D7.4 Report on pilot environmental study results

Alessandro Borgini, Martina Bertoldi, Andrea Tittarelli,
Roberto Lillini, Paolo Baili, Paolo Contiero

V1 – 30th June 2021

The Pilot of Environmental Study in the Province of Alto Adige

PROJECT BACKGROUND

The project "WAter & Soil contamination and Awareness on Breast cancer risk in Young women Project" (WASABY) focuses on the geographical analysis of population-based cancer incidence data in connection with environmental factors. Project's details are at www.wasabysite.it.

Specifically, WASABY aims to evaluate correlation between water & soil pollutants (e.g., arsenic in water, topsoil metals, etc.) and a given health outcome (i.e., breast cancer incidence) with data already available, pre-existing WASABY, collected, checked and validated by official sources.

One of the project objectives was to perform a pilot environmental study on breast cancer and water/soil contaminants to be carried out in one selected European area.

The study area chosen, identified during the course of the project, is the Province of Alto Adige because:

- Information on type of persistent chemical substances and on the general level of chemical pollution in this area is accessible through the use of open source environmental databases and digital archives.
- The target area has a history of contamination and available data that takes into account the delay of time between exposure to contaminants and the diagnosis of breast cancer.
- In this specific area we also have the presence of a Cancer Registry with high-quality geo-coded data, covering the entire province. The Cancer registry participated in the Wasaby Work Package 4.
- With 533349 inhabitants (ISTAT data) and an area of 7398.38 km², the autonomous province of Bolzano - Alto Adige is the second largest province in Italy after the province of Sassari.
- From an economic and productive point of view, in Alto Adige there is a variety in the sectors of productive activity ranging from agriculture, to the food sector to the transport sector.

The present report shows data used, methods applied and results obtained in the environmental pilot study performed for WASABY project.

1.0 – THE TERRITORIAL REFERENCE FRAMEWORK OF THE PROVINCE OF ALTO ADIGE

The autonomous Province of Bolzano - Alto Adige has an area of 7398.38 km²; it borders to the north and east with Austria (Tyrol and Salzburg), to the west with Switzerland (Canton of Grisons), to the south-east with Veneto (province of Belluno), to the south with the autonomous province of Trento (Trentino) and to the south-west (at the Stelvio pass) with Lombardy (province of Sondrio) (Figure 1).

The province includes countless valleys, passes, rivers and lakes that surround the entire South Tyrolean territory. Valley Aurina is the northernmost Valley in all of Italy and Predoi the northernmost municipality, located in the innermost part of the Valley, under the mountain of Italy, on the border with Austria.

Figure 1 – Alto Adige Province.

2.0 – RESIDENT POPULATION AND PRINCIPAL PRODUCTIVE ACTIVITIES

The resident population in the province of Alto Adige is 533,349 inhabitants (ISTAT - Census 2020). From an economic and productive point of view, in Alto Adige the fruit growing sector is very developed: 10% of the apples of the European Union are grown in South Tyrol, on an area of 18000 hectares thanks to the mild climate. Other important production sectors are precisely that of agriculture, food, road haulage and cableways, the wood industry, mountain clothing. In this province a lot of efforts have been made for growth with environmental sustainability. Alto Adige has focused heavily on alternative energy sources: water energy, solar energy and biomass. The Province of Alto Adige is therefore considered to be at the forefront in Italy for having focused heavily on the green economy of its companies.

But by far the most important economic sector is tourism for the renowned ski resorts. The beauty of the Dolomite landscape and the excellent South Tyrolean cuisine attract millions of tourists every year. The province of Bolzano is the first in Italy in terms of overnight stays, ahead of Rimini and Venice.

3.0 – HYDROGEOLOGICAL CLASSIFICATION OF THE TERRITORY OF THE PROVINCE OF ALTO ADIGE

The assessment of the spatial distribution of groundwater bodies must be related to the hydro-geological structures present in the territory; in addition, the identification of underground water bodies also include the direction of groundwater flow and the groundwater feed and delivery areas water.From the hydro-geological point of view, Alto Adige is variously diversified and, in the first analysis, it can be divided into three sectors: the mountains area, the valleys and the alpine passes area.

3.1 – The mountains area

The Province of Alto Adige is an area rich in mountains. On its territory the Central Alps rise, to which the Ortles belongs, the highest mountain in the autonomous province of Bolzano (as well as the entire Trentino-Alto Adige region) with its 3905 mts., and the Eastern Alps. Among the main reliefs there are also the Palla Bianca, the Similaun, the Cima Altissima, the Gran Pilastro.

A part of the Dolomites also belongs to Alto Adige, declared a World Heritage Site in 2009. Among the Dolomite mountains are the Sesto Dolomites with Punta Tre Scarperi, Croda dei Baranci, and the Tre Cime di Lavaredo, the Dolomites of Gardena (Sassolungo, Sassopiatto, Sella, Piz Boè) and the other Dolomite reliefs such as Plan de Corones, Latemar, Plose, Catinaccio, Sciliar.

3.2 – The valleys area

Adige Valley identifies the stretch of the Valley crossed by the Adige river that goes from Merano to Rovereto, in Trentino. Adige Valley is the most densely populated area of the province, rising from the metropolitan area of Bolzano. Below there is a list of the main Valleys and their tributaries.

- ***Adige Valley:*** Passiria Valley, d'Ultimo Valley.
- ***Isarco Valley:*** Tires Valley, Gardena Valley, Ridanna Valley, Sarentino Valley, d'Ega Valley, Vizze Valley, Racines Valley, Giovo Valley, Fleres Valley.
- ***Pusteria Valley:*** Tures Valley, Aurina Valley, Badia Valley, Casies Valley, Fiscalina Valley, Anterselva Valley, Landro Valley, Braies Valley.
- ***Venosta Valley:*** Monastero Valley, Martello Valley, Senales Valley.

3.3 – The alpine passes

The Brenner Pass is the main border crossing between Italy and Austria. It lies on the watershed between the basins of the Adriatic Sea and the Black Sea. On the Italian side the Isarco Valley descends, on the Austrian side the Sill river.

Among the many other Alpine passes are: Pass of Gardena, Pass of Nigra, Pass of Sella, Pass of Stalle, Pass of the Rombo, Pass of the Mendola, Pass of the Erbe, Pass of the Palade, Pass of the Stelvio, Pass of Campolongo, Pass of Costalunga, Pass of Monte Croce of Comelico, Pass of Monte Giovo, Passo di Pampeago, Pass of Pennes, Pass of Resia, Pass of Valley Parola, Pass San Lugano.

3.4 – Rivers

The Alto Adige area is essentially characterized by the high basin of the Adige river, the second Italian river after the Po, located north of the Salorno strait. The most important tributaries are the Isarco, the Talvera, the Rienza, the Passirio and the Rio Ram, whose sources are located in Switzerland. Also located in South Tyrol are the sources of the Drava, which later develops in Central-Eastern Europe and flows into the Danube.

3.5 – Lakes

In the province of Bolzano there are 176 natural water basins with a length greater than or equal to 100 mts. Most of these basins are located at altitudes above 2000 mts. There are 13 natural lakes with an area greater than five hectares: of these only three, Lake Caldaro (the largest natural lake in South Tyrol) and the two lakes of Monticolo, are located below 1000 mts.

