each 100 m cell results in similar patterns as above and further supports that there is an over-representation of sPTC in densely populated areas near major cities, but it also suggests elevated risks in several rural areas. The level of detail (in spatial terms) is substantially higher compared with the spatial scan statistics analysis, but the results cannot be verified by statistical tests, or be otherwise validated due to the limited number of cases, and should therefore be considered as indicative. Using the nearest 10,000 or 50,000 neighbours' results in similar spatial patterns, but the former results in a clearer delineation of hotspots than the latter. The analysis of prevalence among the nearest neighbours to each 100 m cell reveals some additional detail, indicating that there is an over-representation of sPTC in densely populated areas near the major

cities, but also in some rural areas (Fig. 3D and E).

3.2. Relationships between sPTC and gamma radiation

Differences between our study cohort and the entire population of Sweden in terms of radiation levels cannot be determined by visual assessment of boxplots (Fig. 4). There are, however, some notable differences between metastatic and non-metastatic patients in this respect, indicating that patients with metastatic sPTC are generally subject to higher levels of gamma radiation from U-238 and Th-232 and lower levels of radiation from Cs-137 (Fig. 4).

The statistical analysis reveals some significant, and unexpected, differences between the groups (Table 2):

- sPTC patients, in general, are not subject to significantly higher levels of gamma radiation from any of the assessed sources, although they have significantly *lower* levels of gamma radiation from U-238 and Cs-137, compared to the Swedish population as a whole.
- Patients with non-metastatic sPTC are subject to significantly lower levels of gamma radiation from K-40, compared to the Swedish population as a whole.
- Patients with metastatic sPTC are subject to significantly higher levels of gamma radiation from Th-232 and U-238, but significantly lower levels of gamma radiation from Cs-137, compared to the Swedish population as a whole.
- Patients with metastatic sPTC are subject to significantly higher levels of gamma radiation from Th-232, U-238, and K-40, but significantly lower levels of gamma radiation from Cs-137, compared to patients with non-metastatic sPTC.

