



WASABY

**Project WASABY: WAter and Soil contamination and Awareness on
Breast cancer risk in Young women**

2nd Steering Committee

-

16.12.2020

3

Spatial analysis on breast cancer incidence rates

Participating CR and data provided

Nation	CR	Cases	CR	Cases	Nation
Germany	Schleswig-Holstein	8013	Gironde	2308	France
Italy	Alto Adige	747	Poitou-Charentes	2513	
	Napoli 3 South	1514	Loire-Atlantique et Vendée	3621	
	Palermo	1914	Haute Vienne	477	
	Parma	381	Calvados	1285	
	Ragusa	515	Manche	831	
	Siracusa	587	Somme	1018	
	Trapani	574	Lille et sa région	1286	
	Trento	946	Bas-Rhin	1654	
	Umbria	2008	Haut-Rhin	1253	
	Varese	2679	Doubs et Territoire de Belfort	840	
Lithuania	Lithuania	3797	Cancers gynécologiques de Côte-d'Or	891	
Portugal	Central Portugal	3820	Isère	2223	
	Northern Portugal	5561	Hérault	2075	
Slovenia	Slovenia	2464	Tarn	672	
Spain	Basque Country	3512	Guadeloupe	n.a.	Poland
	Castellon-Valencia	936	Martinique	n.a.	
	Girona	903	Greater Poland	3358	
	Granada	1253	Kielce	915	
	Murcia	2813	Silesia	3651	
	Navarra	1057	Masovia	4323	
UK	Northern Ireland	n.a.	Podkarpackie	1392	
Total 82580					

Summary for each cancer registry

1. Data collection	1
2. Dataset preparation	2
2.1 Merging observed 0-49 cases by Census Tract with shapefile and 2001 EDI dataset	2
2.2 Shapefile and translation in Stata format	2
3. Computing the expected cases and the SIRs by CT	4
3.1 Direct standardization	4
3.2 Indirect standardization	7
4. SARAR models: Direct standardization with world standard	9
4.1 SARAR models: Building of a space-weighted matrix	9
4.2 Inferential estimation with SARAR model (spatial-autoregressive model with SAR disturbances)	9
5. SARAR models: Indirect standardization	12
5.1 SARAR models: Building of a space-weighted matrix	12
5.2 Inferential estimation with SARAR model (spatial-autoregressive model with SAR disturbances)	12
6. SaTScan: Detection of spatial disease clusters	15
7. Bayesian estimation of SIRs	21
7.1 Bayesian estimation of SIRs by BYM models	21
7.2 Bayesian estimation of SIRs by INLA models	21
8. Identification of industrial facilities and water wells	22

1. Data collection & 2. Dataset preparation

Country	Number of CRs	Status of dataset
France	17	 Data collected by CR. Data not yet sent to INT
Germany	1	
Italy	10	Data sent to INT
Lithuania	1	Data sent to INT
Poland	5	Data collected by CR
Portugal	2	Data sent to INT
Slovenia	1	Next presentation
Spain	6	Data sent to INT
Northern Ireland	1	Delay in data transfer agreement signature
Total	44	-

3. Computing the expected cases and the SIRs

Country	Number of CRs	Direct std. performed	Indirect std. performed
France	17	0	0
Germany	1	Locally analyzed	Locally analyzed
Italy	10	9	3
Lithuania	1	0	0
Poland	5	Locally analyzed	Locally analyzed
Portugal	2	2	2
Slovenia	1	Locally analyzed	Locally analyzed
Spain	6	5	5
Northern Ireland	1	0	0
Total	44	16	10

4. SARAR direct & 5. SARAR indirect & 6. SATSCAN

Country	Number of CRs	SARAR & SATSCAN direct std. Performed	SARAR & SATSCAN indirect std. Performed
France	17	0	0
Germany	1	Locally analyzed	Locally analyzed
Italy	10	9	3
Lithuania	1	0	0
Poland	5	Locally analyzed	Locally analyzed
Portugal	2	2	2
Slovenia	1	Locally analyzed	Locally analyzed
Spain	6	5	5
Northern Ireland	1	0	0
Total	44	16	10

7.1 BYM models

An attempt of developing BYM models for many (or all) areas analyzed at WP4 is on-going.

