WASABY - Water And Soil contamination and Awareness on Breast cancer risk in Young Women

Paolo Baili – Alessandro Borgini – Roberto Lillini

CALL

CALL FOR PROPOSALS FOR A PILOT PROJECT

ON Primary prevention courses for girls living in areas with higher risk of breast cancer.

2. OBJECTIVES

- The project should identify a certain number of geographical regions in the European Union presenting higher breast cancer rates due to identifiable environmental factors. The selection of the pilot areas to implement the project should keep a geographical diversity and include several Member States;

- The key objective of this pilot project is to establish the scientific evidence for environmental problems in relation to areas with higher breast cancer rates, with the aim to improve the situation for the citizens living in this area in a cross-sectorial manner.

- Priority in the analysis and on the actions to be proposed should be given to youngest cohorts of age in full compatibility and complementary with the European Union breast cancer policy on prevention and early detection of breast cancer.

- The project will focus on soil and water contamination by polluting agents and include specifically illegal disposal of toxic waste as well as on industrial areas. The pilot project is targeted at areas, as described above, where breast cancer incidence rates are statistically much higher than average.

- The project will be carried out in form of a study in order to review the evidence of polluting agents and their impact on health, specifically breast cancer with full scientific justification of any action proposed.

APPLICATION HISTORY

- How we interpreted the call requirements:
  - Identify areas at risk for BC → use population-based data from cancer registries

CALL

CALL FOR PROPOSALS FOR A PILOT PROJECT

ON Primary prevention courses for girls living in areas with higher risk of breast cancer.

2. OBJECTIVES

- The project should identify a certain number of geographical regions in the European Union presenting higher breast cancer rates due to identifiable environmental factors. The selection of the pilot areas to implement the project should keep a geographical diversity and include several Member States;

- The key objective of this pilot project is to establish the scientific evidence for environmental problems in relation to areas with higher breast cancer rates, with the aim to improve the situation for the citizens living in this area in a cross-sectorial manner.

- Priority in the analysis and on the actions to be proposed should be given to youngest cohorts of age in full compatibility and complementary with the European Union breast cancer policy on prevention and early detection of breast cancer.

- The project will focus on soil and water contamination by polluting agents and include specifically illegal disposal of toxic waste as well as on industrial areas. The pilot project is targeted at areas, as described above, where breast cancer incidence rates are statistically much higher than average.

- The project will be carried out in form of a study in order to review the evidence of polluting agents and their impact on health, specifically breast cancer with full scientific justification of any action proposed.

APPLICATION HISTORY

- How we interpreted the call requirements:
  - Identify areas at risk for BC → use population-based data from cancer registries
  - Promote BC preventive actions for girls in areas at higher risk → not necessarily on W&S contaminants

CALL

CALL FOR PROPOSALS FOR A PILOT PROJECT

ON Primary prevention courses for girls living in areas with higher risk of breast cancer.

2. OBJECTIVES

- The project should identify a certain number of geographical regions in the European Union presenting higher breast cancer rates due to identifiable environmental factors. The selection of the pilot areas to implement the project should keep a geographical diversity and include several Member States;

- The key objective of this pilot project is to establish the scientific evidence for environmental problems in relation to areas with higher breast cancer rates, with the aim to improve the situation for the citizens living in this area in a cross-sectorial manner.

- Priority in the analysis and on the actions to be proposed should be given to youngest cohorts of age in full compatibility and complementary with the European Union breast cancer policy on prevention and early detection of breast cancer.

- The project will focus on soil and water contamination by polluting agents and include specifically illegal disposal of toxic waste as well as on industrial areas. The pilot project is targeted at areas, as described above, where breast cancer incidence rates are statistically much higher than average.

- The project will be carried out in form of a study in order to review the evidence of polluting agents and their impact on health, specifically breast cancer with full scientific justification of any action proposed.

APPLICATION HISTORY

- How we interpreted the call requirements:
  - Identify areas at risk for BC → use population-based data from cancer registries
  - Promote BC preventive actions for girls in areas at higher risk → not necessarily on W&S contaminants
  - Call for pilot studies → ecological studies on association between WS and BC risk
  - Call in public health → use the example of BC to promote cancer registry activities on spatial analysis

CALL

CALL FOR PROPOSALS FOR A PILOT PROJECT

ON Primary prevention courses for girls living in areas with higher risk of breast cancer.

2. OBJECTIVES

- The project should identify a certain number of geographical regions in the European Union presenting higher breast cancer rates due to identifiable environmental factors. The selection of the pilot areas to implement the project should keep a geographical diversity and include several Member States;

- The key objective of this pilot project is to establish the scientific evidence for environmental problems in relation to areas with higher breast cancer rates, with the aim to improve the situation for the citizens living in this area in a cross-sectorial manner.

- Priority in the analysis and on the actions to be proposed should be given to youngest cohorts of age in full compatibility and complementary with the European Union breast cancer policy on prevention and early detection of breast cancer.

- The project will focus on soil and water contamination by polluting agents and include specifically illegal disposal of toxic waste as well as on industrial areas. The pilot project is targeted at areas, as described above, where breast cancer incidence rates are statistically much higher than average.

- The project will be carried out in form of a study in order to review the evidence of polluting agents and their impact on health, specifically breast cancer with full scientific justification of any action proposed.

WHAT DO WE EXPECT?

- We do not want to compare CR data among countries (as in EUROCARE, CIVC, CONCORD, …)
- We would like to try to compare CR data in a country, but also this request is not mandatory

- So each CR can have different data
- Different methods can be applied

In the first 2 pages of the application

- "Strange" call: an Horizon subject for a Public Health DG Santé call
  - Water & Soil (WS) → Breast cancer (BC)

Table 1.1. Population prevention strategies for breast cancer [Colditz G, NPJ Breast Cancer, 2015]

Establish and maintain childhood dietary intake of vegetables, fruits, whole grains
Reduce animal protein
Increase and maintain physical activity through childhood, adolescence, and adult years
Avoid and reduce weight gain during the adult years
Achieve a sustained 10% weight loss among postmenopausal women
Reduce or eliminate alcohol intake between menarche and first birth
Reduce or eliminate alcohol intake among adult women throughout the life course
Reduce use of estrogen plus progestin hormone therapy (a known carcinogen)

No action connected with environmental risk is listed in the above table, as no public health consensus was reached to date on the relationship between environment and breast cancer. […]

Given the above, and in consideration of the scopes of a DG SANTE call (i.e. excluding analytic studies such as cohort or case-control studies which could aim to evaluate cause-and-effect relationship), the proposed action does not have the ambition to change the public health view on the relationship between environmental risk factors and breast cancer.

Who are we?

