



WASABY

Estimation of incidence of breast cancer in young women with SARAR models. First results of the WASABY Project

Lillini Roberto¹, Bertoldi Martina¹, Žagar Tina², Zadnik Vesna², Launoy Guy³, Guillaume Elodie³, Launay Ludivine³, Contiero Paolo¹, Borgini Alessandro¹, Ritchie David⁴, Katalinic Alexander⁵, Pritzkeleit Ron⁵, Tittarelli Andrea¹, Modonesi Carlo¹, Tagliabue Giovanna¹, Amati Camilla¹, Sant Milena¹, Baili Paolo¹, the WASABY Working Group

1. *Fondazione IRCCS "Istituto Nazionale Tumori", Milan, Italy*
2. *Slovenian Cancer Registry, Institute of Oncology Ljubljana, Slovenia*
3. *UIO86 INSERM UCN, University of Caen, France*
4. *Association of European Cancer Leagues, Brussels, Belgium*
5. *Schleswig-Holstein Cancer Registry, Institute Cancer Epidemiology, University of Lübeck, Germany*

IN SYNTHESIS

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



AIMS

- To present the first results on breast cancer incidence risks estimates by Census Tracts (CTs), using cases made available by some of the participating Cancer Registries (CRs).
- To stimulate the discussion about the more affordable methods for spatial analysis with respect to the peculiarities of the WASABY.
- To show the relevance of such methods for territorial surveillance by Cancer Registries and other health institutions.

MATERIALS (1)

- 18 CRs have already sent their data to WP4 for the analysis, regarding 0-49 years female breast cancer incident cases.
- 23 CRs will develop the analysis on their own, according to the methodology and techniques that will be proposed by WP4 and WP6, after discussing the most reliable method for every specific situation.
- Data from 4 CRs are still pending for different reasons.

MATERIALS (2)

- To date, a total of 78717 incident cases from the CRs were considered for the analysis.
- 222 (0.3%) of them were excluded for incomplete and unrecoverable geo-coding information:
- The cases are distributed along the 1996-2016 period.
- The table reports the European Deprivation Index (EDI) and local indices situation:

Nations	EDI 2001	EDI 2011	Local indices
France	X	X	
Germany			X
Italy	X	In development	X
Lithuania		In development	
Poland		In development	X
Portugal	X	X	
Slovenia	X	X	
Spain	X	In development	
Northern Ireland, UK			X

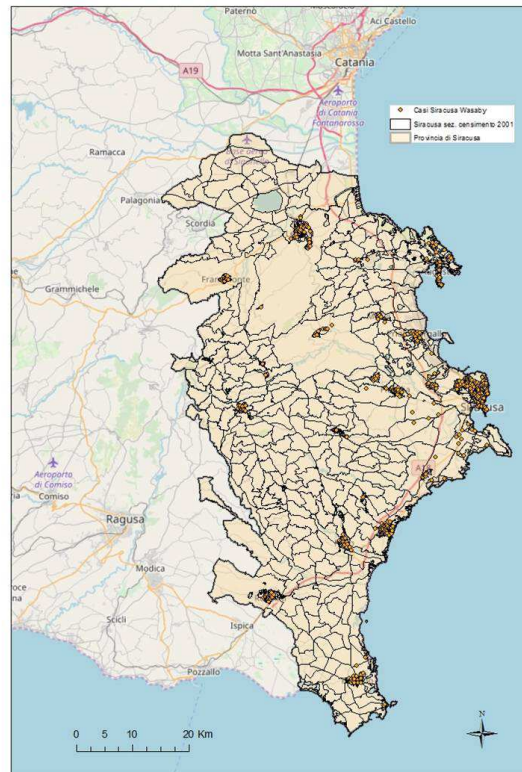
METHODS – GEO-CODING (1)

For some CRs, the geo-coding of the cases was performed by WP4 team, according to the following procedure.

- Recovery of complete address at street number for all cases.
- Every address was geo-referenced through a free service on the Web that "captures" information from Google Maps and returns it in Excel format (Excel Geocoding Tool).
- The spatial coordinates of all addresses (latitude and longitude) were therefore projected on geographical maps, using both the ArcGis software and the Qgis freeware software.
- The accuracy of the result was verified through comparisons and overlaps with maps of administrative boundaries, with individual punctual Google Maps queries, even in streetview mode, and with toponymic verification.

