



Discussion

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

Milan, 19.02.2018

WASABY's strategy for each CR

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

Protocol

The broad INT protocol will include three main options

- Send data to Milan/Slovenia:
 - Individual level or
 - Aggregated data (number of BC cases) by S.U. level
- Do not send data to Milan/Slovenia:
 - CR will perform in-house analysis

Incidence data

- Calendar years: 10-years period
- CRs do not need to have the same calendar years
- Only first breast cancer case
- Add Subtype (not mandatory)
- Stage, age, invasive/insitu, screening: see later

Background activity with each CR

- Changes in shape files in various calendar years
- Geocode reference
- Ethical committee problems with confidentiality
- Intrinsic interpretation of the S.U.
- Maps details to be shared online, in publications....
- Calendar year availability
- Data storage

Calendar years

	SES	Incidence	Problems
Option 1	2001	1996-2005	Geocode < 2001
Option 2	2001	2001-2010	
Option 3	2011	2006-2015	Incidence > 2013
Option 4	2001	2001-2005	
	2011	2006-2010	

We have a lot of work....

- To reach the WASABY objective we have a lot of work and discussion to be performed in each CR
- We are worried about the feasibility of reaching our target indicators and our main aims
- At this first stage I would like to stress only on:
 - Invasive cancers
 - Not stage
 - Age < 50 (and include all ages not mandatory)

Appendix

1. Synthesis - Data collection protocol: proposal

FILE WITH BREAST CANCER CASES

All primary invasive female breast cancer (ICD9 174*, ICD10 C50*), selected from cancer registries data during a specific period (ex: 2000 to 2009). Cancer registration criteria follow IARC rules.

Variable name	Description	Data type
CR	Cancer Registry name	Alphanumeric variable
PATIENT_ID	Patient identification code assigned by Cancer Registry. It is necessary not only to identify a single subject, but also to retrieve all necessary health and administrative data Note: synchronous breast cancer cases must be counted once	Numeric/Alphanumeric variable
DATE OF BIRTH	Date of birth of the patient	DD/MM/YYYY
DATE OF DIAGNOSIS	Incidence date based on histological or cytological confirmation of the malignancy, if available	DD/MM/YYYY
AGE	Age at diagnosis	Numeric variable
ICD_9	Complete ICD-9 code of incident case	Alphanumeric variable
ICD_10	Complete ICD-10 code of incident case	Alphanumeric variable
ICDO3_M	ICDO3 morphology code of incident case	Alphanumeric variable
STAGE	Stage at diagnosis according to TNM stage grouping I II III IV unknown	Alphanumeric variable

2. Methods and software of analysis

Example for discussing issues about spatial analysis

Area: Varese CR.

Incidence: 1586 cases of primary breast tumours (ICD10=C50), aged 0-49 years, from the period 2000-2009 (follow up till to 31/12/2014).

Covariates: stage at diagnosis, ICDO3-M classification, EDI 2001 z-score at census tract level (CT).

Software: Stata 14 modules shp2dta, spmat, spreg (with post-estimation) for Bayesian spatial analysis.

2. Methods and software of analysis

Example

In the analysis only 1509 cases were considered, without missing data for all covariates (EDI 2001 included)

Varese shapefile was transformed in Stata format by shp2dta.

The spatial-weighting proximity matrix was computed by spmat:

Summary of spatial-weighting object varemat:

Matrix	Description
Dimensions	1002 x 1002
Stored as	1002 x 1002
Links	
total	2118
min	0
mean	2.113772
max	12

2. Methods and software of analysis

Example

The model was computed by spreg:

Spatial autoregressive model
(GS2SLS estimates)