Fondazione IRCCS Istituto Nazionale dei Tumori

WP1: Coordination	Paolo Baili, Camilla Amati
WP4: Data management	Roberto Lillini
WP7: Environment& BC	Paolo Contiero, Giovanna Tagliabue, Alessandro Borgini, Martina Bertoldi, Andrea Tittarelli, Carlo Modonesi
WP2: Dissemination	David Ritchie
WP3: Evaluation	Alexander Katalinic, Ron Pritzkuleit
WP5: Deprivation indexes	Guy Launoy, Elodie Guillaume, Ludivine Launoy, Oliver Dejardin
WP6: Methods &Analysis	Tina Zagar, Vesna Zadnik

ECL Association of European Cancer Leagues

Universität zu Lübeck

UNICAEN université de Caen Basse-Normandie

Onkološki inštitut Ljubljana / Institute of Oncology Ljubljana

WASABY HIGHLIGHTS

POPULATION-BASED SPATIAL ANALYSIS OF INCIDENCE RISK

USE OF DEPRIVATION INDEX IN SPATIAL ANALYSIS

BREAST CANCER

ONLINE COURSES ON PREVENTION

CONNECTION BETWEEN ENVIRONMENTAL DB AND CANCER REGISTRY DB

WATER & SOIL POLLUTION

Target Group- 1

Public Health Managers

POPULATION BASED DATA

- WASABY aims to study, standardize and promote new analytical methods that optimize the use of population-based CR data for the improvement of the European citizens' health
- European cancer information system improvement as an added value

RISK MAPS

- National and regional health policy makers receive a tool enabling them to target breast cancer preventive actions
- WASABY method can be translated to any cancer site

Target Group-2

Cancer Registries

- WP4, WP5, WP6 and WP7 are built around a topic of public health high relevance: **the study of cancer risk using geographical analysis at a European level**

In WASABY, cancer registries are actors and receivers: their activity is promoted and innovative methodology is developed

1) the definition of a replicable model of spatial analysis to be used by CRs
2) methodology for correlating CRs data with environmental data

Target Group-3

Young Girls & Adolescents

- In Year 3, WASABY is expected to launch online courses and a smartphone app aimed to increasing awareness about breast cancer risk among young girls

WHAT DO WE EXPECT?

- Increase spatial analysis activities across European CRs
- Improve the use of deprivation indexes in spatial analysis
- Enlarge the utilization of the European Deprivation Index
- Prepare a useful web instrument to promote BC-preventive actions among young girls
- Pilot the connection between databases of WS pollutants and CR databases

OUR GOALS

Specific Objective	Identify geographical areas in the European Union with higher breast cancer risk (through spatial analysis of CR data)
Process Indicator(s)	**Target**
Number of CRs participating in the CR survey (see methods and means)	>=20
Number of CR countries participating in the survey	>=7
Output Indicator(s)	**Target**
Number of CRs sending data for spatial analysis on breast cancer	>=15
Number of CR countries sending data for spatial analysis on breast cancer	>=6
Outcome/Impact Indicator(s)	**Target**
Number of CRs performing spatial analysis for the first time	>=4
Number of CR countries performing spatial analysis for the first time	>=1

OUR GOALS

Specific Objective	Deprivation indexes estimates to be considered in spatial analysis of breast cancer incidence as SES mediator
Process Indicator(s)	**Target**
See SO1 process indicators (CR survey will cover also SES topic)	-
Output Indicator(s)	**Target**
Number of CRs for which deprivation indexes is used in breast cancer spatial analysis	>=15
Number of CR countries for which deprivation indexes is used in breast cancer spatial analysis	>=6
Number of countries for which EDI is used in breast cancer spatial analysis	>=5
Outcome/Impact Indicator(s)	**Target**
Number of CRs using deprivation indexes for the first time	>=4
Number of CR countries using EDI for the first time	>=2

OUR GOALS

Specific Objective	Review the evidence of polluting agents and their impact on breast cancer
Process Indicator(s)	**Target**
Number of published scientific articles considered	>=50
Output Indicator(s)	**Target**
Number of water/soil pollutants studied in relation with breast cancer	>=15
Outcome/Impact Indicator(s)	**Target**
Number of reports on literature review	>=1

OUR GOALS

Specific Objective	Environmental study on correlation between soil and water contamination and breast cancer risk
Process Indicator(s)	**Target**
Number of European databases with available data on water/soil pollutants	>=2
Number of national databases with available data on water/soil pollutants	>=3
Number of pollutant indicators available in different countries	>=5
Output Indicator(s)	**Target**
Number of pollutant indicators available for pilot environmental study	>=1
Outcome/Impact Indicator(s)	**Target**
Number of pilot environmental studies performed	>=1