The statistical modelling of sPTC cases relative to different radiation levels among a set number of "nearest neighbours" to each sPTC patient is generally inconclusive (Fig. 5). There appears to be a greater risk of
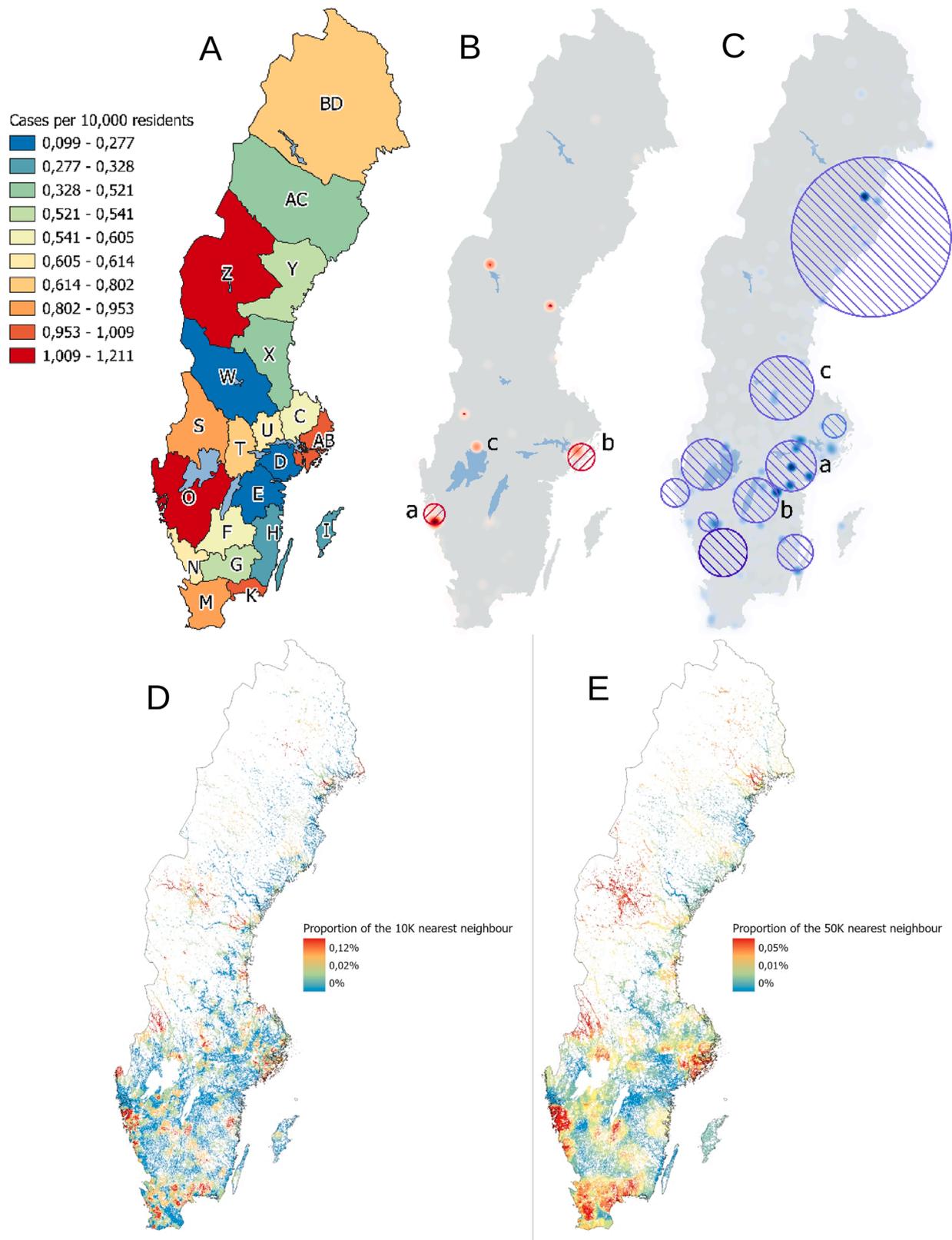
Fig. 3. (A) The prevalence by county (number cases per 10 000), letter = county, see Fig. 2; (B) Clusters of higher-than-expected prevalence from the spatial scan statistics analysis (population at risk = 5%) is indicated as striped circles with red outline. Corresponding results from the heatmap analysis is indicated in red. (C) Clusters of lower-than-expected prevalence from the spatial scan statistics (population at risk = 5%) analysis is indicated as striped circles with blue outline. Corresponding results from the heatmap analysis is indicated in blue. On the second row, the prevalence among the (D) 10,000 and (E) 50,000 nearest population to each 100 m populated cell in Sweden is indicated.
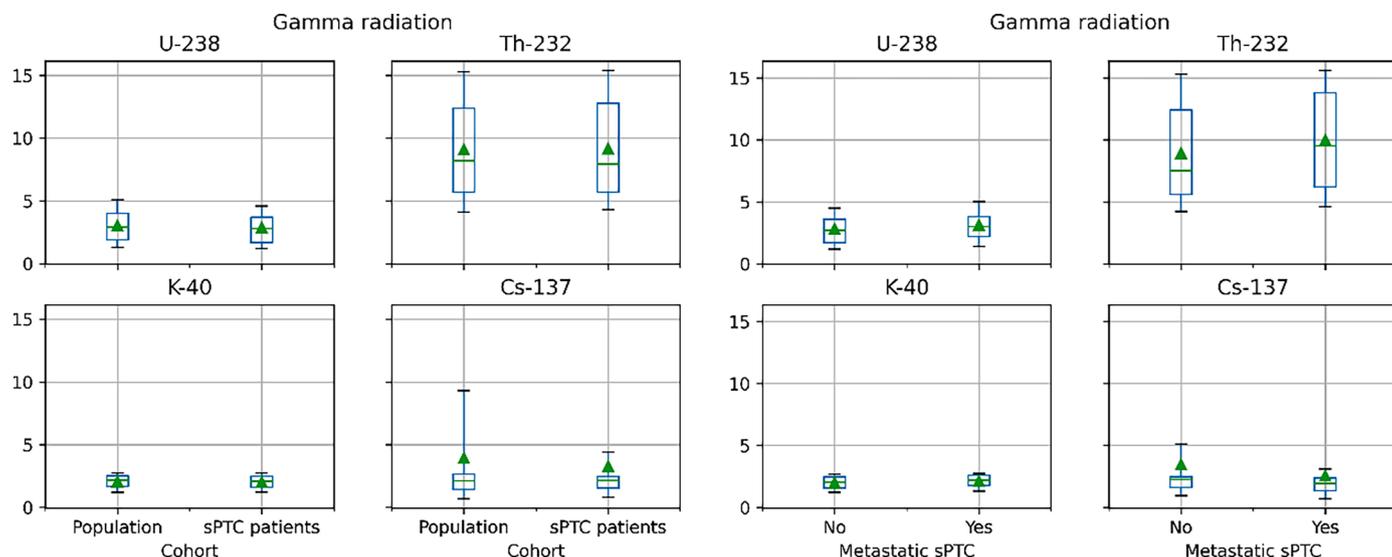
Fig. 4. Whisker diagrams illustrating the statistical degree of gamma radiation from the assessed sources, for the patient sample and the entire Swedish population, respectively (left), and patients with metastatic and non-metastatic sPTC, respectively (right). Box: median, 1st to 3rd quartiles; Whiskers: 10th and 90th percentile; Triangle: mean; Units: U-238 (ppm eU), Th-232 (ppm eTh), K-40 (%), Cs-137 (kBq/m2). Metastatic sPTC refers to small thyroid papillary cancers with central and/or lateral lymph node metastasis, indicating a more advanced disease.