Other main natural lakes include Lake Anterselva, Lake Braies, Lake Carezza, Lake Costalovara, Lake Dobbiaco, Lake Favogna, Lake Fiè, Lake Santa Maria, Lake of San Valentino alla Muta, Lake Landro and Lake Varna.

There are also artificial lakes, some of which are of considerable size. Among the main ones we remember the Resia lake (by far the largest lake in the province), the Zoccolo lake, the Fortezza lake, the Rio di Pusteria lake and the Valdaora lake.

4.0 – DATA USED FOR WASABY PILOT ENVIRONMENTAL STUDY

The WASABY Pilot Environmental Study is a feasibility study that evaluates the relationship between health outcomes (breast cancer incidence), environmental factors (water pollutants) and the socio-economic indicators through quantitative and qualitative approaches using the health data of the cancer registry of the Province of Alto Adige and the environmental data already available in the international, national and local pollutants databases.

Several data sources were considered to develop the WASABY pilot environmental study.

4.1 – Cancer registry of the province of Alto Adige

The Alto Adige Cancer Registry (CR) routinely collects breast cancer cases in order to estimate cancer epidemiological indicators (i.e., breast cancer incidence). For WASABY, this data have been reported individually (XY coordinates). A geographic trans-coding procedure, performed by a specific Excel tool, allowed to report the coordinates and the cases to the corresponding Census Tract (CT).

Primary invasive female breast cancer (ICD9 174*, ICD10 C50*), selected from cancer registries data during a specific ten years period (specifically, 2004 to 2013) are included in the project. According to the goals of the project, only cases with age at diagnosis less than 50 years of age have been considered. Synchronous and metachronous breast cancer cases have been counted once. Residence addresses at diagnosis have been retrieved from the individual data reference of the CR.

Incident cases were diagnosed in a period of 10 years, from 1st January 2004 to 31st December 2013, for a total number of 747 female breast cancer patients aged 0-49 year. 6 cases (0.8%) were not geo-coded for missing or mistaken geographic information and were excluded from the analysis.

For this pilot environmental study, WASABY have used the two different populations for standardization purposes:

- the Segi's World Standard Population, used for the direct standardization;
- the 2001 population of the Province of Alto Adige at CT level. More specifically, the female population data by 5-year age groups, calendar year within time period and CT have been considered.

4.2 – Shapefiles collected with maps of the cancer registry of the province of Alto Adige

For the selected CR of the Province of Alto Adige, WASABY needed to have a complete shapefile of the geographic area covered by its activity. The shapefile format is a digital vector storage format for storing geometric location and associated attribute information. It consists of a collection of files with a common filename prefix Alto Adige.shp, Alto Adige.dbf, Alto Adige.shx, stored in the same directory, with mandatory and optional files.

Files must be combined with information on calendar years of validity (in case of administrative changes of CT in the studied incidence years).

4.3 – European Deprivation Index for socio-economic deprivation

The socio-economic status (SES) of the patients has been proven in literature to be a relevant covariate able to influence the health status of a population and, in WASABY, it has been considered the covariate to be studied along with the environmental pollutants as an element which could have influenced the breast cancer incidence. We didn't have the individual socio-economic characteristics of the cases, therefore a socio-economic index able to describe the SES characteristics of the CT of residence of every case had to be used. The European Deprivation Index (EDI) was used for this analysis.

4.4 – Environmental database sources collected and used for different persistent pollutants

For the pilot environmental study in the Province of Alto Adige we have used various local, national and international databases where the data on the concentrations of the most persistent and dangerous contaminants in the various environmental matrices are available, especially the matrix of ground water coming from wells located in the territory of Alto Adige.

We started using reliable international and, if possible, local databases available in open source mode from the different website to obtain reliable data, checked and certified.

We selected fifteen contaminants that could represent different classes of the most persistent, toxic and harmful contaminants such as chlorinated pesticides, carcinogenic halogenated aliphatic contaminants and the main heavy metals that can cause a risk to breast cancer, these persistent contaminants, divided into different "families o classes of contaminants we had already selected them in an exhaustive bibliographic collection initially requested by the WASABY project (See Wasaby Deliverable D7.1).

The contaminants that we have selected are the following: trichloromethane and trichloromethane (aliphatic chlorinated carcinogens), aldrin, endrin, dieldrin, p-p DDT (organoclorinated insecticide), atrazines and simazine (herbicide of the triazine class), chlorpyrifos (organophosphate pesticide) arsenic, lead, nickel, mercury, chrome, cadmium (heavy metal) (Table 1).

We also used a specific database of the European Environment Agency (EEA) to collect groundwater quality measurements in the Province of Alto Adige for several years: the Waterbase database - Water quality.

Waterbase Water Quality is the generic name given to the EEA database on the status and quality of Europe's rivers, lakes, groundwater and transitional, coastal and marine waters, the quantity of European water resources and emissions in surface waters from the point and widespread sources of pollution.

We have recovered the data relating to deep waters (mainly from wells) at the level of the Province of Alto Adige with the geographical coordinates indicated, in order to report on the map both the location of the wells and the values of the concentrations of persistent contaminants selected through the use of the Geographic Information System (GIS).

Therefore these georeferenced maps indicate both the concentration of the selected organic and/or heavy metal contaminant and the specific site for the detection of the contaminant and the sampling period. The entire database is available at the following link:

https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quality-2

The data series are calculated as the average of the annual average concentrations in the Province of Alto Adige for groundwater bodies (wells).

Table 1 – Characteristics of the pollutants selected for the Environmental Pilot Study.

Category	**Substance**	**CAS NUMBER Substance**	**Unit of measurement (U.M.)**	**Legal limit values**
Heavy Metal	Arsenic	7440-38-2	µg/L	10
Heavy Metal	Cadmium	7440-43-9	µg/L	5
Heavy Metal	Chrome total	7440-47-3	µg/L	50
Heavy Metal	Chrome VI	18540-29-9	µg/L	5
Heavy Metal	Mercury	7439-97-6	µg/L	1
Heavy Metal	Nichel	7440-02-0	µg/L	20
Heavy Metal	Lead	7439-92-1	µg/L	10
Aliphatic Chlorinated Carcinogens	Trichloroethylene	79-01-6	µg/L	10
Aliphatic Chlorinated Carcinogens	Trichloromethane	67-66-3	µg/l	0,15
Organophosphate Pesticide	Clorpirifos	2921-88-2	µg/l	0,1
Organoclorinated Insecticide	Aldrin	309-00-2	µg/L	0,03
Organoclorinated Insecticide	Dieldrin	60-57-1	µg/L	0,03
Organoclorinated Insecticide	Endrin	72-20-8	µg/L	0
Organoclorinated Insecticide	o,p'-DDT	789-02-6	µg/l	0,1
Nitrogen Pesticides	Atrazine	1912-24-9	µg/L	0,1
Nitrogen Pesticides	Simazine	122-34-9	µg/L	0,1

5.0 – METHODS USED FOR WASABY PILOT ENVIRONMENTAL STUDY

5.1 – Estimation of the pollutants distribution using GIS and interpolation method (Kriging)

In this Environmental Pilot Study, ArcGIS 10.4 and QGIS 3.14.16. software were used for the preparation of zoning maps. A specific method was applied to estimate the distribution of the pollutants on the considered area of the Province of Alto Adige, starting from fixed sources of water and pollution.

After an accurate international literature review on the methods used for such studies from the Seventies to today (Lillini et al., 2021), the Kriging interpolation method was chosen.

