LIFE12 ENV/IT/000834 MED HISS

The MED HISS Project

Mediterranean Health Interview Survey study:

Ecologic vs. Individual approach: Methodology and recommendations

“Ecologic”

- In Epidemiology:

Ecologic = aggregate

- Point of confusion when talking to people in other fields.
- Ecologic study:

  “a study in which the units of analyses are populations or groups of people, rather than individuals."

Study designs

- Künzli and Tager, EHP 1997.

		Exposure	
		Individual	**Ecologic**
Outcome	**Individual**	Individual	Semi-individual
	Ecologic	Aggregate	Ecologic

MEDHISS Italy

ESCAPE

MEDHISS Spain, Slovenia

Ecologic studies

Main disadvantages

- Ecological bias, ecological fallacy.
  - exposure-response relationships in group-level data are not the same than at the individual level.
- Ecological inference is theoretically biased when:
  - there is a non-linear relationship between the exposure and risk of outcome, and there is within-area variability in the exposure.
  - There is within-area confounding.
  - Unmeasured factors confound the baseline disease risk for groups or the effect of the risk factor under study.

Main disadvantages

- Any factor that is directly or indirectly related to the grouping process may act as confounder.
- Contextual effects arise when individual responses are influenced not only by their individual characteristics and behaviors, but by characteristics of other individuals in their area or of the area itself.
  - E.g. individual social class vs. average social class of municipality.

Main disadvantages

- Using ecological data alone it is difficult to distinguish between relative risks specific to individuals, and contextual effects.
- More difficult to state a priori what the effect of a confounder can be.
- In sum, direction of biases are difficult, if not impossible, to predict.

Esempio di diluizione del rischio negli studi ecologici

(Deprivation and mortality: A deprivation index suitable for geographical analysis of inequalities Article · Jul 1999 · Epidemiologia e prevenzione)

Tabella 3 (Table 3)

Analisi per indice di deprivazione a Torino (Analysis by deprivation categories in Turin)
Uomini 1991-1995 (Men 1991-1995)
RR per tutte le cause (RR for general mortality)
18-64 aa (ages 18-64)

	INDICE				
livello	*molto ricco*	*ricco*	*medio*	*deprivato*	*molto deprivato*
individuo	1	1.21	1.46	1.48	1.73
sezione	1	1.16	1.18	1.32	1.71
zona statistica	1	1.09	1.23	1.34	1.43
quartiere	1	1.09	1.15	1.25	1.28
circoscrizione	1	1.08	1.20	1.13	1.26

Nota: (Numerosità della popolazione nei cinque livelli)

1) Individuo: 920.000 abitanti al censimento 1991
2) sezione di censimento: 3657 sezioni, pop. media di 250 persone
3) zona statistica: 92 zone statistiche, pop. media di 10.000 persone
4) quartiere: 23 quartieri, pop. media di 40.000 persone
5) circoscrizione: 10 circoscrizioni, pop. media di 92.000 persone

Other difficulties encountered

- Very difficult to obtain data on relevant variables at group level (need information for all the spatial units).
- E.g. no statistics by municipality on:
  - Smoking
  - Diet
  - Physical activity
  - SES and income data limited.

Summary table from Künzli & Tager

Table 3. Limits on causal inference in ecologic studies that do not apply to the semi-individual and individual-level studies

Methodological concern	Specific situation in truly ecologic studies[a]
Confounding	Any variable related to grouping process may confound, not only risk factors for outcome
Regression and correlation coefficients	Weighted average of between-group and individual-level effects Uncontrollable distortion
Model specification	Source of undefined errors Individual level function cannot be assessed Limited number of (ecologic) modeling variables available Available ecogologic variables may be adjusted to different (unknown) standards
Misclassification	Deleterious effects on measures of associations Requires unavailable estimates of specificity/sensitivity of ecologic exposure measures Direction and size of bias unpredictable