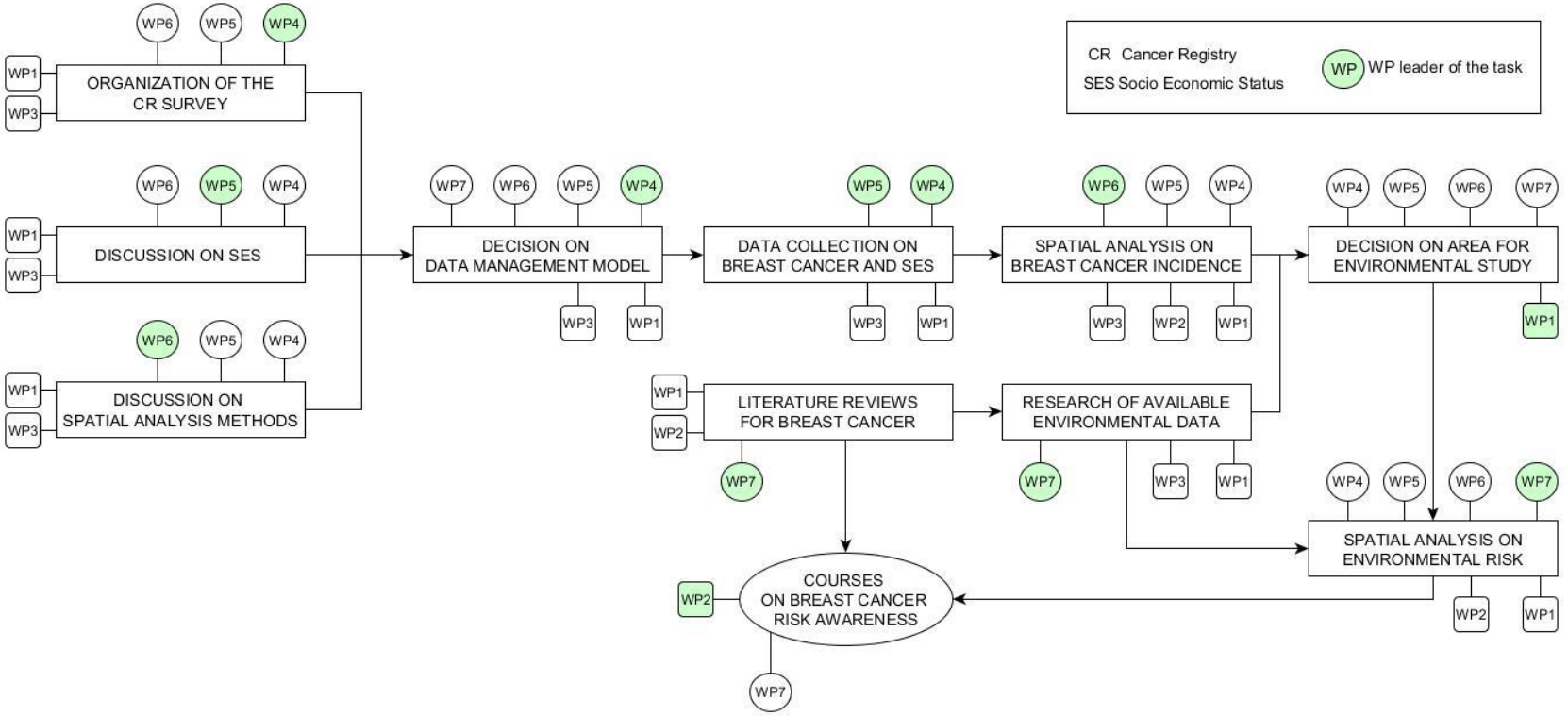
OUR GOALS

Specific Objective	Design courses on breast cancer risk factor awareness for young girls/adolescents
Process Indicator(s)	**Target**
Number of target countries where promoting the online course	>=5
Number of participants per target country taking part in “alpha test” of online course	>=15
Number of participants per target country taking part in final “beta test” of online course	>=15
Mean age of participants to testing phase	14.5 yrs
Output Indicator(s)	**Target**
Number of unique visits to online course per target country	>=5,000
Engagement rate of Facebook page (or other social media metric)	>=50%
Mean age of participants to online course	14.5 yrs
Outcome/Impact Indicator(s)	**Target**
Number of unique visitors completing online course per target country	>=4,000
Number of downloads of additional/complementary information by unique visitor after completion of course per country OR Number of unique visitors following hyperlinks to referral information / partner webpages after completion of course per country	>=1,000

WPs

WP	Title	Description
1	Coordination of the project	Actions undertaken to manage the project and to make sure that it is implemented as planned
2	Dissemination of the project	Actions undertaken to ensure that the results and deliverables of the project will be made available to the target groups
3	Evaluation of the project	Actions undertaken to verify if the project is being implemented as planned and reaches the objectives
4	Data management	Actions undertaken to manage and coordinate data for and from WP-5, WP-6 and WP-7
5	Deprivation indexes	Construction of indicators of social and economic deprivation by the smallest geographical areas of countries
6	Methods and analysis	Actions undertaken to design and validate a method for studying and mapping the variation of cancer risk within CRs areas
7	Environmental risk factors & breast cancer	Actions undertaken to study the relation between pollutants contamination and cancer risk

STRUCTURE


CR Cancer Registry
SES Socio Economic Status
WP
WP leader of the task
ORGANIZATION OF THE CR SURVEY
DISCUSSION ON SES
DISCUSSION ON SPATIAL ANALYSIS METHODS
DECISION ON DATA MANAGEMENT MODEL
DATA COLLECTION ON BREAST CANCER AND SES
SPATIAL ANALYSIS ON BREAST CANCER INCIDENCE
DECISION ON AREA FOR ENVIRONMENTAL STUDY
LITERATURE REVIEWS FOR BREAST CANCER
RESEARCH OF AVAILABLE ENVIRONMENTAL DATA
SPATIAL ANALYSIS ON ENVIRONMENTAL RISK
COURSES ON BREAST CANCER RISK AWARENESS

WASABY's strategy for each CR

- WASABY in the first year will try to involve only those CRs able to analyze data by the smallest unit (S.U.) and not by municipality
- We have to stress the importance of geocoding the residence address at date of diagnosis
- After 10 months we have to evaluate
  - Number of CRs able to participate
  - Number of countries able to participate
  - Number of CRs performing for the first time the spatial analysis
- According to the target indicators (in the contract) we can decide to change our strategy and to enlarge the number of CRs

Protocol

- Each participating CR is required to provide information on **breast cancer cases (coded as C50 according to the ICD-10) diagnosed during a specific ten-year period (to be defined separately for each participating CR, e.g. 2001-2010)**, together with age at diagnosis (or 5-year age groups), morphology and data on the place of residence at the time of diagnosis (exact x and y coordinates or SU).

- **Socio Economic Status (SES) data will be collected as main confounder in the spatial analysis**: National or European Deprivation indexes by SU will be utilized.

- Maps of incidence will be estimated in order to identify CR SU characterized by higher-than-CR average rates.

- Population files **at the same geographic level on which that CR intends to study the incident cases**.

Protocol

Before the start of the data collection, every CR is required provide the following general information, which will be added to the ones collected with the preliminary survey:

- The calendar years by which the CR can provide incidence data at the most disaggregated geographic level (exact x and y coordinates or SU).
- The calendar year by which such geographic level has changed (e.g., census tract changes between two different Census data collection).
- Any confidentiality problems likely to arise if/when pursuing the approval to the Ethical Committees, locally.
- Detail limits in publishing maps (see confidentiality above).

Current participating Cancer Registries

Nation	Cancer Registry	Nation	Cancer Registry
Belgium	Belgium	*Poland*	*Kracow*
Germany	Bremen	Poland	Kielce
Germany	Schleswig-Holstein	Poland	Silesia
Italy	Napoli 3 South	Portugal	Central Portugal
Italy	Palermo	Portugal	Northern Portugal
Italy	*Parma*	Slovenia	Slovenia
Italy	Ragusa	Spain	Basque Country
Italy	Siracusa	Spain	Castellon-Valencia
Italy	Trento	Spain	Girona
Italy	Umbria	Spain	Granada
Italy	Varese	Spain	Murcia
Lithuania	Lithuania	UK	Northern Ireland
Poland	Greater Poland		

Shapefiles

For every CR, WASABY needs 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 (e.g., Varese.shp, Varese.dbf, Varese.shx), stored in the same directory, with mandatory and optional files.

File name	Description	Data type
(CR area).shp	Shape format; the feature geometry itself	Alphanumeric
(CR area).shx	Shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly	Alphanumeric
(CR area).dbf	Attribute format; columnar attributes for each shape, in dBase IV format	Alphanumeric

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

Other optional files, regarding spatial features not reported in the .dbf file, can be added but are not needed for a correct representation.

In the .dbf file an information about the minimum geo-coding level must be reported (i.e., census block, municipality, etc.).

Identification of risk areas across European Countries Background

We will conduct this project using open source GIS software such as QGis [http://www.qgis.org/en/site/].