More specifically, the Empirical Bayesian Kriging was chosen, due to its advantages and to the peculiarities of the available environmental data, and it was used to analyze the data in GIS Software

In a general sense, the Kriging weights are calculated such that points nearby to the location of interest are given more weight than those farther away. Clustering of points is also taken into account, so that clusters of points are weighted less heavily (in effect, they contain less information than single points). This helps to reduce bias in the predictions. The Kriging predictor is an "optimal linear predictor" and an exact interpolator, meaning that each interpolated value is calculated to minimize the prediction error for that point. The value that is generated from the Kriging process for any actually sampled location will be equal to the observed value at this point, and all the interpolated values will be the Best Linear Unbiased Predictors (BLUPs). Kriging will in general not be more effective than simpler methods of interpolation if there is little spatial autocorrelation among the sampled data points (that is, if the values do not co-vary in space). If there is at least moderate spatial autocorrelation, however, Kriging can be a helpful method to preserve spatial variability that would be lost using a simpler method.

Due to the peculiarities of the available data in WASABY, the chosen method among the various Kriging techniques is the Empirical Bayesian Kriging (EBK) Interpolation. EBK is a geostatistical interpolation method that automates the most difficult aspects of building a valid Kriging model. EBK also differs from other Kriging methods by accounting for the error introduced by estimating the underlying semivariogram. Other Kriging methods calculate the semivariogram from known data locations and use this single semivariogram to make predictions at unknown locations; this process implicitly assumes that the estimated semivariogram is the true semivariogram for the interpolation region. By not taking the uncertainty of semivariogram estimation into account, other Kriging methods underestimate the standard errors of prediction.

5.2 – Computing the expected cases and the SIRs by CT

According to the methodologies presented in WASABY WP4, WP5 and WP6, the geo-coding of Alto Adige Cancer Registry (CR) cases was performed at Census Tract (CT) level, integrating some correction about the right attribution of the 2001 CT. The official 2001 shapefile at CT level for the registry-covered area and the 2001 Census Population were both provided by the Italian National Statistics Office (ISTAT); the 2001 European Deprivation Index (EDI) at CT level were provided by WP4. A database with CT as key variable was built, merging the shapefiles (.shp and .dbf dataset from shapefile conversion in STATA format), the 2001 EDI dataset with the geo-coded Alto Adige CR dataset.

The 2001 EDI was considered in its quantitative form and distributed at CT level. The values came from a negative minimum (the lowest deprivation) to a positive maximum (the highest deprivation).

The 741 considered incident cases were distributed in 605 CTs. In 5008 CTs where no cases were present, observed incidence was put to 0 instead of missing value. This choice was due to the need of having no empty cells that could bias the spatial analysis and for mapping the area completely.

The expected cases and the SIRs were computed both by direct and indirect standardization.

When performing the direct standardization, the World population standard was used. The World Age-Standardised Rate (ASR-W) per 100,000 for 0-49 years old female breast cancer patients is the mean obtained by the two values reported by the Cancer Incidence in Five Continents X and XI:

ASR-W (2003-2007) = 30.4
ASR-W (2008-2012) = 26.2

Mean ASR-W (2003-2012) = 28.3

When performing the indirect standardization, age-standardized rates were computed by using the 2001 local population by CT per 10 years (covering the incidence time range).

5.3 – SARAR models

According to the methods proposed by the WP6 in deliverable D6.1 and by the WP4 in deliverable D4.3, the following spatial analysis techniques were performed on this dataset:

- Inferential estimation of SIRs by Spatial-Autoregressive model with SAR disturbances (SARAR) models, using STATA 14 (module: spreg) (applied to both direct and indirect standardization).
- Detection of spatial disease clusters and evaluation of their statistic significance by continuous Poisson model, using SaTScan 9.6 (applied to both, if results from previous step were statistically significant different, as tested by comparison of the 95% Confidence Intervals).

SIR estimation with SARAR models was performed, using a generalized spatial two-stage least-squares (GS2SLS) estimator, considering also the effects of the socio-economic deprivation (by EDI 2001 used in its quantitative format and as an exogenous variable). The procedure was applied to both standardized SIRs and allowed to compute estimated SIRs and evaluate the probability of statistical significant geographic clusters of disease.

5.4 – SaTScan: Detection of spatial disease clusters

The cluster analysis was developed by applying a Discrete Poisson Model. With the discrete Poisson model, the number of cases in each location is Poisson-distributed. Under the null hypothesis, and when there are no covariates, the expected number of cases in each area is proportional to its population size, or to the person-years in that area. Poisson data were analyzed with the purely spatial scan statistics.

The discrete Poisson model required case and population counts for a set of data locations such as census tracts, as well as the geographical coordinates for each of those locations. These were provided using the case, population and coordinates files. The population data needed not be specified continuously over time, but only at one or more specific ‘census times’.

5.5 – Association of the estimated pollutants distribution with the presence of geographic clusters of breast cancer incidence

Finally, the possible connection between pollutant variables and the so-found clusters was evaluated by logistic regression model (Relative Risk Ratios, $p<0.05$).

5.6 – Software used

Stata 14.0 was used for SIR computation, SARAR models application and logistic regression. SaTScan 9.6 allowed developing the cluster analysis.

6.0 – RESULTS

6.1 – Estimation of the pollutants using the Kriging method

The application of the Kriging technique to Alto Adige CR area allowed considering the molecules and other elements and characteristics in the water of the province. The information came from the various water wells for water inspection existing in the area, as collected by the EEA Water Quality Database. All the water wells were geo-coded in terms of X&Y coordinates and attributed to the corresponding Census Tract.
According to the examined literature, also the water wells in a 10 kms-buffer were considered for interpolation.
Seventy elements were found (e.g., molecules, water ph, etc.) and 15 of them were considered for their potential connection with breast cancer incidence (as from literature):

Aldrin

Arsenic and its compounds

Atrazine

Cadmium and its compounds

Chlorpyrifos

Chromium and its compounds

DDTpp1

Dieldrin

Endrin

Lead and its compounds

Mercury and its compounds

Nickel and its compounds

Simazine

Trichloroethylene

Trichloromethane

The following maps (Figures 2a-2d) report the distribution of the interpolated estimation of the quantity of the 15 elements (in deciles). All pollutants were significantly lower than the limits fixed by the law, however it was possible to estimate their mean values distribution by CT for all of them.
They were represented according to the following colour scale:

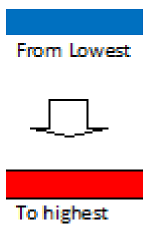


Figure 2a – Kriging mean estimation of aldrin, arsenic and its compounds, atrazine, cadmium and its compounds in Alto Adige province CTs.