The chosen software is CARBayes: Spatial Generalised Linear Mixed Models for Areal Unit Data, a module of R specifically developed for analyzing ecological data by defined geographical polygons (i.e., Census Tracts) (Lee, 2015-2020).

The main issue is the excess of zero-cases areas that characterizes our data. The last version of CARBayes (5.2) should deal with it, but we do not have yet succeeded in applying it.

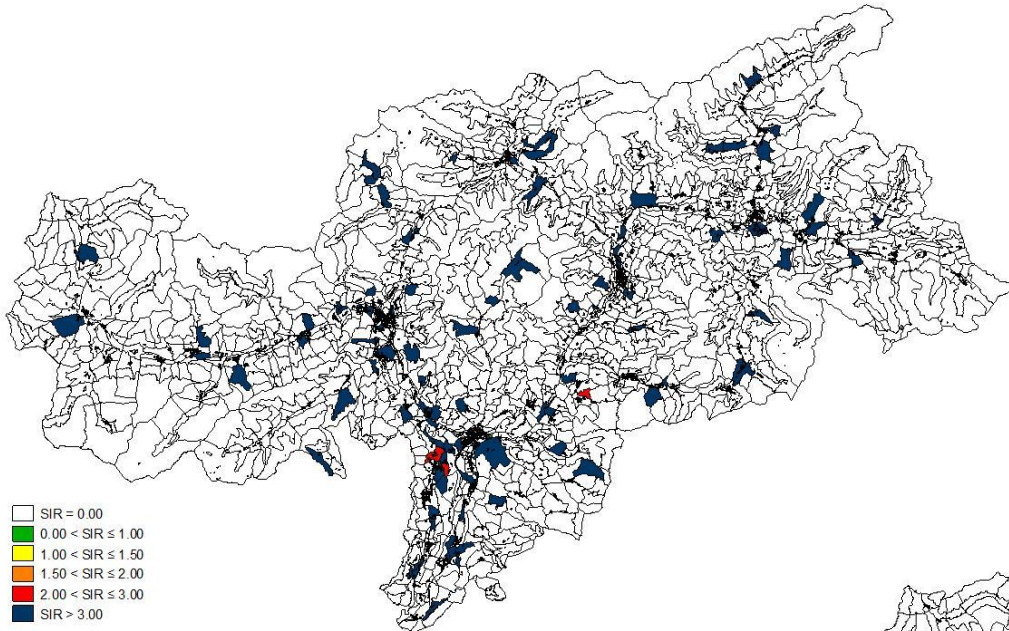
We are in contact with the author, to try to solve this issue.

The module can be found at:

<https://cran.r-project.org/web/packages/CARBayes/index.html>

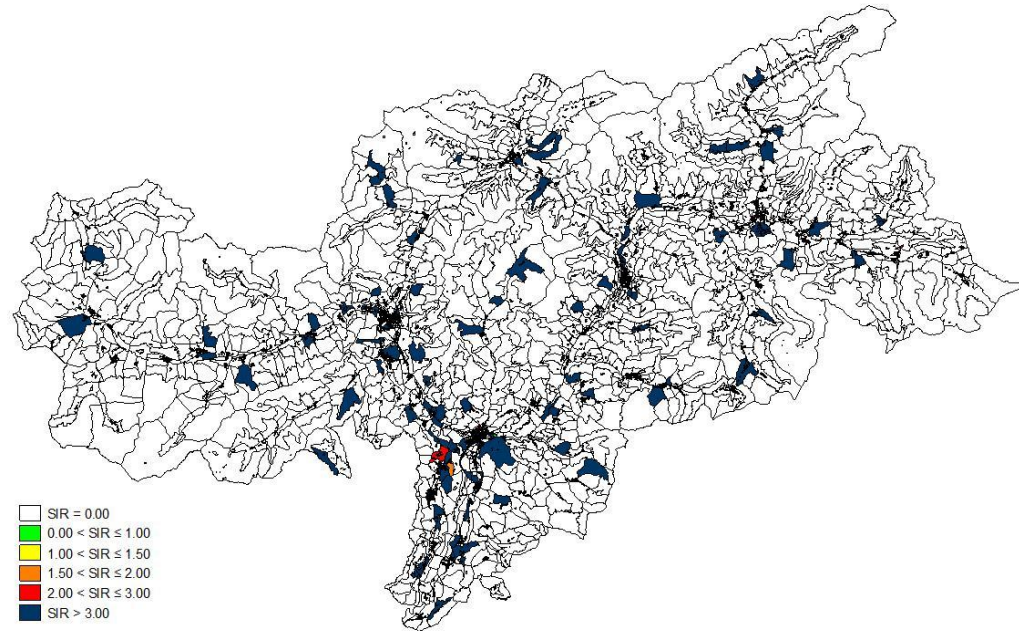
ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR

Observed SIRs distribution



World Segi's Standard

Indirect standardization



ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR

SARAR models results

Results of the SARAR model (Standardization: World)

	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

Results of the SARAR model (Indirect standardization)

	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

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

No socio-economic effect

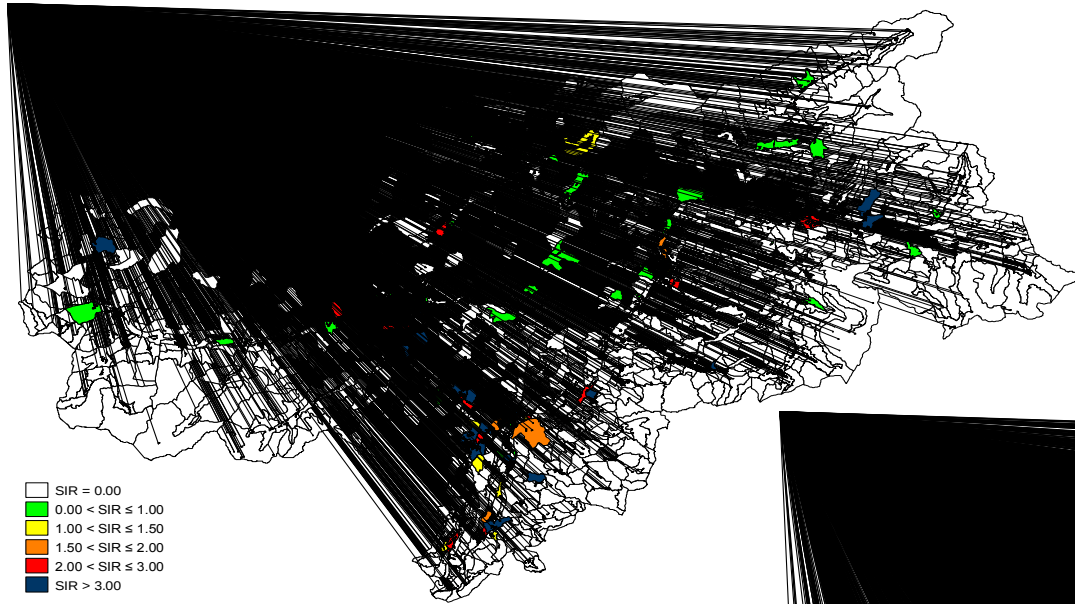
Potential cluster presence

No statistical significant differences between the two standardization. Clusters will be computed only for World standard estimates.



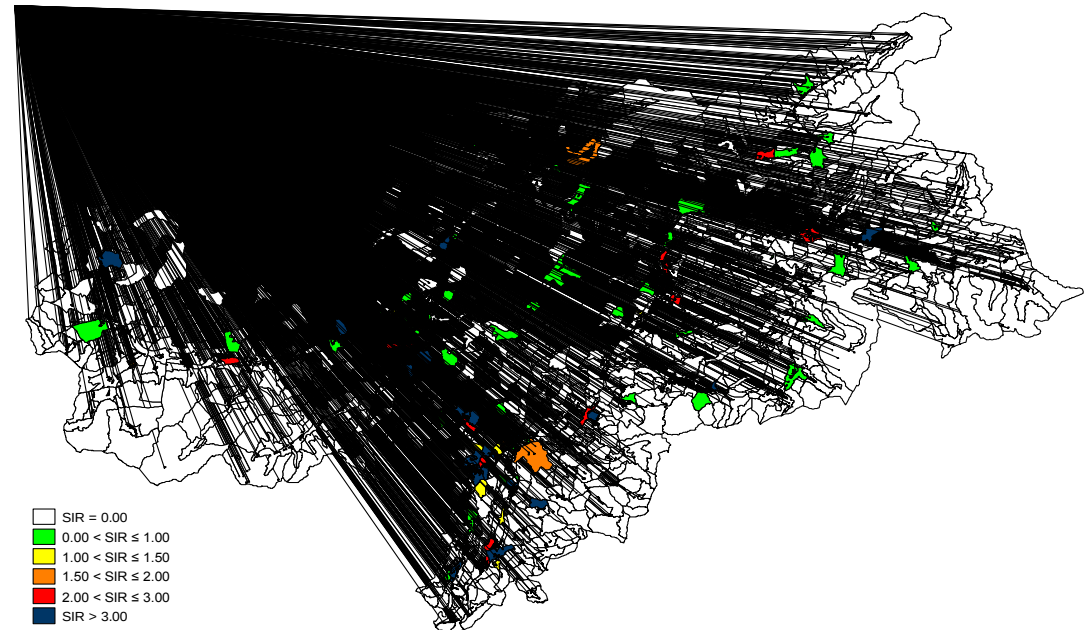
ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR

Estimated SIRs distribution (SARAR)



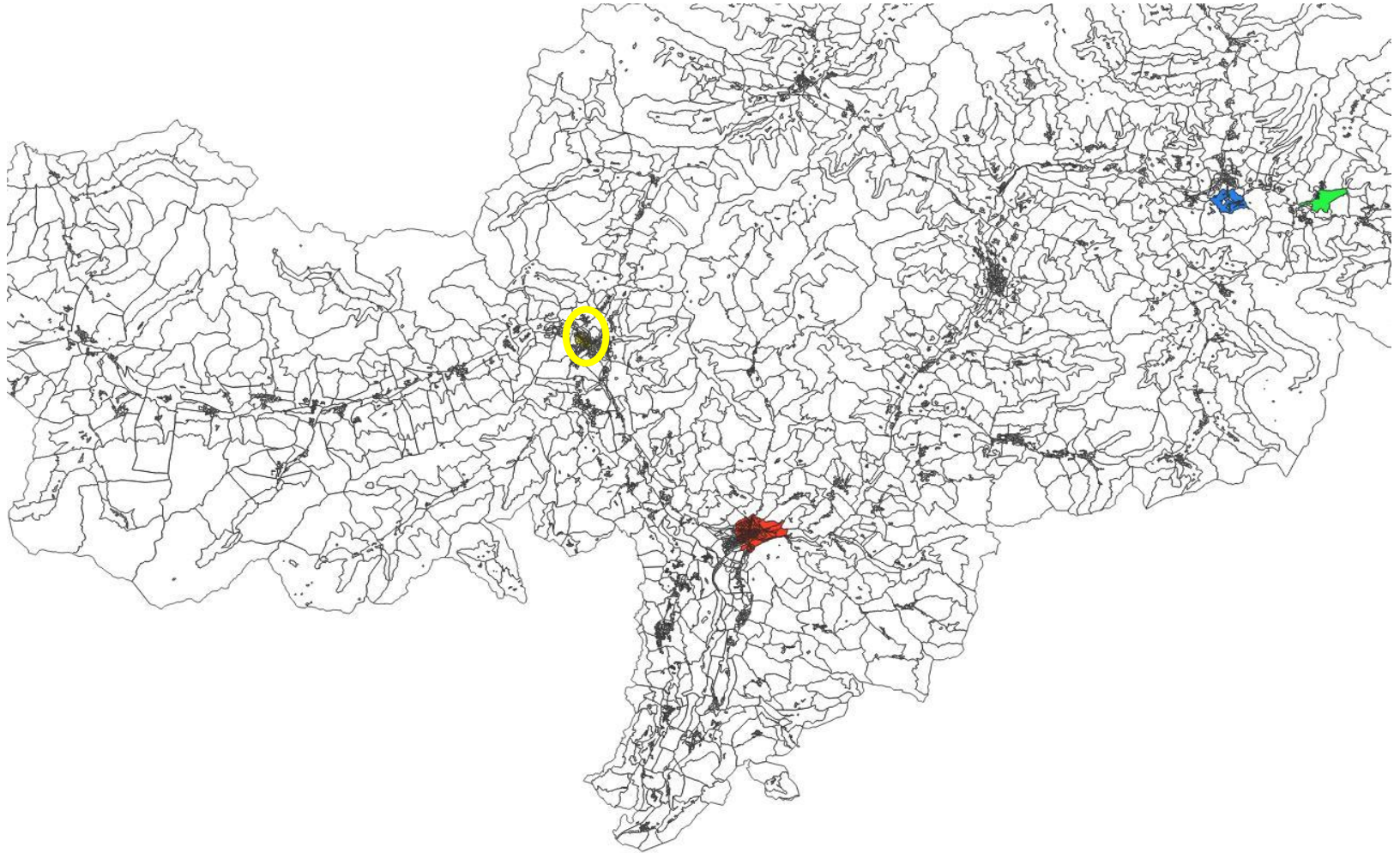
World Segi's Standard

Indirect standardization



ALREADY OBTAINED RESULTS - EXAMPLES ON ALTO ADIGE CR

SatScan clusters



Reports on spatial analysis for each Cancer Registry

Files to download

- A synthesis of main results for each CR is available at:
http://www.wasabysite.it/material/CR_Summary_data.xlsx
- The file reports the following information and will be periodically updated:

COUNTRY	CLUSTER (I.S.)
CANCER REGISTRY	CLUSTER CASES (I.S.)
EDI/SES index availability	WATERBASE POINTS
TOTAL CASES	E-PRTR POINTS
GEOGRAPHIC VARIABLES	Association with pollution sources (W.S.)
Correlation with EDI (W.S.)	Association with pollution sources (I.S.)
CLUSTER (W.S.)	POPULATION
CLUSTER CASES (W.S.)	0-49 FEMALE CENSUS TRACK POPULATION
Association with EDI (I.S.)	0-49 FEMALE TOTAL POPULATION

- All draft reports with analysis already performed are at:
http://www.wasabysite.it/material/CR_Reports.rar

Synthesis of the analyses for each CR

COUNTRY	CANCER REGISTRY	Association with EDI (W.S.)	CLUSTER (W.S.)	CLUSTER CASES (W.S.)	Association with EDI (I.S.)	CLUSTER (I.S.)	CLUSTER CASES (I.S.)	Association with pollution sources (W.S.)
ES	Basque Country	No	Yes. 5	1472	No	Yes. 10	1292	Yes
ES	Castellon-Valencia	No	On-going	On-going	On-going	On-going	On-going	On-going