METHODS – GEO-CODING (2)

The shapefile of the distribution of the cases was superimposed on the shapefile of the 2001 CTs and, using the spatial join function, the corresponding CT was associated to the cases, in order to be able to attribute to each subject a quantitative index of status socio-economic, available at the CT level.



	A	B	C	R
1	ID_CASO	Latitude	Longitude	SEZ2001
2	1	37,069732	15,286629	890170000557
3	2	37,226070	14,873130	890100000016
4	3	37,226511	14,873602	890100000016
5	4	37,227050	14,875980	890100000010
6	5	37,224100	14,876410	890100000016
7	6	37,223160	14,877440	890100000015
8	7	37,229390	14,877960	890100000006
9	8	37,230470	14,878150	890100000003
10	9	37,230040	14,878790	890100000003
11	10	37,223340	14,879250	890100000014
12	11	37,224990	14,879860	890100000014
13	12	37,226840	14,879890	890100000011
14	13	37,225020	14,880120	890100000014
15	14	37,224040	14,880200	890100000014
16	15	37,231605	14,880459	890100000001
17	16	37,231090	14,881850	890100000001

METHODS – SPATIAL ANALYSIS (1)

- The incidence estimates were computed from the observed SIRs (reference for the expected cases: the CR area, standardization on standard world population) through spatial autoregression models with effect of auto-regressive disorder and exogenous covariates (SARAR models).
- The European Deprivation Index (EDI) 2001 or 2011 in its quantitative version is the considered covariate, calculated at Census Tract level.
- The analyses were conducted separately for each CR via the Stata 14.0 SPREG module.

METHODS – SPATIAL ANALYSIS (2)

Characteristics of the SARAR model:

$$y = \lambda W y + X \beta + u \quad (1)$$

$$u = \rho M u + \varepsilon \quad (2)$$

Where:

- y is an $n \times 1$ vector of observations on the dependent variable;
- W and M are $n \times n$ spatial-weighting matrices (with 0 diagonal elements);
- $W y$ and $M u$ are $n \times 1$ vectors typically referred to as spatial lags, and λ and ρ are the corresponding scalar parameters typically referred to as SAR parameters;
- X is an $n \times k$ matrix of observations on k right-hand-side exogenous variables (where some of the variables may be spatial lags of exogenous variables), and β is the corresponding $k \times 1$ parameter vector;
- ε is an $n \times 1$ vector of innovations.

RESULTS – Varese province CR

Spatial autoregressive model
(GS2SLS estimates)

Number of obs = 5194

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

sir049_wor					

sir049_wor					
zediweighted_01	-.1704525	.0742063	-2.30	0.022	-.3158942 - .0250109
_cons	1.301344	.2869289	4.54	0.000	.7389736 1.863714