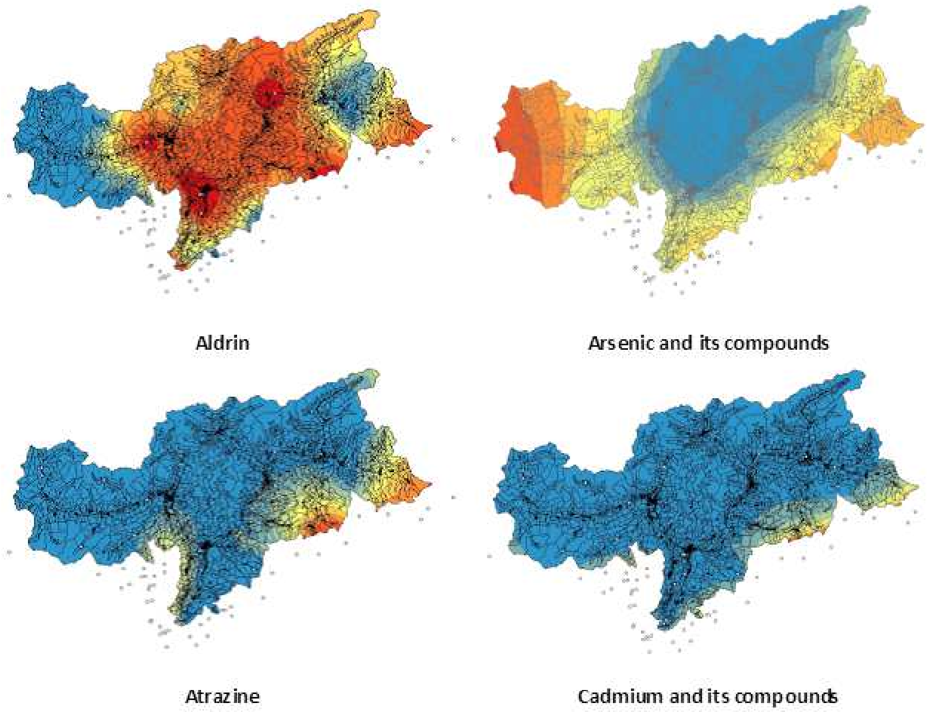


Figure 2b - Kriging mean estimation of chlorpyrifos, chromium and its compounds, DDTpp1, dieldrin in Alto Adige province CTs.

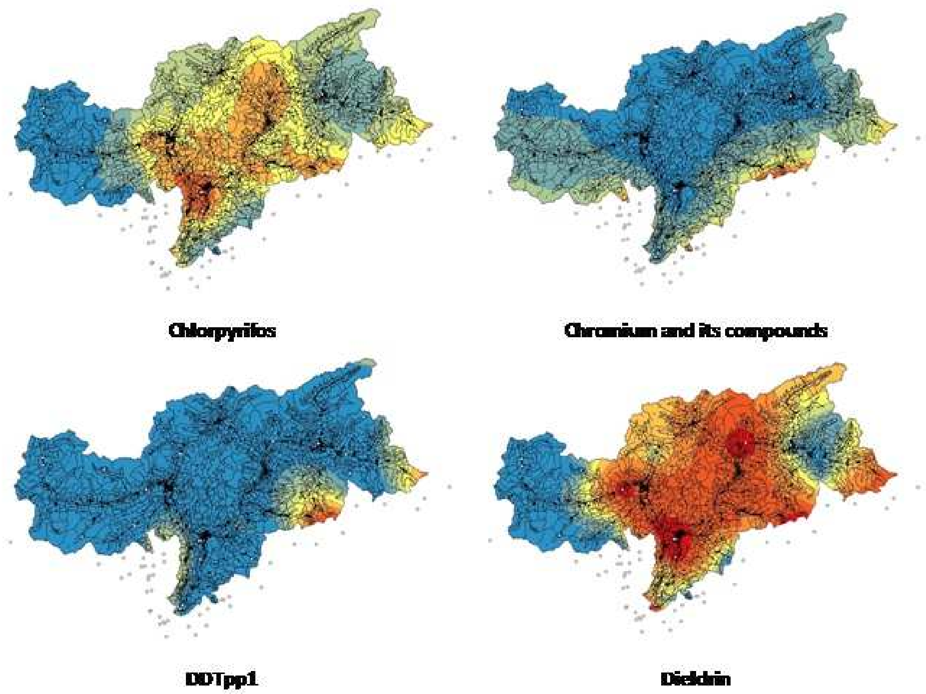


Figure 2c - Kriging mean estimation of endrin, lead and its compounds, mercury and its compounds, nickel and its compounds in Alto Adige province CTs.

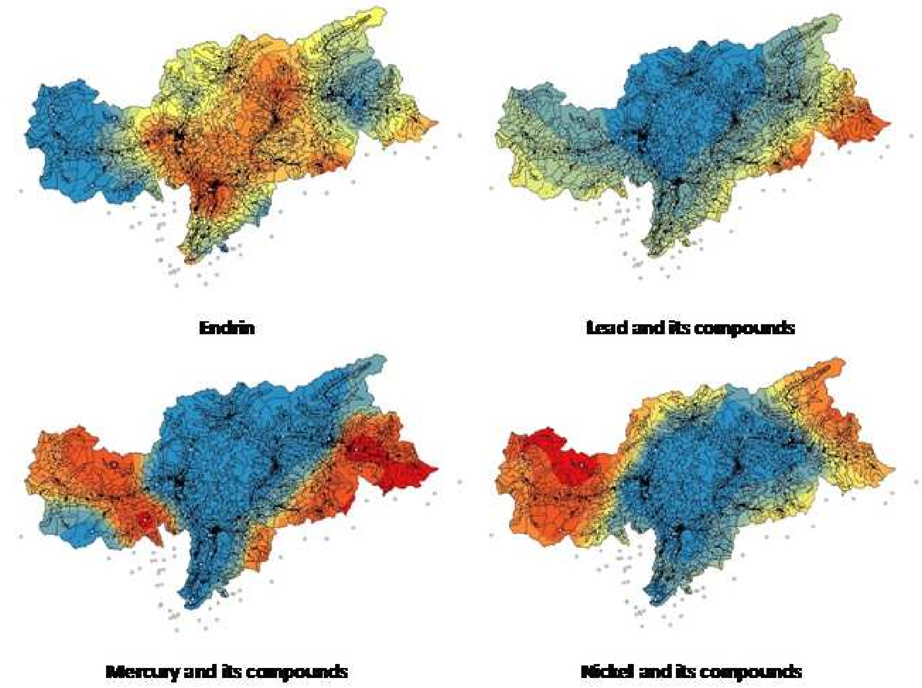
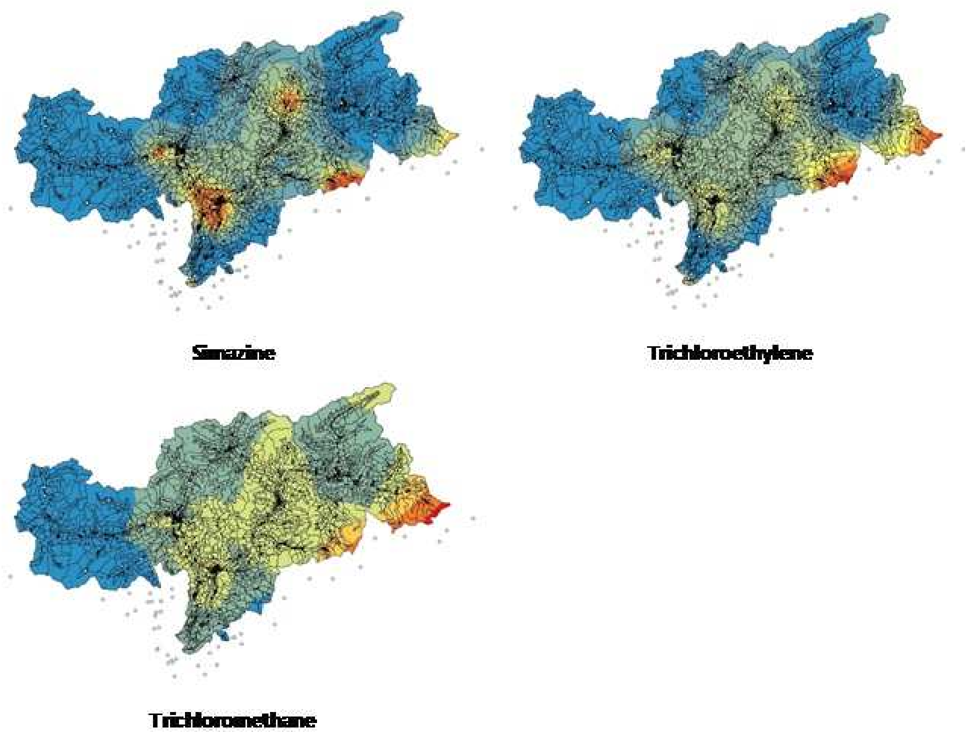


Figure 2d - Kriging mean estimation of sinnazine, trichloroethylene, trichloromethane in Alto Adige province CTs



.