Table 2

Key statistical properties for gamma radiation levels for the assessed sources, for all cohort patients, metastatic cases (N1), non-metastatic cases (N0), and the entire Swedish population. The bold text represents significant results.

		All cases	N1	N0	Entire population
Caesium −137	Mean (kBq/m2)	3.28	2.57	3.49	3.95
	Difference between patients and population	**Lower****	**Lower****	Lower	
	Difference between N1 and N0		**Lower***		
Thorium −232	Mean (ppm eTh)	9.16	9.99	8.92	9.09
	Difference between patients and population	Higher	**Higher****	Lower	
	Difference between N1 and N0		**Higher****		
Uranium −238	Mean (ppm eU)	2.89	3.13	2.83	3.06
	Difference between patients and population	**Lower****	**Higher*****	Lower	
	Difference between N1 and N0		**Higher***		
Potassium −40	Mean (%)	2.04	2.14	2.01	2.06
	Difference between patients and population	Lower	Higher	**Lower***	
	Difference between N1 and N0		**Higher****		

* Significant differences $p < 0.05$,.

** significant differences $p < 0.01$,.

*** significant differences $p < 0.001$.

sPTC at certain levels of radiation from all the different sources, but there are no logical explanations for such "bumps". The results also indicate that (Fig. 5a) very high concentrations of K-40 are associated with a greater risk of sPTC, (Fig. 5b) there is initially a linear increase in sPTC cases with increasing radiation levels from K-40 (which then levels out) and Th-232 (which is then reversed). These findings should, however, be interpreted with caution.

4. Discussion

The different approaches to identifying spatial distribution of sPTC prevalence show some common patterns, although additional studies with larger sample sizes are necessary to provide conclusive results. However, it is obvious that sPTC has causative factors that are neither evenly distributed among the population, nor geographically.

The results show that metastatic sPTCs are associated with significantly higher levels of gamma radiation from Th-232, U-238 and K-40, indicating that increased levels of gamma radiation from the environment could be a risk factor for developing metastatic sPTC. There is, however, no evidence to suggest that it could be a risk factor for developing non-metastatic sPTC. It was unexpected that sPTC patients, according to our analysis, are subject to lower levels of gamma radiation from Cs-137, but this may be explained by methodological limitations. While gamma radiation from U-238, Th-232, and K-40 can be considered relatively constant over time, the Cs-137 data we used in the study refer to levels at the time of disposal following the Chernobyl accident, which (given the relatively short half-life of Cs-137) may not be representative of current conditions. Future studies of connections between sPTC and Cs-137 should take such temporal aspects into consideration.