Furthermore, a single ArcGis desktop 10.0 license is also available for INT group and it will be used during the project to improve, if necessary, geographical and spatial analysis.

GIS system software allows users to create maps with many layers (raster or vector) using different map projections. The vector data is stored as either point, line, or polygon-feature.

Different kinds of raster images are supported, and the software can georeference images. Maps can be assembled in different formats and for different uses.

SES and other confounders

Since this study includes different European countries, *it is important that measurement of* ***socioeconomic deprivation be comparable or at least transferable between different European countries*****, despite their socio-cultural differences**, to improve the comparability and reproducibility across countries.

The European Deprivation Index (EDI) measures the social environment in a comparable manner across countries, **despite the differences in the census variables available**, and to incorporate the social and cultural specificities of each country concerned.

Individual factors, e.g. ethnicity, family history, age, reproductive factors, alcohol intake, weight, physical activity, hormone therapy and oral contraceptives, have been found to influence the risk of breast cancer.

Adherence to organized screening programmes in areas covered by cancer registries, lead to an increment of incidence in those areas; **such information, however, is not available at individual level**.

Where possible, **information on adherence to organized cancer screening is to be collected at SU level**. If data are collected only for ages <50 screening adherence is not required.

Spatial analysis

Two different methods of mapping smoothed estimates in WASABY:

- Estimate of a census block level breast cancer incidence risk using **Generalized Additive Models (GAMs)**, a form of non-parametric or semi-parametric regression offering the possibility to analyze contextual data while adjusting for covariates and taking into account spatial autocorrelation. With this model it is possible to estimate the relative risk by adjusting for covariates.

- Estimate of a census block relative breast cancer risk, by using the **Besag, York and Mollié (BYM) model**, since it assumes the existence of two sources of extra variation, one spatial and the other non-spatial. The BYM model can be specified as a generalized linear mixed model (GLMM) with Poisson response variables. The BYM model enables us to obtain smoothed estimates in each sub-area and to estimate the effects of possible explanatory variables, such as the deprivation index.

Open source software is used for data manipulation and statistical analyses such as R and WinBUGS.

Background

A third method have been developed and proposed by Slovenia Cancer Registry, but it is available only for CRs that have/can their population completely geo-coded (x&y coordinates).

In Slovenia the geographic coordinates or cancer cases as well as for population (~ 2 mio) are available.

The CR developed a procedure to estimate the standardized incidence ratio (SIR), based on the georeferenced data.

Data

- Cancer registry of RS:
  1,033 female patients with cervical cancer 2003–2008 (6 years)
- Central Population Register:
  1,030,654 females (in 2006)

PIN
(x,y) coordinates
gender
age

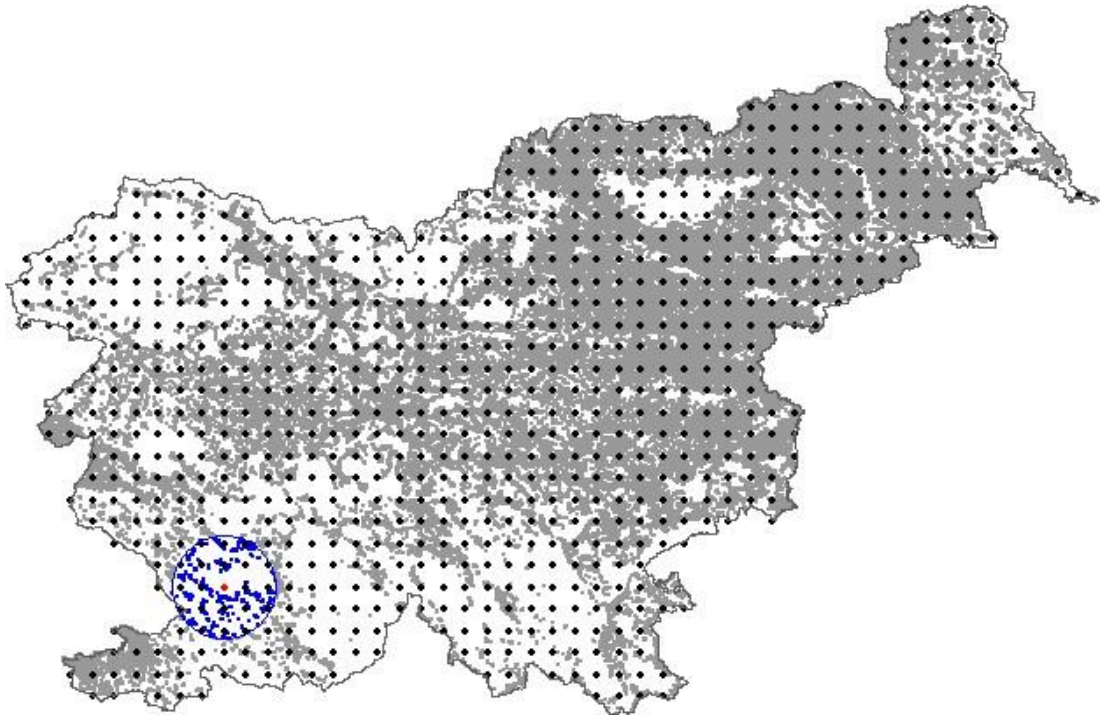
- Surveying and Mapping Authority of the RS:
  193 municipalities
  (valid from 1990 to 2005)

We wanted to compare several types of maps based on the same underlying data set.