ES	Girona	No	No	0	No	No	0	No
ES	Granada	No	Yes. 2	386	No	Yes. 1	341	Yes
ES	Murcia	Yes	Yes. 1	724	Yes	Yes. 1	1258	Yes
ES	Navarra	No	No	0	No	Yes. 1	588	No
IT	Alto Adige	No	Yes. 4	160	No	Yes. 4	160	Yes
IT	Asl Napoli 3	No	Yes. 25	252	On-going	On-going	On-going	On-going
IT	Palermo	On-going	On-going	On-going	On-going	On-going	On-going	On-going
IT	Parma	No	No	0	No	No	No	No
IT	Ragusa	Yes	Yes. 7	101	No	No	No	No
IT	Siracusa	No	Yes. 8	90	On-going	On-going	On-going	Yes
IT	Trapani	No	Yes. 10	126	On-going	On-going	On-going	Yes
IT	Trento	Yes	Yes. 3	562	On-going	On-going	On-going	Yes
IT	Umbria	Yes	Yes. 11	611	On-going	On-going	On-going	Yes
IT	Varese	No	Yes. 12	414	On-going	On-going	On-going	Yes
LT	Lithuania	On-going	On-going	On-going	On-going	On-going	On-going	On-going
PT	Central Portugal	No	No	0	No	Yes. 3	2163	No
PT	Northern Portugal	No	Yes. 3	2776	No	Yes. 1	2903	Yes
Legenda								
W.S.	World Standard	ES	Spain					
I.S.	Indirect Standardization	IT	Italy					
		LT	Lithuania					
		PT	Portugal					

Analyses and conclusion: forecast timeline

- SARAR and SATSCAN direct and indirect: end of January 2021, apart for France and Northern Ireland (due to delay in data sending).
- SARAR and SATSCAN direct and indirect for France and Northern Ireland: end of March 2021.
- CARBayes analyses: April 2021, depending on the solving of the zero-inflated models issues.