6.2 – SIRs estimates

Figure 3 reports the geographic distribution of the estimated SIRs by the two standardizations, grouped by the following thresholds:

SIR = 0.00

0.00 < SIR ≤ 1.00

1.00 < SIR ≤ 1.50

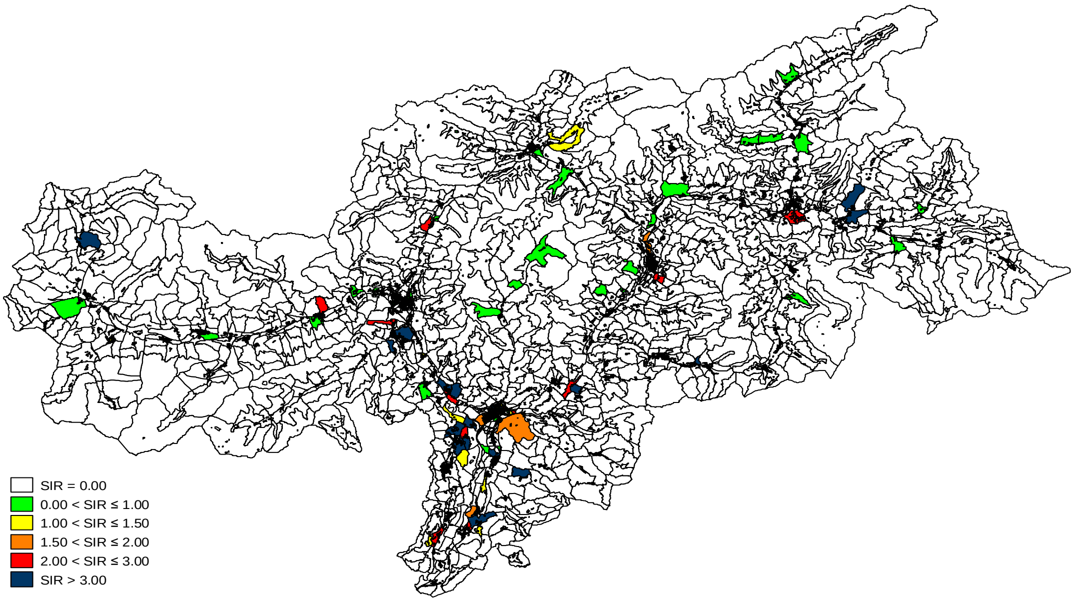
1.50 < SIR ≤ 2.00

2.00 < SIR ≤ 3.00

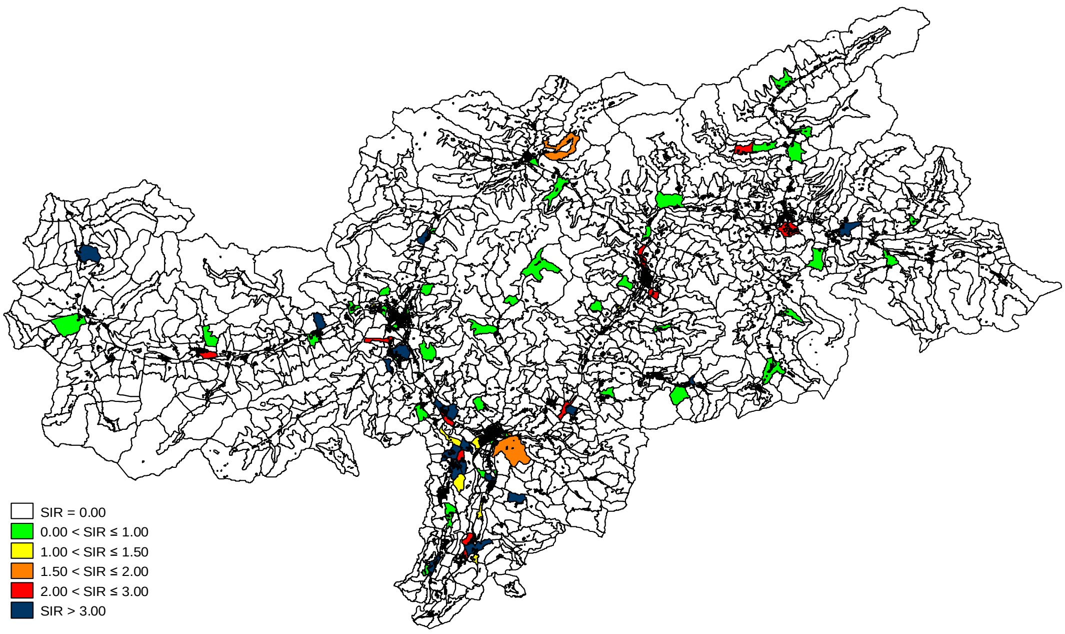
SIR > 3.00

Figure 3 – Estimated SIRs by the two standardizations.

Direct standardization



Indirect standardization



6.3 – SARAR estimates

Considering the results from the SARAR models applied at directly and indirectly standardized SIRs, the potential presence of geographic clusters was found in both of the models (Table 2).

The estimated λ (the coefficient which indicates the value of the spatial auto-correlation) was positive and significant, indicating moderate SAR dependence in the observed SIR. In other words, the observed SIR for a given CT was affected by the SIRs of the neighbouring CTs.

The estimated ρ (the coefficient which indicates the value of the spatial auto-correlation error) coefficient was negative and significant, indicating moderate SAR dependence in the error term. In other words, an exogenous shock to one CT would cause moderate changes in the observed SIR in the neighboring CTs.

The observed effects for λ and ρ coefficients remarked the potential presence of local clusters; there was no significant effects from the EDI (zeditestweighted) in both models.

Table 2 – Results of the SARAR Models.

Direct Standardization: World

sir049f_wor_final	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sir049f_wor_final						
zeditestweighted	-.0466084	.0442601	-1.05	0.292	-.1333566	.0401398
_cons	-.0466326	.3309047	-0.14	0.888	-.6951939	.6019288
lambda						
_cons	4.926798	.9758536	5.05	0.000	3.01416	6.839436
rho						
_cons	-5.988781	.8816294	-6.79	0.000	-7.716743	-4.260819

Indirect standardization

sir049f_ind_final	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sir049f_ind_final						
zeditestweighted	-.1167034	.0619538	-1.88	0.060	-.2381306	.0047238
_cons	.0779195	.3716032	0.21	0.834	-.6504094	.8062484
lambda						
_cons	4.684553	.9287472	5.04	0.000	2.864242	6.504864
rho						
_cons	-5.362461	.7907083	-6.78	0.000	-6.912221	-3.812701

The mean estimated SIR by the SARAR models was reported in Table 3: They were quite similar through the two standardization; confidence intervals were partially overlapped.