The relationship with population density was analysed due to the results shown by a previous study that regional differences in the sPTC incidence in Sweden were observed (Bayadsi et al., 2020). The proportion of the patients with the metastatic sPTC was highest in Stockholm region. The study however could not distinguish whether these incidence differences were due to environmental causes or whether, at least in part, they are influenced by differences in diagnostic work-up and the frequency of thyroid surgery performed in different regions of Sweden. This study warranted further investigation to find out the reason for these regional differences. Thyroid cancer management in Sweden is very standardized according to the national guidelines and
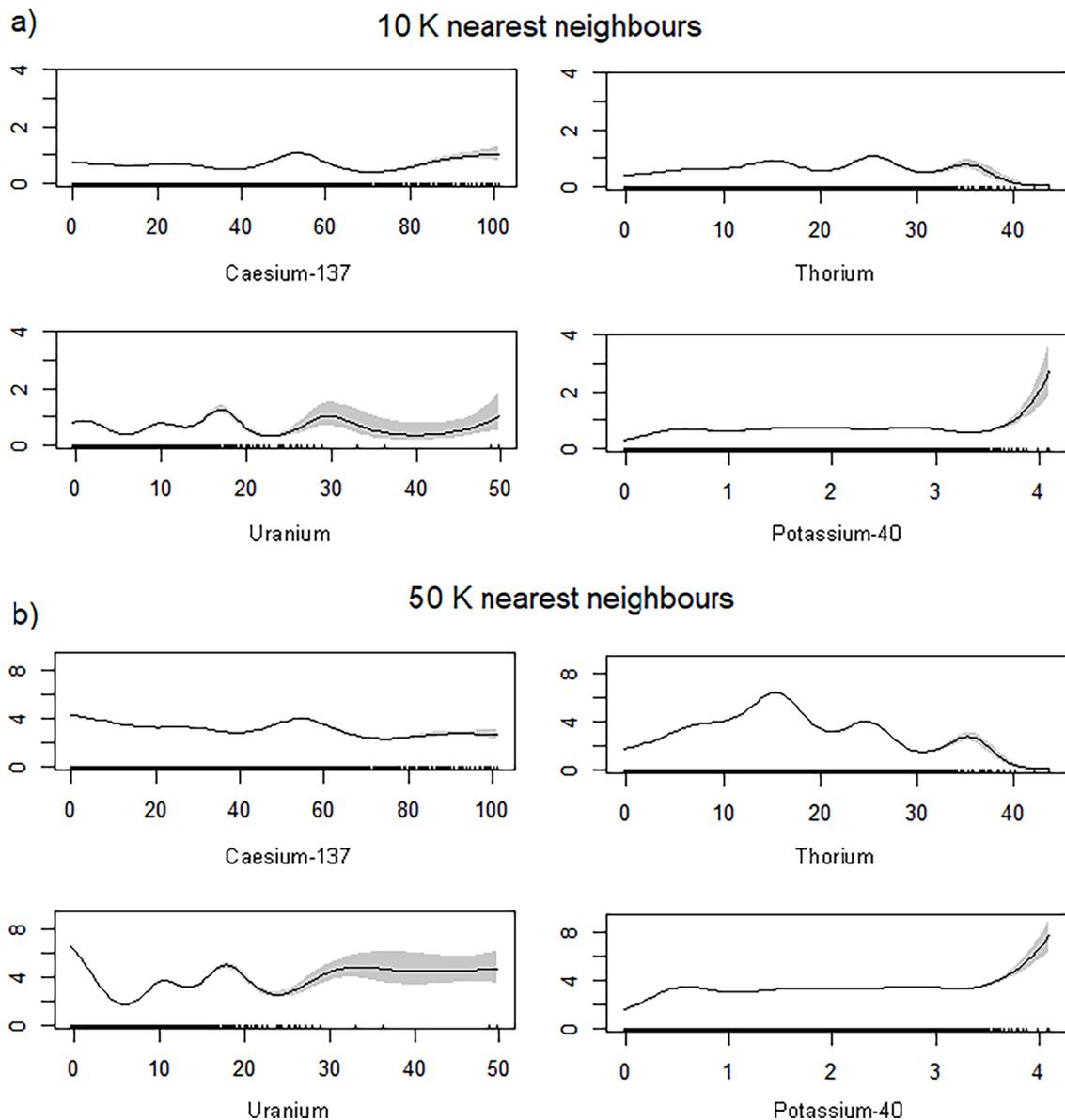
Fig. 5. The partial response plots of the four radiation variables in the GAM models. a) shows the partial responses from the model using each cell's 10,000 nearest neighbours and b) those with the 50,000 nearest neighbours. The x-axis shows the value of the radiation layer in each cell and the y-axis shows the modelled number of patients within the chosen number of nearest neighbours.