lambda					
_cons	6.301074	1.297145	4.86	0.000	3.758716 8.843432

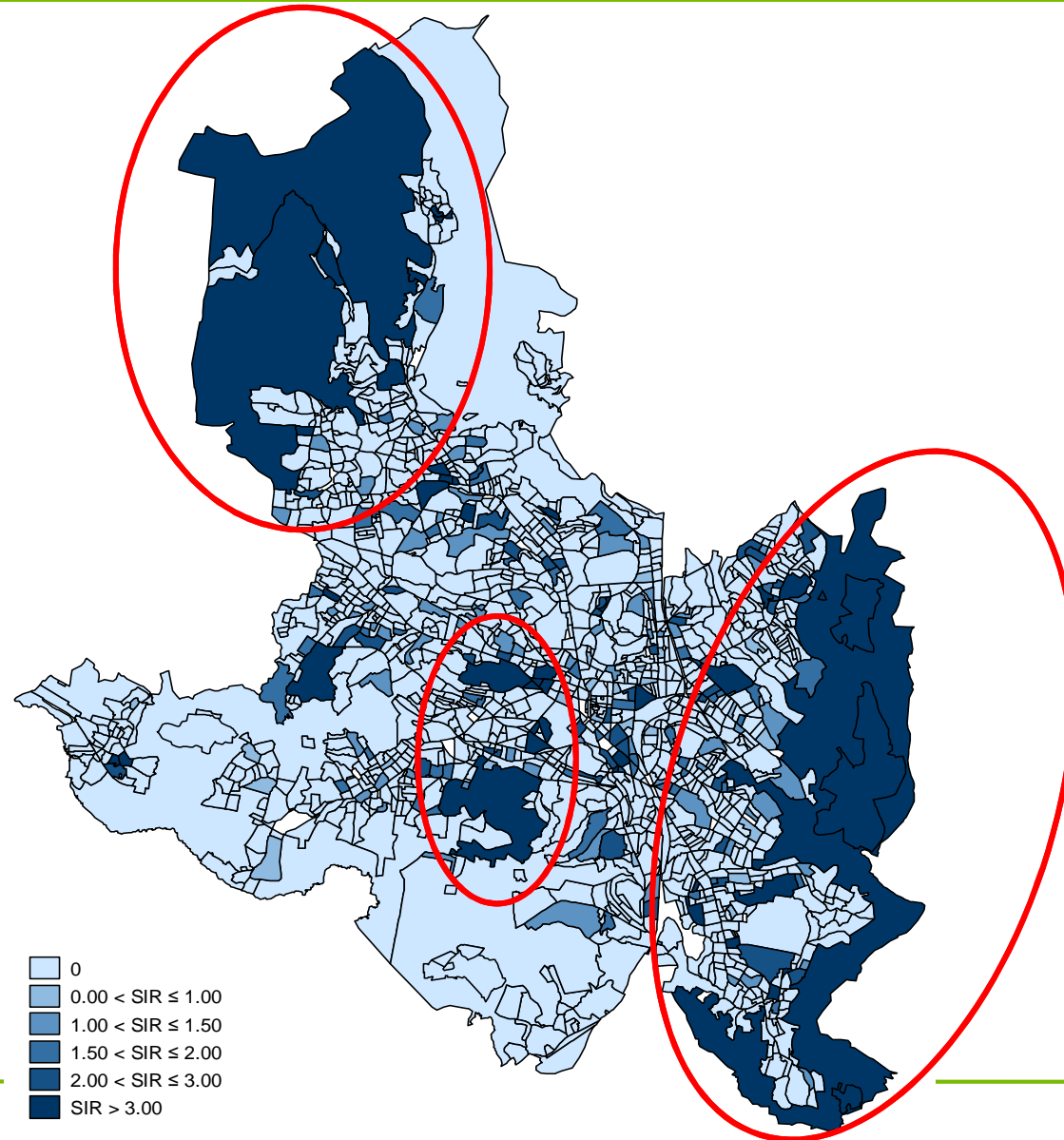
rho					
_cons	-12.38115	1.510662	-8.20	0.000	-15.34199 -9.420304

A statistically significant effect was observed for the covariate.

Lambda and rho remarked a statistically significant spatial clustering effect.

RESULTS – Varese province CR

Zoom on Varese municipality



RESULTS – Granada province CR

Spatial autoregressive model
(GS2SLS estimates)

Number of obs = 606

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

sir049f_wor						
EDI2001	.0272248	.0139049	1.96	0.050	-.0000284	.054478
_cons	-.3626361	.2445907	-1.48	0.138	-.842025	.1167527