Local estimation of SIR – Data requirements

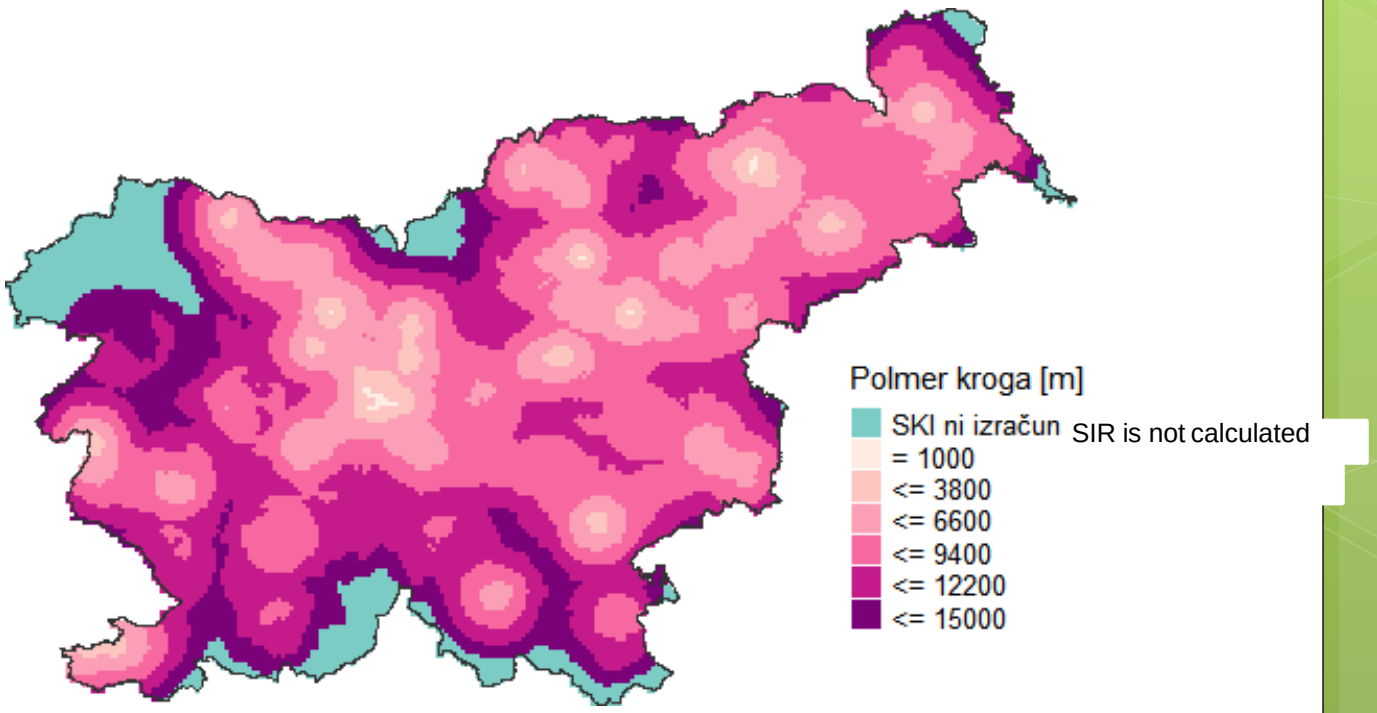
- X and Y coordinates are required for each person included in the analysis – cancer patients and general population.

- 20,614 grid points are 1 km apart.
- Grid points are the centers of circular "moving" window.
- SIR is calculated for each grid point.



Local estimation of SIR – Radius

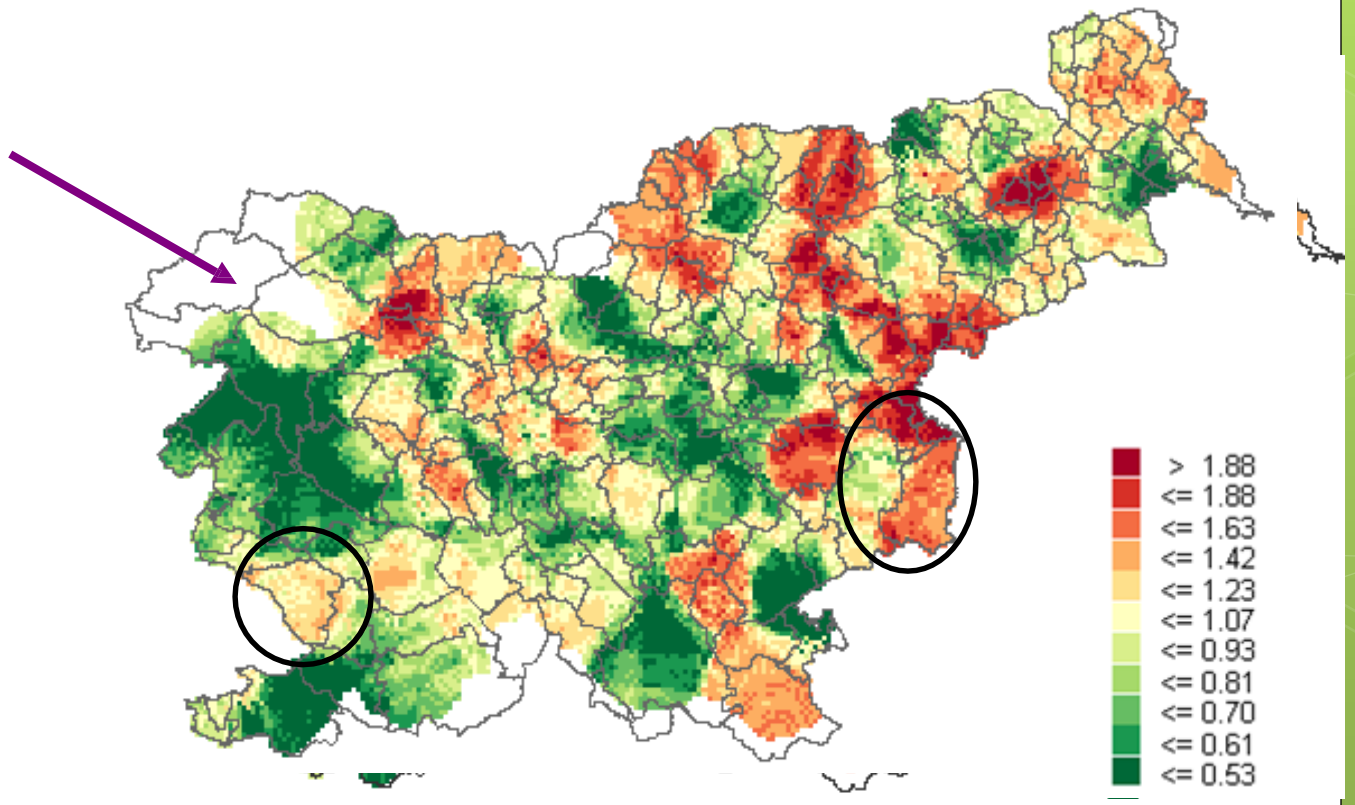
- The circle radius is not fixed in advance but is changing from 1 km (D_{min}) to 15 km (D_{max}) with step 1 km until predetermined minimum population of 7,000 is reached.
- The local SIR estimates are based on (approximately) the same population size, which makes them more stable and more comparable.