Therefore no significant difference in cluster distribution should be assumed and cluster analysis will be applied only for one of the two standardizations (namely the direct standardization).

Table 3 – Mean estimated SIRs by the SARAR models.

	Mean	Std. Err.	[95% Conf.	Interval]
estsir049f_wor_final	.1610468	.0147647	.1321022	.1899914
estsir049f_ind_final	.1968976	.0200317	.1576277	.2361674

6.4 – SaTScan Estimates

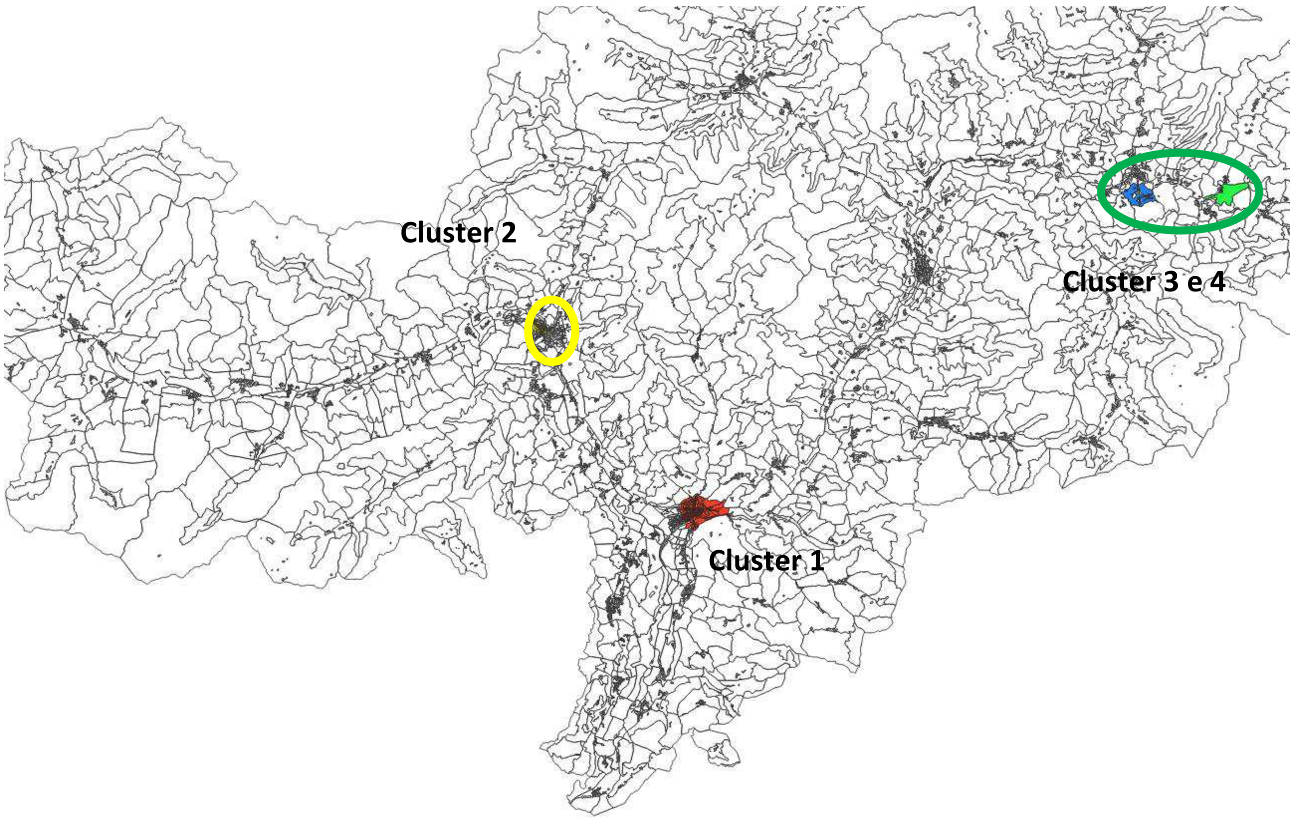
Starting from the previous results, the potential geographic clusters were defined by SaTScan software, applying the following model and conditions to the World standardized data.

Four clusters were identified; the number of cases observed within the clusters was 160, equal to 21.6% of the considered observed cases.

Cluster 1 was statistically significant and regarded most of the Bolzano city. Cluster 2 was not statistically significant but very interesting in terms of position and number of cases; it was located in the western part of Merano city. It should be investigated. Clusters 3 and 4 concerned Riscone, a hamlet of Brunico near Plan de Corones, and Rasun di Sotto, a hamlet in the municipality of Rasun at the crossroads between the Pusteria Valley and the Anterselva Valley. They were not statistically significant and the combination of low number of cases and low population density suggested a low relevance of these two clusters.

The application of the same analysis to indirectly standardized data didn't produce appreciable difference (data not shown). Graphically, the position of the clusters was represented in Figure 4.

Figure 4 – Geographic representation of the four clusters.



6.5 – Association of the estimated pollutants distribution with the presence of geographic clusters of breast cancer incidence

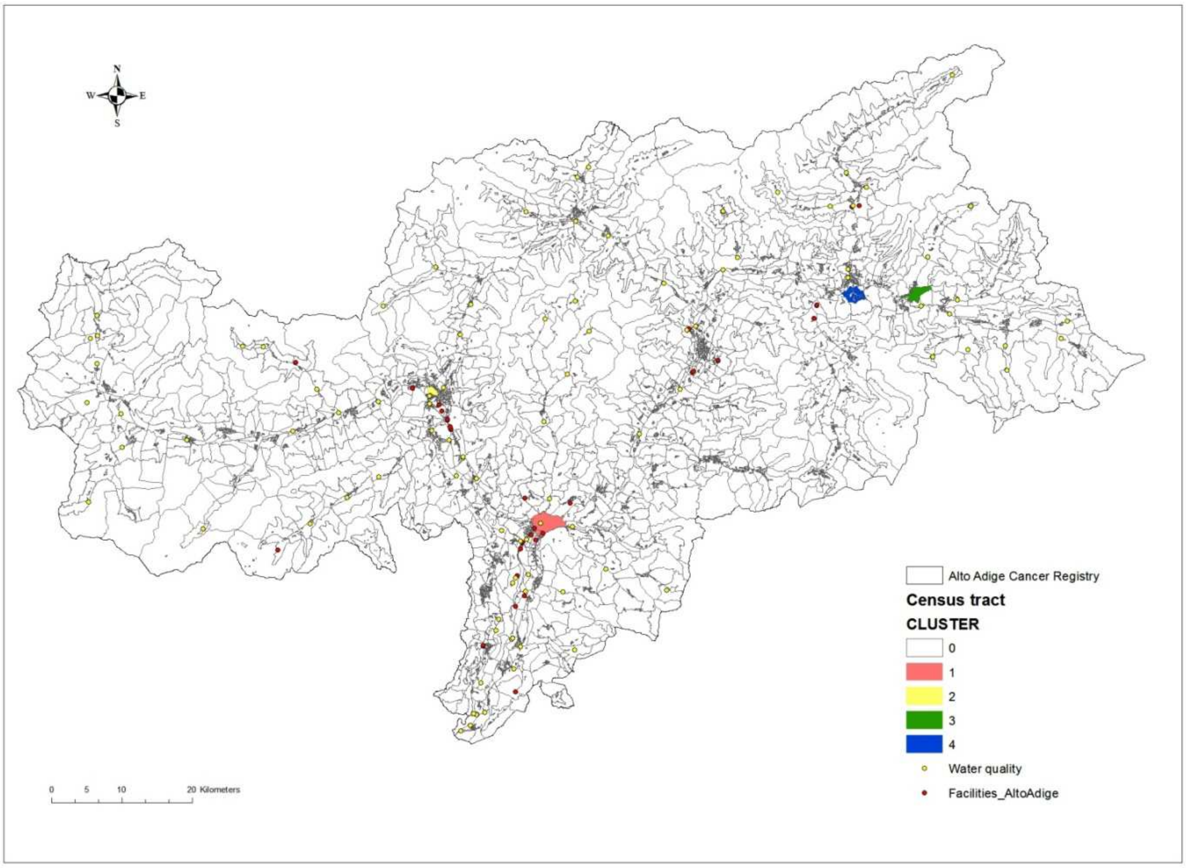
A first evaluation of the probability of connection between water wells and facilities with their positioning in the cluster was tested by a χ2 test ($p<0.05$) these geo-referenced by CTs sources and their belonging or not to a cluster. The result is the following:

Pearson chi2(3) = 7.2483 **Pr = 0.050 (both standardization)**

There is a borderline statistical significance between the presence of the sources and their belonging to the clusters, therefore further investigation should be performed.

In the following map, t clusters identified for Alto Adige CR are combined with the industrial and other emitting activities facilities registered by E-RPTR (red points) and the geo-referenced ground water wells from Waterbase Water Quality (yellow points) (Figure 5).

Figure 5 - Clusters, industrial and other emitting activities facilities registered in E-RPTR and geo-referenced ground water wells from Waterbase Water Quality by CT.



Finally, the possible connection between pollutant variables and clusters was evaluated by logistic regression model (Relative Risk Ratios, $p<0.05$). Table 4 reports only the elements statistically significant associated with the clusters.

Table 4 – Logistic regression model: clusters and water pollutants association.

CLUSTER	**Elements**	**RRR**	***p***	**[95% Conf. Interval]**	
No cluster		*ref.*			
Cluster 1	Arsenic and its compounds	0.000000434	*0.000*	0.00	0.000024
	Chlorpyrifos	8.60	*0.000*	1.95	78.00
	Endrin	3.04	*0.000*	2.12	5.13
	Simazine	3.26	*0.000*	1.33	32.00
	Trichloroethylene	2.19	*0.000*	1.40	11.90
	Trichloromethane	6.61	*0.000*	3.99	10.90
Cluster 2	Aldrin	1.82	*0.000*	1.24	2.67
	Arsenic and its compounds	0.000000303	*0.000*	0.00	0.000479
	Dieldrin	2.02	*0.000*	1.35	3.01
	Endrin	8.03	*0.000*	3.29	19.60
	Simazine	1.66	*0.000*	1.23	12.10
	Trichloroethylene	7.30	*0.000*	2.31	23.10
	Trichloromethane	0.000000121	*0.000*	0.00	0.0007
Cluster 3	*The clusters 3 and 4 were not statistically significant defined; no connection with elements was found*				
Cluster 4					

The pollutants reporting Relative Risk Ratios (RRR) > 1 remarked a positive association with the clusters, while the RRR < 1 stressed a negative association.

7.0 – CONCLUSION

Our pilot environmental study has shown that cancer registries can investigate the relationship between the exposure of environmental contaminants and the effects on health using open source databases of persistent environmental pollutants, different methodologies (statistical models and use of geographical maps) using different software.

Studies of this nature can assist in public health decision-making. In particular, geographical analyses of the distribution of risk factors can be useful in prioritizing preventive measures. Disease mapping is useful for health service provision and targeting interventions if avoidable risk factors are known.

Specifically, the pilot study considers the data of environmental monitoring surveys carried out over the years by the Bolzano environmental protection agency for the assessment in the deep waters of the presence of dangerous chemical contaminants and even pesticides. These data are also included in the different National reports of pesticides in water (ISPRA reports) with environmental monitoring data of the various Italian territories. It is highlighted that in this specific territory the different persistent environmental contaminants considered never exceed the national legislative limits for the protection of deep waters, and this is an absolutely important data also for the question of public health.

The possible connection between the different persistent contaminants present in underground water bodies and clusters was evaluated by the logistic regression model (Relative Risk Ratios, $p < 0.05$) found a positive relative risk (RR> 1). A positive association with the clusters was found for contaminants such as Chlorpyrifos (RR=8.60), Endrin (RR=8.03), Trichloroethylene (RR=7.30), Trichloromethane (RR=6.61). These results show only an association and not a causal relationship between incident clusters and water pollutants. The causality should be investigated with deeper and more structured analyses, however the biological plausibility of such association is corroborated by literature.

In fact, endrin and other chlorinated pesticides are endocrine disruptors, the hypothesis is that organochlorine residues in breast cancer tissue generate an estrogenic microenvironment that can influence biological behavior of the tumor through ERα activation and ERα-dependent proliferation.

A recent metanalysis conducted by Rocha et al. found exposure to endocrine disruptors increased the risk of breast cancer. In addition, chlorinated solvents such as trichloroethylene and trichloromethane are carcinogens for several organs.

On the contrary, arsenic and its compounds in both clusters and trichloromethane in cluster 2 recorded a “strange” protective effect (0.000000434 and 0.000000303 for the first pollutant and 0.000000121 for the second one), notwithstanding their proven cancerousness.

Such protective effect could a statistical randomness linked to the quantity of those specific substances very low according to the estimation of the interpolating model. This result strengthen the concept that similar studies highlights associations rather than causal relationships, but that they are extremely useful to find clues to possible public health problems and to address specific more in-depth studies.