recommendations regarding diagnosis, and treatment. This makes it highly unlikely that the differences in sPTC prevalence in Sweden depends on differences in detection or diagnosis rates only.

Several additional methodological limitations may have affected the results. First, the analysis lacks information regarding where the patients were born and where they lived before the time of diagnosis. The average Swedish resident moves about 10 times during a lifetime (Statistics Sweden SCB, 2022), and it can therefore be assumed that some patients have moved several times from birth to time of diagnosis. The gamma radiation levels in the location where they lived at the time of diagnosis may thus not be representative of the gamma radiation they were subjected to during their childhood. Since there is a higher risk of developing PTC when exposed to ionizing radiation in childhood than in adulthood (Morton et al., 2021; Wakeford, 2004), this is an important limitation of this study. Future studies should attempt to track the movement of the patients and aggregate radiation levels from different residence locations.

The limited sample size probably explains why the results are inconclusive and sometimes contradictory. Considering the geographical prevalence, this is particularly problematic in a country like Sweden, with large variations in population density across the country. One single case in a sparsely populated area could mean an over-representation compared to what can be expected based on the population in the area used for aggregation. It would therefore be misleading to report differences in prevalence for areas that are too small, in terms of population. A larger sample size would facilitate the use of smaller aggregation units but would also enable a better basis for geostatistical analysis. Considering radiation, or other potential risk factors that can be similarly assessed, a larger sample size would make the spatiotemporal concerns less problematic. However, provided that the sample is truly random and that the spatiotemporal concerns can be properly addressed, the limited size of the sample should not significantly impact the results.

There are multiple challenges in determining suitable methods for assessing spatial differences in prevalence. All three spatial analysis methods presented here are sensitive to sample size, and the location of very few cases in sparsely inhabited areas can, questionably, cause large areas to be identified as "hot spots". There are also other methodological

concerns that should be noted. Heatmaps, for example, have been used in a few similar studies (Kiani et al., 2021), but they are principally a kind of logical model (Fatima et al., 2021), requiring careful attention to detail in design and parameterization, as well as empirical validation, to be useful. Both the former and the latter are difficult. However, if done appropriately, they can be effective in identifying areas with an over- or under-representation, especially since they are not limited to a pre-defined aggregation unit, which may or may not be appropriate given underlying (unknown) causative factors. That is, models that are not dependent on predefined borders for aggregation may be more effective in identifying geographical patterns. Regardless of the method used to identify spatial differences in prevalence (for any disease), it is imperative to use appropriate methods and to validate (or at least evaluate) the results using empirical data and statistical analysis. The latter is the main strength of Spatial Scan Statistics, which should be considered a more robust method for identifying spatial clusters. In many cases, it may be wise to use multiple methods that provide complementary information, such as demonstrated here.

Spatially explicit register-based studies, such as shown here, are essential to studying any disease with unknown causative factors associated with the physical environment, or otherwise spatially dependent, and may thus have a notable potential for future research. One example where this kind of study could provide important scientific advances is Type 1 Diabetes Mellitus (DM), for which Sweden has the second highest incidence in the world. Since this cannot be explained by genetics, the possibility of environmental risk factors has been raised multiple times (Rewers and Ludvigsson, 2016). Spatially explicit register-based studies can be used to eliminate or confirm suspected environmental risk factors but also, given the abundance of geographical data, can identify new causative factors associated with the physical environment, or new spatial patterns that can provide additional information for better understanding the cause of the disease. Another strength of this kind of study is that it allows us to use the entire population as a comparative cohort, without requiring any information about individuals.