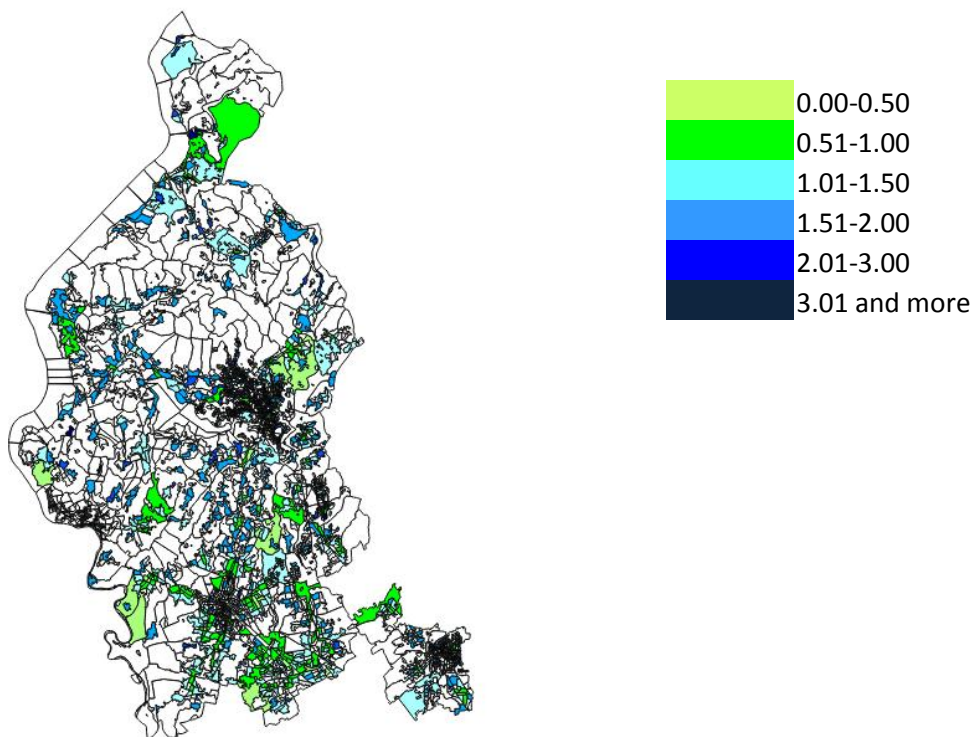
Number of obs = 1002

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
sirf049						
zeditestweighted	-.2806605	.1922421	-1.46	0.144	-.6574482	.0961271
_cons	3.125324	.6351175	4.92	0.000	1.880517	4.370132
-----+-----						
lambda						
_cons	-.7720799	1.574308	-0.49	0.624	-3.857668	2.313508
-----+-----						
rho						
_cons	-.4379243	1.056824	-0.41	0.679	-2.509261	1.633413

2. Methods and software of analysis

Example

Map of the estimated SIR:



2. Methods and software of analysis

Open questions

Our very simple example is only an opportunity to discuss about:

- What methods? Here we have used a GAM approach, a form of non-parametric or semi-parametric regression offering the possibility to analyse contextual data while adjusting for covariates and taking into account spatial autocorrelation.

2. Methods and software of analysis

Open questions

Our very simple example is only an opportunity to discuss about:

- What software package? Stata 14 doesn't allow, to our knowledge, to use BYM models, therefore R packages (e.g., CARBayes) or other software solution should be considered.

2. Methods and software of analysis

Open questions

Our very simple example is only an opportunity to discuss about:

- Other aspects conditioning the models.

3. Nations and Cancer Registries: Belgium

#of CRs	1 (national)
Geo-code level	Municipality
Geo-referentiation linkage by	Cancer Registry
Reference population available at	Cancer Registry by calendar year
Maps available at	Cancer Registry
SES	EDI (in development?)

3. Nations and Cancer Registries: Germany

#of CRs	2
Geo-code level	Urban District (Bremen)
Geo-referentiation linkage by	Cancer Registry
Reference population available at	Cancer Registry by calendar year
Maps available at	Cancer Registry
SES	Local index (until 2009)

3. Nations and Cancer Registries: Italy

#of CRs

7 + 1 pending (Parma)

Geo-code level

Census Block (some Municipality)

**Geo-referentiation linkage by
Reference population available at**

Cancer Registry
Cancer Registry by calendar year

**Maps available at
SES**

Cancer Registry/ISTAT
EDI 2001+national & local

3. Nations and Cancer Registries: Lithuania

#of CRs	1 pending (national)
Geo-code level	Municipality
Geo-referentiation linkage by	Don't know
Reference population available at	Cancer Registry by calendar year
Maps available at	Don't know
SES	Don't know

3. Nations and Cancer Registries: Poland

#of CRs	3 +1 pending (Krakow)
Geo-code level	Municipality
Geo-referentiation linkage by	Various (CR or external paid resources)
Reference population available at	Cancer Registry by calendar year
Maps available at	Various (CR or external free resources)
SES	Local index

3. Nations and Cancer Registries: Portugal

#of CRs	2
Geo-code level	Various (Municipality, Parish)
Geo-referentiation linkage by	Cancer Registry
Reference population available at	Cancer Registry by calendar year
Maps available at	Cancer Registry/INE
SES	EDI 2001

3. Nations and Cancer Registries: Slovenia

#of CRs	1 (national)
Geo-code level	x, y coordinates
Geo-referentiation linkage by	Cancer Registry
Reference population available at	Cancer Registry by calendar year
Maps available at	Cancer Registry
SES	EDI (in development)

3. Nations and Cancer Registries: Spain

#of CRs

4 + 1 pending (Murcia)

Geo-code level

Census Block (Municipality for Granada)

**Geo-referentiation linkage by
Reference population available at**

External paid resources
Cancer Registry by
calendar year (some
sporadic)

**Maps available at
SES**

Cancer Registry
EDI 2001

3. Nations and Cancer Registries: UK

#of CRs	1 (Northern Ireland)
Geo-code level	Zip Code
Geo-referentiation linkage by	Cancer Registry
Reference population available at	Cancer Registry by calendar
Maps available at	Don't know
SES	Local index

3. Participating Cancer Registries

Open questions

- Are the participating CRs' nations enough (from the project we need at least 6 nations)?
However, would it be interesting to include new countries from a scientific point of view?
(e.g., FRANCE?)
- Should we try to extend Cancer Registry adherence, e.g., re-contacting the ones which never replied to the questionnaire circulated in 2017?

4. Preliminary survey

Open questions

- After consulting the copy of the questionnaire, do you think that further information should be asked (also considering the possible extension to other CRs)?

5. Data storage

Issues to be discussed

- Data will be centrally stored at FONDAZIONE IRCCS ISTITUTO NAZIONALE DEI TUMORI, MILAN.
- For a list of cancer registries data will be shared with the ONKOLOSKI INSTITUT, LJUBLJANA for the analysis.
- Data will be stored individually, but anonymously.
- 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)