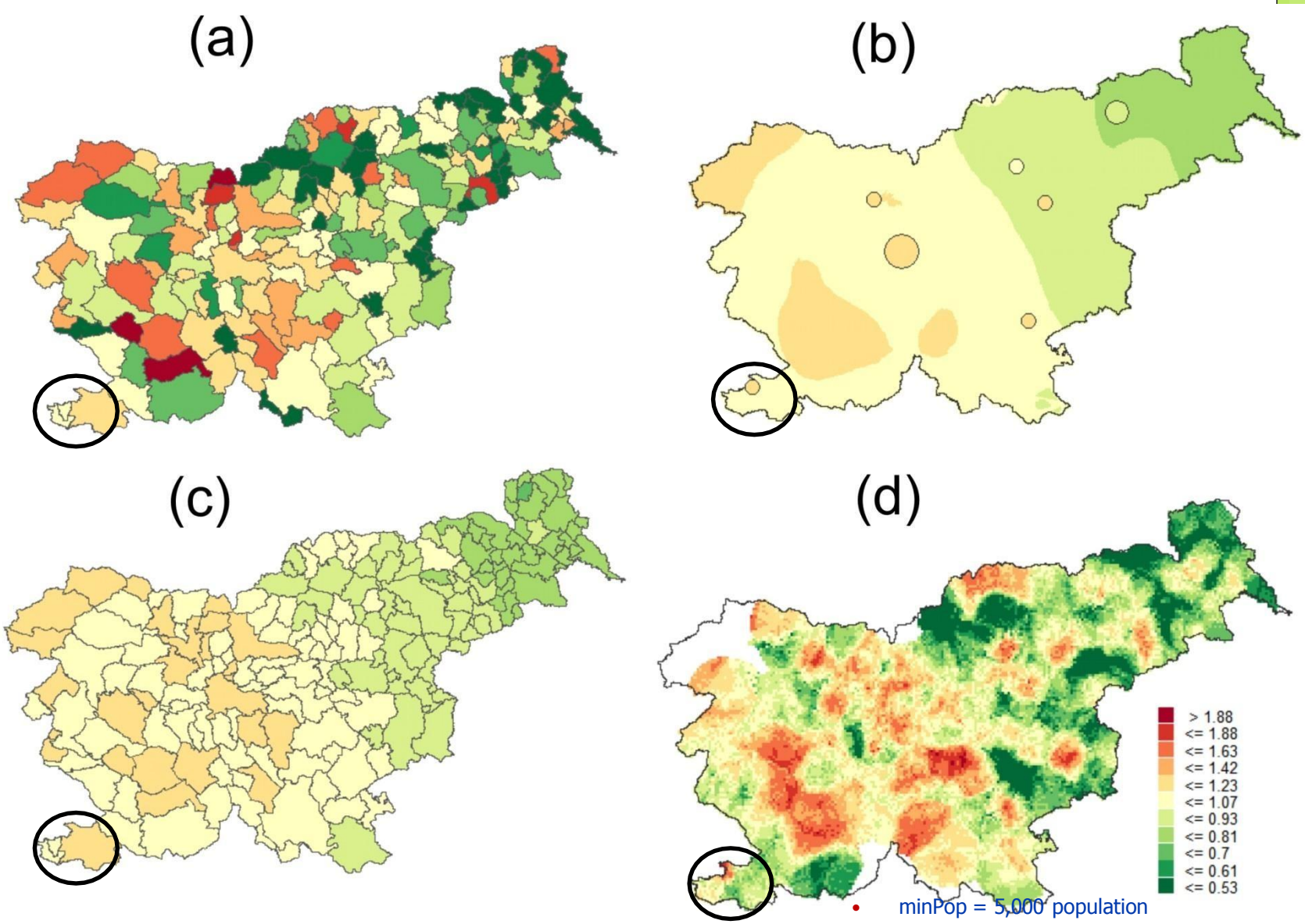
Local estimation of SIR – The map

- minPop = 7,000 population
- Radius is increasing from 1 km to 15 km

sparsely populated areas



Breast cancer incidence for females, Slovenia 2002–2004



- minPop = 5,000 population
- Radius is increasing from 1 km to 15 km

Local estimation of SIR – DISADVANTAGES

- Data availability and its quality is the first practical constraint for the method selection.
- The exact coordinates for the residence of population are usually not available for the routine analyses.
- Problems with estimation near the area border.
- Producing a map of locally estimated SIRs is laborious.

Local estimation of SIR – ADVANTAGES

- Map of local SIR estimates is a high resolution map and emphasizes local patterns.
- The local SIR estimates are based on (approximately) the same population size which makes them more stable and more comparable.
- We are not using the arbitrary administrative units (regarding cancer mapping).

Environmental risk factors and breast cancer

1. To review the scientific literature on breast cancer and environmental factors. Studies with both positive and negative results will be considered.

2. To search for European and national environmental databanks.

3. To design a protocol for pilot environmental study.

Data collection and analysis of pilot environmental study in the area will be identified by the Steering Committee.

Literature review on environmental factors related with breast cancer

1. Research of scientific articles in the PubMed, Scopus, and Embase, and other databases, using the term ‘‘**breast cancer**’’ in combination with **environmental persistent organic pollutants (POPs)** chosen in the Wasaby European project like:

   - **Polychlorinated Biphenyls (PCBs)**
   - **DDT, DDD, DDE, HCB and Chlordane and Dioxins (TCDD)**
   - **PAHs**
   - **Perfluoroalkylated substances (PFAs)**
   - **Triazine**
   - **Heavy Metals (In principle Cadmium)**

Literature review on environmental factors related with breast cancer

Persistent organic pollutants (POPs) are a group of **heterogeneous compounds** of both natural and **anthropogenic origin** with **highly persistent** and **bioaccumulative properties** and common properties like **lipophilic compounds** accumulate in fat, resulting **in bioaccumulation** and **biomagnification** up the food chain.

There is growing concern that POPs **may increase breast cancer risk** due to their **xenoestrogenic properties.** They are also **resistant to photolytic**, **biological or chemical degradation** and remain in the environment for a long period (Damstra 2002).

However, **most of the studies have examined exposure** to the pollutants **after diagnosis of breast cancer**, overlooking exposure during critical windows of vulnerability.

Literature review on environmental factors related with breast cancer

2. We're going to put **in excel schematic table** the scientific articles found on exposure to various environmental persistent contaminants (in particular in water and soil) and the risk of breast cancer as written in the Wasaby proposal**:**

- the name of the researchers of the scientific and scientific journal;
- the place and the study design;
- the exposure to the main contaminants;
- the years of the study interest;
- the relative risk or odds ratio
- the relevant comments for the scientific article.

Literature review on environmental factors related with breast cancer

At the moment we have found the following scientific articles that correlate the BC with the different **environmental (POPs) .**

Contaminants	Number of scientific articles
DDT, DDD, DDE, HCB and PCBs	9
Dioxins (TCDD)	12
PAHs	8
Perfluoroalkylated substances (PFAs)	5
Triazine	4
Heavy Metals (in principle Cadmium)	7
Review BC and POPs	17
Meta-Analysis	1
	TOT. 63

Water & soil environmental databases

1. In Wasaby we will use a geochemical baseline data to examine the distribution of high levels of persistent organic pollutants (POPs) and heavy metals in Europe.

2. In this project we need specific geographic information systems (GIS) database containing European mapping and environmental monitoring data in water and soil of the main environmental contaminants that persist in the identified environmental matrices.

3. In order to identify possible indicators already available in the mentioned databases, we start to investigate some online databases to be used during the project.