Given the above discussion, further analysis of sPTC requires a larger sample size and more information about all addresses of all patients from their time of birth to their time of diagnosis. Future studies should also seek to statistically determine potential gender differences and explore additional methods for geostatistically analysing the prevalence across the country.

5. Conclusions

Metastatic sPTCs are associated with significantly higher levels of gamma radiation from Th-232, U-238 and K-40, indicating that increased levels of gamma radiation from the environment could be a risk factor for developing metastatic sPTC. There is, however, no evidence for suggesting that it could be a risk factor for developing non-metastatic sPTC. sPTC is significantly more common in densely populated areas and spatial clusters with higher-than-expected prevalence could be identified around the two largest cities in Sweden. The results in this study clearly indicate that sPTC has causative factors that are neither evenly distributed among the population, nor geographically. Spatially dependent factors could be either demographic or environmental. Further analysis of sPTC requires a larger sample size and more information about all addresses of all patients from their time of birth to their time of diagnosis.

Spatially explicit register-based studies, such as shown here, are essential to studying any disease with unknown causative factors associated with the physical environment, or otherwise dependent on location, and may thus have significant potential for future research.

Author contributions

Conceptualization and methodology: H. Bayadsi, P. Van Den Brink, M. Erlandsson, P. Nordin, J. Hennings and O. Englund. Funding acquisition: H. Bayadsi. Data curation: H. Bayadsi and S. Koorem. Formal data analysis: H. Bayadsi, P. Van Den Brink, M. Erlandsson and O. Englund. Project administration & supervision: O. Englund. Writing - orginal draft: H. Bayadsi, S. Gudbjornsson, S. Sebraoui and O. Englund. Writing - review & editing: All authors commented on previous versions of the manuscript and read, editied and approved the final manuscript.

Ethics

This study was approved by the Swedish Ethical Review Authority, permit number 2019-04983, and was carried out in accordance with the EU's General Data Protection Regulation (GDPR) rules.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Haytham Bayadsi reports financial support was provided by Unit of Research Education and Development, Jämtland-Härjedalen Region, Sweden. Haytham Bayadsi reports financial support was provided by Jämtland County Cancer and Care Fund.

Data availability

Data will be made available on request.

References

Agency for Toxic Substances and Disease Registry (ATSDR). (2019). *ToxFAQs™ for Thorium.* Centres for Disease Control and Prevention (CDC). Cited December 15, 2022, Available from: https://wwwn.cdc.gov/TSP/ToxFAQs/ToxFAQsDetails.aspx?faqid=659&toxid=121.

Geophysical flight measurements, Gamma radiation (overview). 2016 [cited 2022; Available from: https://resource.sgu.se/dokument/produkter/geofysiska-flygmatningar-gammastralning-oversiktlig-beskrivning.pdf.

The American Joint Committee on Cancer (AJCC) (2010). *AJCC 7th Ed Cancer Staging Manual.* Cancer Staging System Products. Cited December 1, 2022, Available from: https://www.facs.org/quality-programs/cancer-programs/american-joint-committee-on-cancer/cancer-staging-systems/ajcc-cancer-staging-system-products/.

Albi, E., et al., 2017. Radiation and thyroid cancer. Int. J. Mol. Sci. 18 (5).

Alinaghizadeh, H., et al., 2016. Total cancer incidence in relation to 137Cs fallout in the most contaminated counties in Sweden after the Chernobyl nuclear power plant accident: a register-based study. BMJ Open 6 (12), e011924.

Aneja, S., et al., 2011. Geographical information systems: applications and limitations in oncology research. Oncology (Williston Park) 25 (12), 1221–1225.