References

1. Aballay LR, Díaz Mdel P, Francisca FM, Muñoz SE. Cancer incidence and pattern of arsenic concentration in drinking water wells in Córdoba, Argentina. Int J Environ Health Res, 2012, 22(3):220-231.
2. Aoki Y. Evaluation of in vivo mutagenesis for assessing the health risk of air pollutants. Genes Environ., 2017, Apr 1;39:16.
3. Asadi S, Vuppala P, Redd, MA. Remote sensing and GIS techniques for evaluation of groundwater quality in municipal corporation of Hyderabad (Zone-V), India, Int. J. Environ. Res. Public Health, 2007, 4, 45–52. doi: 10.3390/ijerph2007010008.
4. Colak EH, Yomralioglu T, Nisanci R, Yildirim V, Duran C. Geostatistical analysis of the relationship between heavy metals in drinking water and cancer incidence in residential areas in the Black Sea region of Turkey. J Environ Health, 2015, 77(6),6-93. pmid: 25619041.
5. Dios M, Moran M, Carrera F, Pombo C, Souto JA, Casares JJ, Diaz AA, Saez A. PRTRval: a software tool for the validation of E-PRTR emissions data. International Journal of Environment and Pollution (IJEP), 2014, Vol. 55, No. 1/2/3/4.
6. Egbueri JC, Mgbenu CN. Chemometric analysis for pollution source identification and human health risk assessment of water resources in Ojoto Province, southeast Nigeriael modelling approach. Applied Water Science, 2020, Vol. 10.
7. Fernández-Navarro P, García-Pérez J, Ramis R, Boldo E, López-Abente G. Industrial pollution and cancer in Spain: An important public health issue. Environ Res, 2017, Nov;159:555-563. doi: 10.1016/j.envres.2017.08.049.
8. Guillaume E, Pornet C, Dejardin O, Launay L, Lillini R, Vercelli M, Marí-Dell'Olmo M, Fernández Fontelo A, Borrell C, Ribeiro AI, Pina MF, Mayer A, Delpierre C, Rachet B, Launoy G. Development of a cross-cultural deprivation index in five European countries. J Epidemiol Community Health. 2016 May;70(5):493-9. doi: 10.1136/jech-2015-205729. PMID: 26659762.
9. Hackeloeer A, Klasing K, Krisp JM, Meng L. Georeferencing: a review of methods and applications. Annals of GIS. 2014. 20 (1): 61–69. doi:10.1080/19475683.2013.
10. ISPRA: Rapporto Nazionale pesticidi nelle acque dati 2009 – 2010. RAPPORTI 175/2013.
11. ISPRA: Rapporto Nazionale pesticidi nelle acque dati 2015 – 2016. RAPPORTI 282/2018.
12. Jephcote, C, Mah, A. Regional inequalities in benzene exposures across the European petrochemical industry: A Bayesian multilevel modelling approach. Environ Int. 2019 Nov;132:104812. doi: 10.1016/j.envint.2019.05.006.
13. Kahn LG, Philippat C, Nakayama SF, Slama R, Trasande L. Endocrine-disrupting chemicals: implications for human health. Lancet Diabetes Endocrinol. 2020 Aug;8(8):703-718. doi: 10.1016/S2213-8587(20)30129-7. PMID: 32707118; PMCID: PMC7437820.
14. Khosravi R, Eslami H, Almodaresi SA, Heidari M, Fallahzadeh RA, Taghavi M, Khodadadi M, Peirovi R. Use of geographic information system and water quality index to assess groundwater quality for drinking purpose in Birjand City, Iran. Desalination and Water Treatment, 2017, 67, 74–83. doi: 10.5004/dwt.2017.20458.
15. Lillini R, Tittarelli A, Bertoldi M, Ritchie D, Katalinic A, Pritzkuleit R, Launoy G, Launay L, Guillaume E, Žagar T, Modonesi C, Meneghini E, Amati C, Di Salvo F, Contiero P, Borgini A, Baili P. Water and Soil

Pollution: Ecological Environmental Study Methodologies Useful for Public Health Projects. A Literature Review. Rev Environ Contam Toxicol. 2021 Apr 19. doi: 10.1007/398_2020_58. Online ahead of print. PMID: 33866420.

16. Lin WC, Lin YP, Wang YC, Chang TK, Chiang LC. Assessing and mapping spatial associations among oral cancer mortality rates, concentrations of heavy metals in soil, and land use types based on multiple scale data. Int J Environ Res Public Health, 2014, 11(2), 2148-2168. doi: 10.3390/ijerph110202148.
17. López-Abente G, García-Pérez J, Fernández-Navarro P, Boldo E, Ramis R. Colorectal cancer mortality and industrial pollution in Spain. BMC Public Health. 2012 Aug 1;12:589. doi: 10.1186/1471-2458-12-589.
18. López-Abente G, Locutura-Rupérez J, Fernández-Navarro P, Martín-Méndez I, Bel-Lan A, Núñez O. Compositional analysis of topsoil metals and its associations with cancer mortality using spatial misaligned data. Environ Geochem Health, 2018, 40(1), 283-294. doi: 10.1007/s10653-016-9904-3.
19. Munoz de-Toro M, Durando M, Beldoménico PM, Beldoménico HR, Kass L, García SR, Luque EH. Estrogenic microenvironment generated by organochlorine residues in adipose mammary tissue modulates biomarker expression in ERα-positive breast carcinomas. Breast Cancer Research, 2006, volume 8, Article number:R47.
20. Naing NN. Easy way to learn standardization : direct and indirect methods. Malays J Med Sci. 2000 Jan;7(1):10-15. PMID: 22844209 PMCID: PMC3406211.
21. Núñez O, Fernández-Navarro P, Martín-Méndez I, Bel-Lan A, Locutura Rupérez JF, López-Abente G. Association between heavy metal and metalloid levels in topsoil and cancer mortality in Spain. Environ Sci Pollut Res Int, 2017 24(8), 7413-7421. doi: 10.1007/s11356-017-8418-6.
22. Núñez O, Fernández-Navarro P, Martín-Méndez I, Bel-Lan A, Locutura JF, López-Abente G. Arsenic and chromium topsoil levels and cancer mortality in Spain. Environ Sci Pollut Res Int, 2016 23(17),17664-17675. doi: 10.1007/s11356-016-6806-y.
23. Pistocchi A, Dorati C, Aloe A, Ginebreda A, Marcé R. River pollution by priority chemical substances under the Water Framework Directive: A provisional pan-European assessment. Science of The Total Environment Volume 66220 April 2019 Pages 434-445.
24. Purdue MP. Trichloroethylene and cancer. J Natl Cancer Inst. 2013 Jun 19;105(12):844-6. doi: 10.1093/jnci/djt131. Epub 2013 May 30. PMID: 23723421; PMCID: PMC3687372.
25. Rocha PRS, Oliveira VD, Vasques CI, Dos Reis PED, Amato, AA. Exposure to endocrine disruptors and risk of breast cancer: A systematic review. Crit Rev Oncol Hematol. 2021 May;161:103330. doi: 10.1016/j.critrevonc.2021.103330. Epub 2021 Apr 20. PMID: 33862246.
26. Rüttenauer T. Neighbours matter: A nation-wide small-area assessment of environmental inequality in Germany. Soc Sci Res 2018 Feb;70:198-211. doi:10.1016/j.ssresearch.2017.11.009.
27. Shabbir R, Ahmad SS. Use of geographic information system and water quality index to assess groundwater quality in Rawalpindi and Islamabad, Arab. J. Sci. Eng., 2015, 40, 2033–2047.
28. Shaddick G, Ranzi A, Thomas ML, Aguirre-Perez R, Dunbar Bekker-Nielsen M, Parmagnani F, Martuzzi M. Towards an assessment of the health impact of industrially contaminated sites: waste landfills in Europe. Epidemiol Prev. 2018 Sep-Dec;42(5-6 Suppl 1):69-75.

29. Sörme L., Palma V, Finnveden G. Using E-PRTR data on point source emissions to air and water—First steps towards a national chemical footprint. Environmental Impact Assessment Review. Volume 56, January 2016, Pages 102-112.
30. Vach W. Regression Models as a Tool in Medical Research. USA: Chapman and Hall/CRC. ISBN-13 : 978-1466517486. 2012.
31. Yoshida H, Clavreul J, Scheutz C, Christensen TH. Influence of data collection schemes on the Life Cycle Assessment of a municipal wastewater treatment plant. Water Research Volume 56, 1 June 2014, Pages 292-303.

WASABY Deliverables (available at http://www.wasabysite.it/)