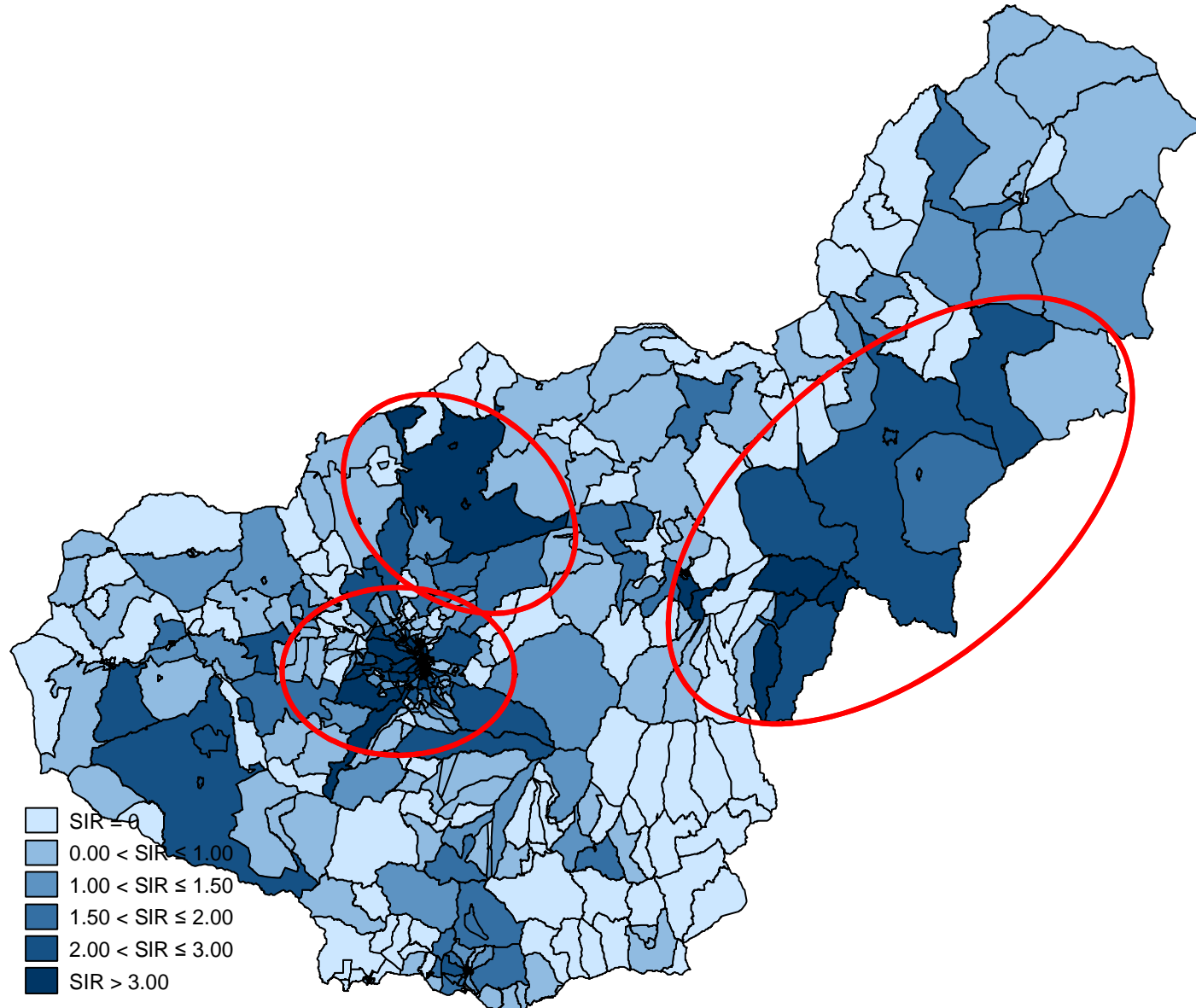
lambda						
_cons	3.705674	.5489546	6.75	0.000	2.629743	4.781606

rho						
_cons	-2.933967	.2437493	-12.04	0.000	-3.411707	-2.456227

No statistically significant effect has been observed for the covariate.

Lambda and rho remark a statistically significant spatial clustering effect.

RESULTS – Granada province CR



DISCUSSION AND CONCLUSION

- These first results are intended as a basis for discussion on the validity of SARAR models, in order to select the best model for the study, starting from the second half of 2019.
- The map of the estimated excess risk, excluding or including the socio-economic covariate with respect to its statistical significance, allows to hypothesize other risk factors spread over the territory.
- Thus, the so-found spatial clusters can be suggestive of areas to be investigated for environmental exposition.

OPEN QUESTIONS AND NEXT STEPS

Questions:

- what are the best SIR's cutoffs for considering “real” problems and avoiding false alarms (too low number of cases), in order to avoid wasting resources?
- what are the reliable methodologies and techniques for associating the clusters of risk excess to pollutants exposure?

Next steps:

- collection of data from the last participating CRs;
- procedures for supporting CRs in geo-coding and preparation of cancer maps and/or EDI;
- completion of the EDI computing in the participating countries;
- choice of the best model for analysis, after the methodological discussion (WP6), and data analysis;
- after the WP7 literature review of the pollutants interesting for breast cancer (finished in March and available at the WASABY website), choice and collection of the available geo-coded pollutant database for the pilot environmental study.

“Depending on when in her lifespan a woman integrates risk-reduction behaviours, the majority of breast cancer can be prevented”

Colditz G, CA Cancer J Clin, 2014

Visit our website to learn more www.wasabysite.it



Co-funded by
the Health Programme
of the European Union

EU funding disclaimer: This project has received
funding from the 3rd European Union Health
Programme 201-2020 under Grant Agreement PP-2-5-
2016 (# 769767)