Bayadsi, H., et al., 2020. Invasiveness and metastatic aggressiveness in small differentiated thyroid cancers: demography of small papillary thyroid carcinomas in the Swedish population. World J. Surg. 44 (2), 461–468.

Connor, N. (2019). *What is Potassium-40 – Characteristics – Half-life – Definition.* Radiation Dosimetry. Cited December 1, 2022, Available from: https://www.radiation-dosimetry.org/what-is-potassium-40-characteristics-half-life-definition/.

Dahlberg, J. (2022). *Thyroid cancer, annual report.* Swedish National Quality Registry, Regional Cancer Centres (RCC). Cited December 1, 2022, from https://cancercentrum.se/samverkan/cancerdiagnoser/skoldkortel/kvalitetsregisterNational.

Dempsey, C. (2002). *What is GIS?* GIS Lounge. Cited December 1, 2022, Available from: https://www.gislounge.com/what-is-gis/.

Evers, J. (2022). *GIS (Geographic Information System).* National Geographic Society. Cited December 1, 2022, Available from: https://education.nationalgeographic.org/resource/geographic-information-system-gis/.

Fatima, M., et al., 2021. Geospatial analysis of COVID-19: a scoping review. Int. J. Environ. Res. Public Health 18 (5).

Furukawa, K., et al., 2013. Long-term trend of thyroid cancer risk among Japanese atomic-bomb survivors: 60 years after exposure. Int. J. Cancer 132 (5), 1222–1226.

Hayashi, Y., et al., 2010. Papillary microcarcinoma of the thyroid among atomic bomb survivors: tumor characteristics and radiation risk. Cancer 116 (7), 1646–1655.

Imaizumi, M., et al., 2015. Association of radiation dose with prevalence of thyroid nodules among atomic bomb survivors exposed in childhood (2007-2011). JAMA Intern. Med. 175 (2), 228–236.

Kazakov, V.S., Demidchik, E.P., Astakhova, L.N., 1992. Thyroid cancer after Chernobyl. Nature 359 (6390), 21.

Kiani, B., et al., 2021. Association between heavy metals and colon cancer: an ecological study based on geographical information systems in North-Eastern Iran. BMC Cancer 21 (1), 414.

Kulldorff, M., 1997. A spatial scan statistic. Commun. Statist. - Theory Method. 26 (6), 1481–1496.

Kulldorff, M., 1999. Geographic information systems (GIS) and community health: some statistical issues. J. Public Health Manag. Pract. 5 (2), 100–106.

Lim, H., et al., 2017. Trends in thyroid cancer incidence and mortality in the United States, 1974-2013. JAMA 317 (13), 1338–1348.

Ludvigsson, J.F., et al., 2009. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur. J. Epidemiol. 24 (11), 659–667.

Moberg, L., 2001. The Chernobyl accident. A summary fifteen years after the accident. In: , p. 21.

Molinaro, E., Campopiano, M.C., Elisei, R., 2021. Management of endocrine disease: papillary thyroid microcarcinoma: toward an active surveillance strategy. Eur. J. Endocrinol. 185 (1), R23–R34.

Morton, L.M., et al., 2021. Radiation-related genomic profile of papillary thyroid carcinoma after the Chernobyl accident. Science 372 (6543).

National Cancer Institute (NCI) (2022). *Thorium.* Cited December 15, 2022, Available from: https://www.cancer.gov/about-cancer/causes-prevention/risk/substances/thorium.

National Center for Environmental Health (NCEH) (2018). *Radioisotope Brief: Cesium-137 (Cs-137).* Centres for Disease Control and Prevention (CDC). Cited December 1, 2022, Available from: https://www.cdc.gov/nceh/radiation/emergencies/isotopes/cesium.htm.

National Health Institute (NHI) (2022, March 30). *Cancer Stat Facts: Thyroid Cancer.* National Cancer Institute (NCI), The Surveillance, Epidemiology, and End Results (SEER) Database. Cited January 30, 2023, Available from: https://seer.cancer.gov/statfacts/html/thyro.html.

Nikiforov, Y.E., 2010. Is ionizing radiation responsible for the increasing incidence of thyroid cancer? Cancer 116 (7), 1626–1628.