Water & soil environmental databases

Code	Name	Argument	Web address	Organization	Countries included	Years covered	Collection points
1	CENSUS_UNITS_2011_RG	European administrative boundaries	http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units	EUROSTAT	33 (EU and EFTA)	ND	ND
2	Global Administrative Area	European administrative boundaries	http://www.gadm.org/	University of California	World	ND	ND
3	Waterbase groundwater	Water quality data	http://www.eea.europa.eu/data-and-maps/data/waterbase-groundwater-6	EEA	38	1990-2012	28,000
4	European Pollutant Release and Transfer Register E-PRTR	Industrial sites Waste	http://prtr.ec.europa.eu/	EEA	33 (EU and EFTA)	2007-2011	30,000
5	European Environment Agency	Environmental pollutants	http://www.eea.europa.eu/	EEA	33 (EU and EFTA)	ND	ND
6	Geochemical Atlas of Europe	Soil quality data	http://weppi.gtk.fi/publ/foregsatlas/	FOREGS	33 (EU and EFTA)	ND	ND
7	EuroGeoSurveys	Water and Soil resources	http://www.eurogeosurveys.org/		33 (EU and EFTA)	ND	ND
8	Geochemical Atlas of Europe	Soil quality data	http://eusoils.jrc.ec.europa.eu/content/heavy-metals-topsoils#tabs-0-description=1	European Soil Data Centre (ESDAC)	26	2008	1,588 topsoil samples

Water & soil environmental databases

Pollutants	Indicator	Database code	Matrix	Measure periodicity	Spatial resolution
Cadmium	µg/l	Waterbase_groundwater_v14	Water	annual	Sample points
	mg/kg	Geochemical Atlas of Europe	Soil	2008	5km grid
DDT	µg/l	Waterbase_groundwater_v14	Water	annual	Sample points
Atrazine	µg/l	Waterbase_groundwater_v14	Water	annual	Sample points
PAH	µg/l	Waterbase_groundwater_v14	Water	annual	Sample points

Example of data and maps

The **FOREGS Geochemical Baseline Mapping** Programme was initiated in 1998 to provide high quality environmental geochemical baseline data in Europe (link:**http://weppi.gtk.fi/publ/foregsatlas/**

The **FOREGS databases and material archives comprise**:

- archived sample materials (topsoil, subsoil, ecc.);
- access databases for field observations;
- analytical data files;
- databases of combined field and analytical data;
- GIS layers;
- work maps and tables;
- collections of field photographs;.

The **Geological Survey of Finland (GTK)** has been responsible for database management and map production for the FOREGS group.

Example of data and maps

The **FOREGS database structure.**

Basic information collected for **all sample sites** included:

• sample identifiers (using agreed FOREGS sampling code);

• date of collection and name of sampler;

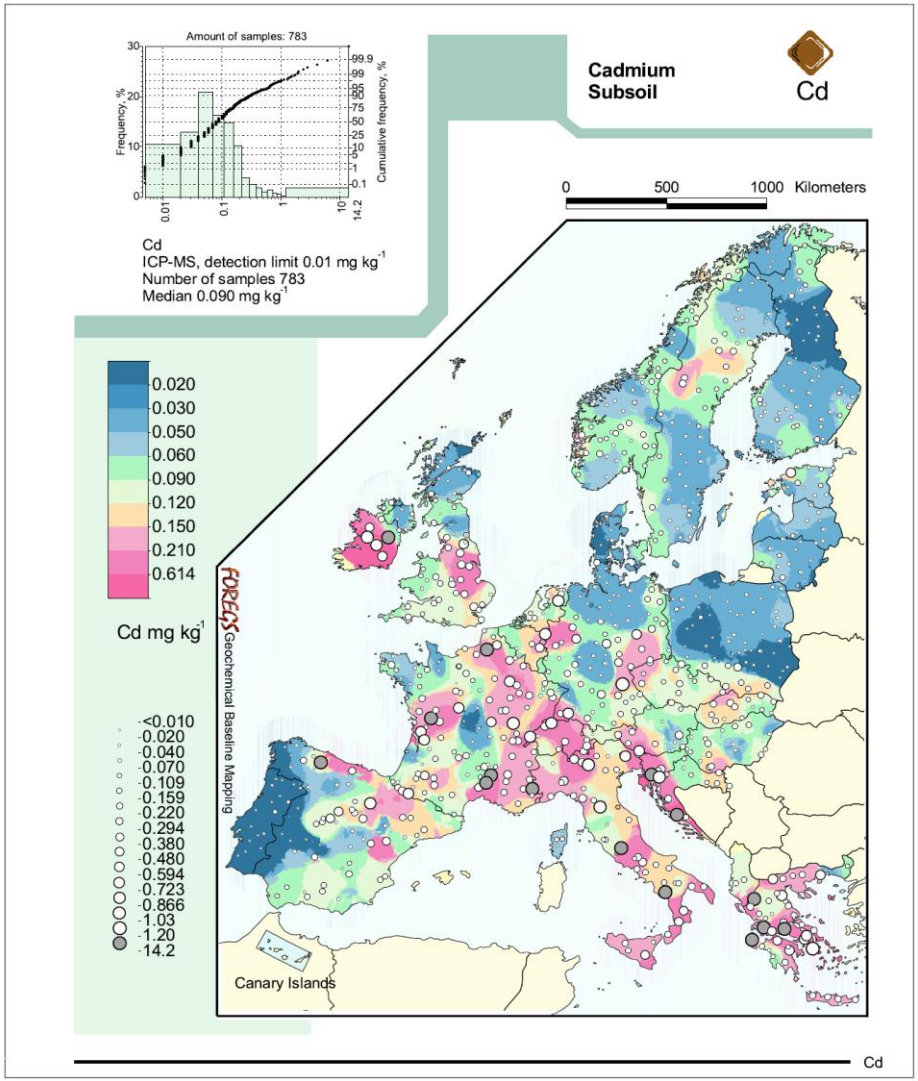
• sampling site location, including geographical region, map sheet, sampling coordinates (easting and northing and latitude and longitude) and altitude;

• number of subsites and a site description, including, landscape/topography, land use (agriculture, pasture, forest, wetland, *etc), bedrock lithology and type of overburden)*.

Analytical data from **nine European geochemical laboratories** were combined with the field data.

GIS data layers were then created and preliminary dot maps, basic tables and distribution graphs prepared.

Distribution of cadmium in subsoil.



Map Produced using the **ArcView GIS®**;

- The data were interpolated to generate a regular grid with a 6 km x 6 km output cell size.
- A 10-grade colour scale was selected to present the distribution and concentration of Cd in mg/Kg.
- Dot size scaled for Subsoil Cadmium concentrations.
- The histogram and cumulative distribution function curve for Cadmium in subsoil.