Nordenström, E. (2021). *SQRTPA Annual Report 2021.* Scandinavian Quality Register for Thyroid, Parathyroid and Adrenal Surgery (SQRTPA). Cited December 1, 2022, Available from: https://sqrtpa.se/arsrapporter.

Open Source Geospatial Foundation (OSGeo) (2022). *QGIS - A Free and Open Source Geographic Information System.* QGIS. Cited December 1, 2022, Available from: https://www.qgis.org/en/site/.

Östh, J., 2014. Introducing the EquiPop software: An application For the Calculation of K-Nearest Neighbour Contexts/Neighbourhoods. Uppsala University, Department of Social and Economic Geography, Uppsala, Sweden.

Powers, A.E., et al., 2019. Changes in trends in thyroid cancer incidence in the United States, 1992 to 2016. JAMA 322 (24), 2440–2441.

Rao, D., *Editorial*, 2012. Radiat. Protect. Environ. 35 (2), 57–58.

Rewers, M., Ludvigsson, J., 2016. Environmental risk factors for type 1 diabetes. Lancet 387 (10035), 2340–2348.

Ruedas, T., 2017. Radioactive heat production of six geologically important nuclides. Geochem. Geophys. Geosyst. 18 (9), 3530–3541.

Geological Survey of Sweden (SGU) (2019). *Gamma Radiation, Potassium, map viewer.* Cited December 1, 2022, Available from: https://www.sgu.se/produkter-och-tjanster/kartor/kartvisaren/geofysikkartvisare/gammastralning-kalium/.

Siegel, R.L., Miller, K.D., Jemal, A., 2019. Cancer statistics, 2019. CA Cancer J. Clin. 69 (1), 7–34.

SSM, 1986. Chernobyl - Its Impact on Sweden. Swedish Radiation Safety Authority (SSM), Stockholm, Sweden, p. 24.

Statistics Sweden (SCB) (2022). *Open data for grid statistics.* Cited December 1, 2022, from https://www.scb.se/en/services/open-data-api/open-geodata/grid-statistics/.

Statistics Sweden (SCB) (2022). *Moving within Sweden.* Cited December 1, 2022, Available from: https://www.scb.se/hitta-statistik/sverige-i-siffror/manniskorna-i-sverige/flyttar-inom-sverige/.

Swedish Radiation Safety Authority (SSM) (2017). *Gamma radiation at ground level.* Cited December 1, 2022, Available from: https://www.stralsakerhetsmyndigheten.se/omraden/miljoovervakning/radioaktiva-amnen/gammastralning-vid-marknivan/.

Takahashi, T., et al., 2003. The relationship of thyroid cancer with radiation exposure from nuclear weapon testing in the Marshall Islands. J. Epidemiol. 13 (2), 99–107.

Taylor, D.M., Taylor, S.K., 1997. Environmental uranium and human health. Rev. Environ. Health 12 (3), 147–157.

Thyroid cancer, Swedish national guidelines. 2021 2021-09-07 [cited 2022; Available from: https://kunskapsbanken.cancercentrum.se/diagnoser/skoldkortelcancer/vardprogram/.

Uddin, S., et al., 2022. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci. Rep. 12 (1), 6256.

Uddin, S., et al., 2019. Comparing different supervised machine learning algorithms for disease prediction. BMC Med. Inf. Decis. Making 19 (1), 281.

United-Nations, 2013. Report of the United Nations Scientific Committee on the effects of atomic radiation. General Assembly, Official Records, pp. 13–22.

Vaccarella, S., et al., 2016. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. N. Engl. J. Med. 375 (7), 614–617.

van Gerwen, M., et al., 2020. Association between uranium exposure and thyroid health: a National Health and Nutrition examination survey analysis and ecological study. Int. J. Environ. Res. Public Health 17 (3).

Wakeford, R., 2004. The cancer epidemiology of radiation. Oncogene 23 (38), 6404–6428.

Wood, S., 2017. Generalized additive models. In: An Introduction With R, 66. Chapman and Hall/CRC, New York, p. 496.