Under construction

Wasaby logo

WASABY

Wasaby website: **www.wasabysite.it**

Appendix

Data to be collected for each cancer registry

- Primary invasive female breast cancer (ICD9 174*, ICD10 C50*), selected from cancer registries data during a specific ten years period (ex: 2001 to 2010) are included in the project. It is mandatory to collect data with age at diagnosis less than 50 years of age, while it is not mandatory to collect data for all ages. Synchronous and metachronous breast cancer cases must be counted once. Cancer registration criteria must follow European Network of Cancer Registries (ENCR) rules.

- Residence addresses at diagnosis retrieved from the National or local Security system or from the personal data reference of each registry will be collected.

Data can be collected in two different modalities.

Option 1 – Individual level

	Variable name	Description	Data type	Mandatory
BREAST CANCER VARIABLES	CR	Cancer Registry name	Alphanumeric variable	Yes
	PATIENT_ID	Patient identification code assigned by Cancer Registry.	Numeric / Alphanumeric variable	Yes
	DATE OF DIAGNOSIS	Incidence date based on histological or cytological confirmation of the malignancy	DD/MM/YYYY	Yes
	DATE OF BIRTH	Date of birth of the patient	DD/MM/YYYY	Yes (one of the two variables)
	AGE	Age at diagnosis	Numeric variable	
	ICDO3_M	ICDO3 morphology code of incident case	Alphanumeric variable	Yes
	SUBTYPE_ER	Estrogen Receptor value at diagnosis	Numeric / Alphanumeric variable	No
	SUBTYPE_PGR	Progesterone Receptor value at diagnosis	Numeric / Alphanumeric variable	No
	SUBTYPE_HER2	HER-2 value at diagnosis	Numeric / Alphanumeric variable	No
	SUBTYPE_KI67	KI-67 value at diagnosis	Numeric / Alphanumeric variable	No
	SUBTYPE_FISH	FISH value at diagnosis	Numeric / Alphanumeric variable	No

Option 1 – Individual level

GEOGRAPHIC VARIABLES	OPTION A	X	Longitude coordinate referred to the address where the patient was residing at the moment of the breast cancer diagnosis	Numeric variable	
		Y	Latitude coordinate referred to the address where the patient was residing at the moment of the breast cancer diagnosis	Numeric variable	
		Reference	The coordinate system used for X and Y: UTM WGS84 32N vs. UTM ED 1950 32N	Alphanumeric variable	
	OPTION B	SU	Smallest administrative unit (SU) where the patient was residing at the moment of the breast cancer diagnosis	Alphanumeric variable	
	OPTION C	MUNICIPALITY_ CODE	Code of the Municipality where the patient was residing at the moment of the breast cancer diagnosis	Alphanumeric variable	
		MUNICIPALITY	Name of the Municipality where the patient was residing at the moment of the breast cancer diagnosis	Alphanumeric variable	Yes, data from one option

Option 2 – Aggregated level

		Variable name	Description	Data type	Mandatory
BREAST CANCER VARIABLES		CR	Cancer Registry name	Alphanumeric variable	Yes
		YEAR DIAGNOSIS	Incidence year based on histological or cytological confirmation of the malignancy	Numeric variable	Yes
		AGE	Age class at diagnosis	Alphanumeric variable	Yes
		ICDO3_M	ICDO3 morphology code of incident case	Alphanumeric variable	Yes
		SUBTYPE_ER	Estrogen Receptor value at diagnosis	Numeric / Alphanumeric variable	No
		SUBTYPE_PGR	Progesterone Receptor value at diagnosis	Numeric / Alphanumeric variable	No
		SUBTYPE_HER2	HER-2 value at diagnosis	Numeric / Alphanumeric variable	No
		SUBTYPE_KI67	KI-67 value at diagnosis	Numeric / Alphanumeric variable	No
		SUBTYPE_FISH	FISH value at diagnosis	Numeric / Alphanumeric variable	No

Option 2 – Aggregated level

GEOGRAPHIC VARIABLES	OPTION B	SU	Smallest administrative unit (SU) where the patient was residing at the moment of the breast cancer diagnosis	Alphanumeric variable	Yes, data from one option
	OPTION C	MUNICIPALITY_	Code of the Municipality where the patient was residing at the moment of the breast cancer diagnosis	Alphanumeric variable	
		CODE			
		MUNICIPALITY	Name of the Municipality where the patient was residing at the moment of the breast cancer diagnosis	Alphanumeric variable	
DATA		NR_CASES	Number of primary invasive female breast cancer by all the previous variables	Numeric variable	Yes

Population files

For every CR, WASABY needs the reference population at the same geographic level on which that CR intends to study the incident cases. More specifically, the population files must contain the female population data by 5-year age groups, calendar year within time period and SU (sub-areas refer to the smallest geographical area for which required data are available and may be different across countries). All the variables are mandatory.

Variable name	Description	Data type
CR	Cancer Registry name	Alphanumeric variable
AGE_CLASS	5-year age class	Numeric/Alphanumeric variable
YEAR	Calendar year	Numeric/Alphanumeric variable
REF_DATE	Reference date of population data (1st Jan, 31st Dec, ecc)	Date/Alphanumeric variable
SU	MUNICIPALITY_CODE OR SU INDICATED IN THE FILE WITH GEOGRAPHIC DATA (SEE PAGES 5 OR 6)	Alphanumeric variable
POP	Female population by 5-year age groups, calendar year within time period and sub-area on which the incidence data would be estimated	Numeric

SES and other confounders

Variable name	Description	Data type	Mandatory
COUNTRY	Country name	Alphanumeric variable	Yes
SUB_AREA	MUNICIPALITY_CODE OR SU INDICATED IN THE FILE WITH GEOGRAPHIC DATA (SEE PAGES 5 OR 6)	Alphanumeric variable	Yes
SES_SCALE	European Deprivation Index or specific national deprivation indices (according to the availability in the specific CR) by SU of incidence data. This is a scale variable	Numeric - Scale	Yes
SES_ORDINAL	European Deprivation Index or specific national deprivation indices (according to the availability in the specific CR) by SU of incidence data, classified by deprivation groups. This is an ordinal variable	Numeric - Ordinal	Yes
SCR_ADH	% of screening adherence by SU of incidence data.	Numeric - Scale	No

Data storage

Two modalities of data storage can be envisaged:

- OPTION 1
  - Data will be centrally stored at FONDAZIONE IRCCS ISTITUTO NAZIONALE DEI TUMORI and, only for the selected number of cancer registries involved in WP6 analyses, data will be shared with the ONKOLOSKI INSTITUT LJUBLJANA
  - Data will be stored individually (but anonymously). If a CR is to send breast cancer cases by SU, data will be stored at aggregated level
  - Data will be stored in a dedicated server not connected to the web, and according to the standard requirements for data security
  - Data handling will be conform with the EC General Data Protection Regulation (2016/679)

- OPTION 2
  - Only results of the analysis (performed by the CR) will be shared in the WASABY project. This will be the case of CRs with the entire population geocoded.