Semi-individual studies

Semi-individual study

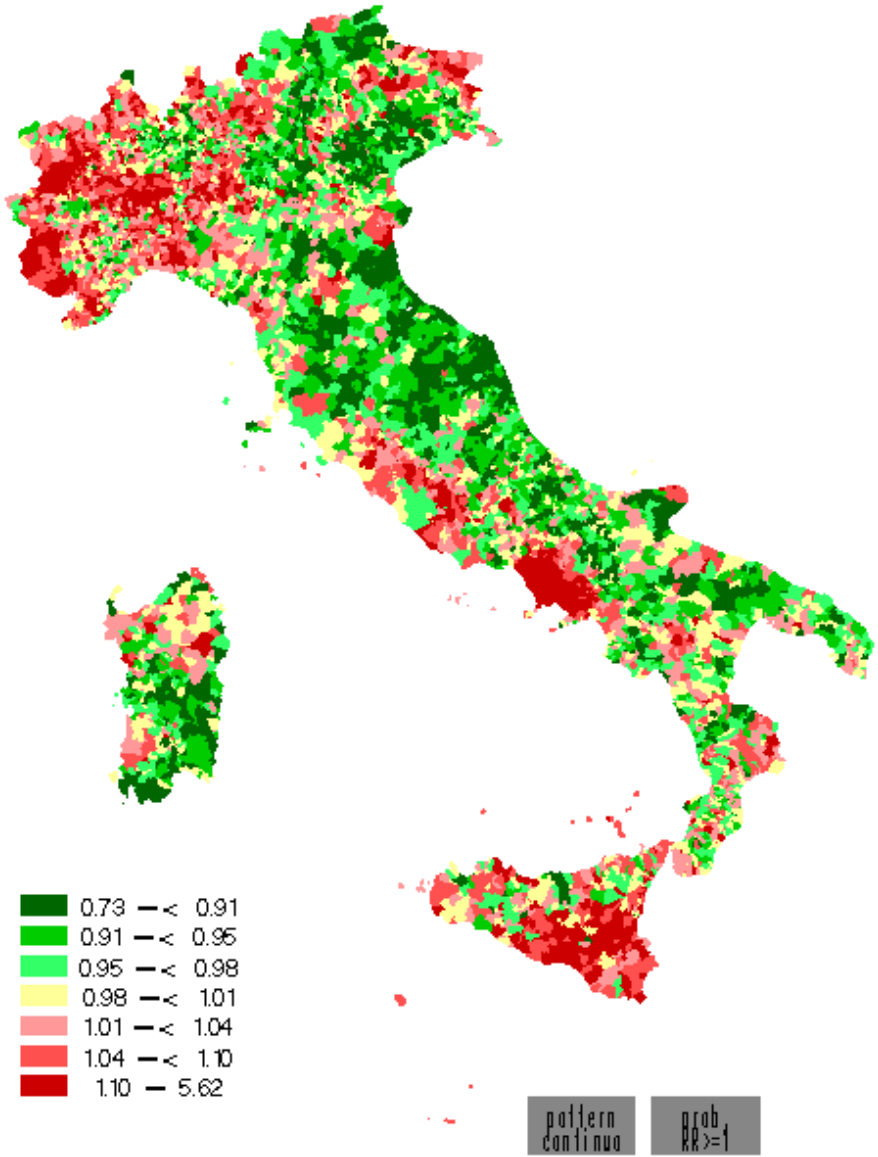
- Main advantage: control for individual level confounders.
- Major caveat: if only a small number of areas are included – difficult to control aggregate level confounding.
- Properties of semi-individual study very similar to individual study.
- Errors in assigning the exposure most likely to be of Berkson type – they do not produce bias in the results.

Results from Italy

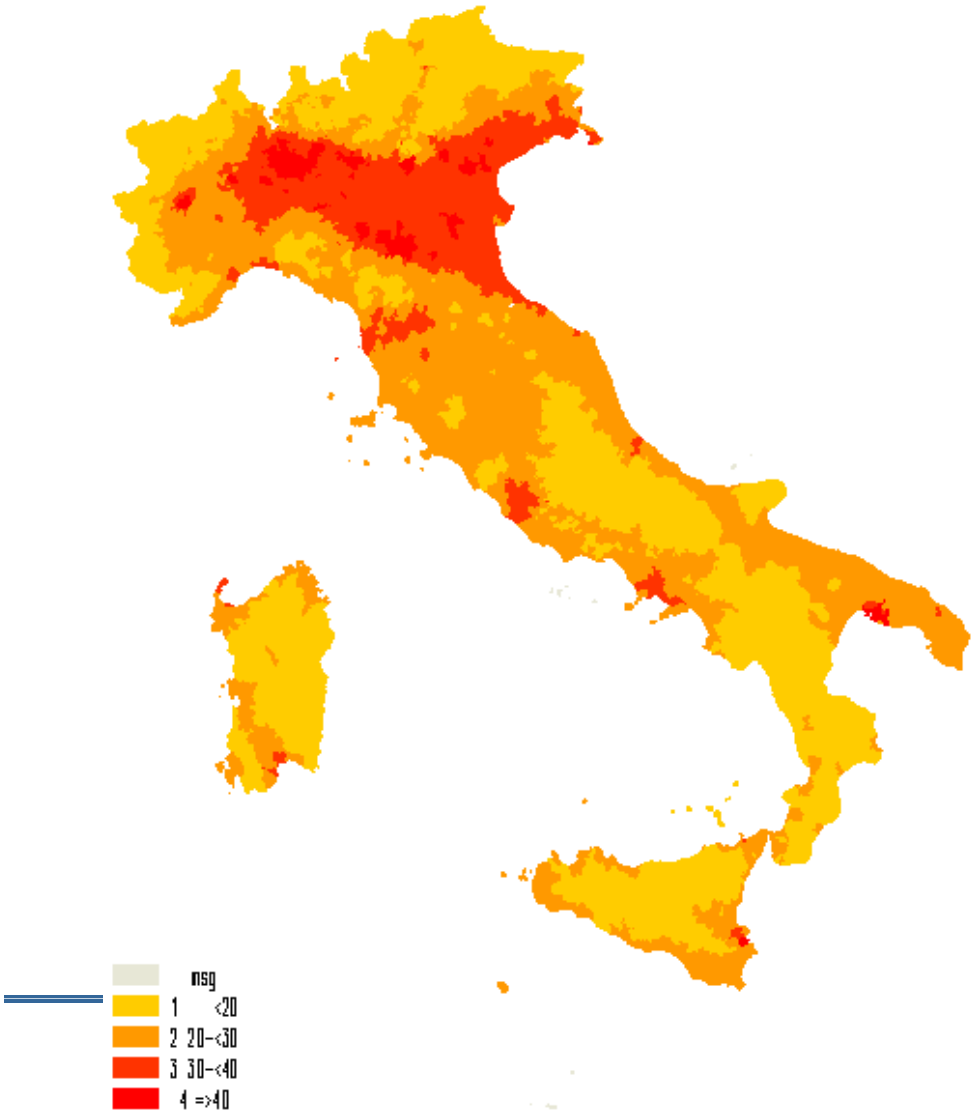
Health data

- In Italy we have the opportunity to compare the results of both approaches.
- Ecologic analysis, similar to analysis in Spain:
  - Italian mortality, 2000-2012, 35 years and over, both genders.
  - Adjusted for population, deprivation index and lung cancer mortality rates

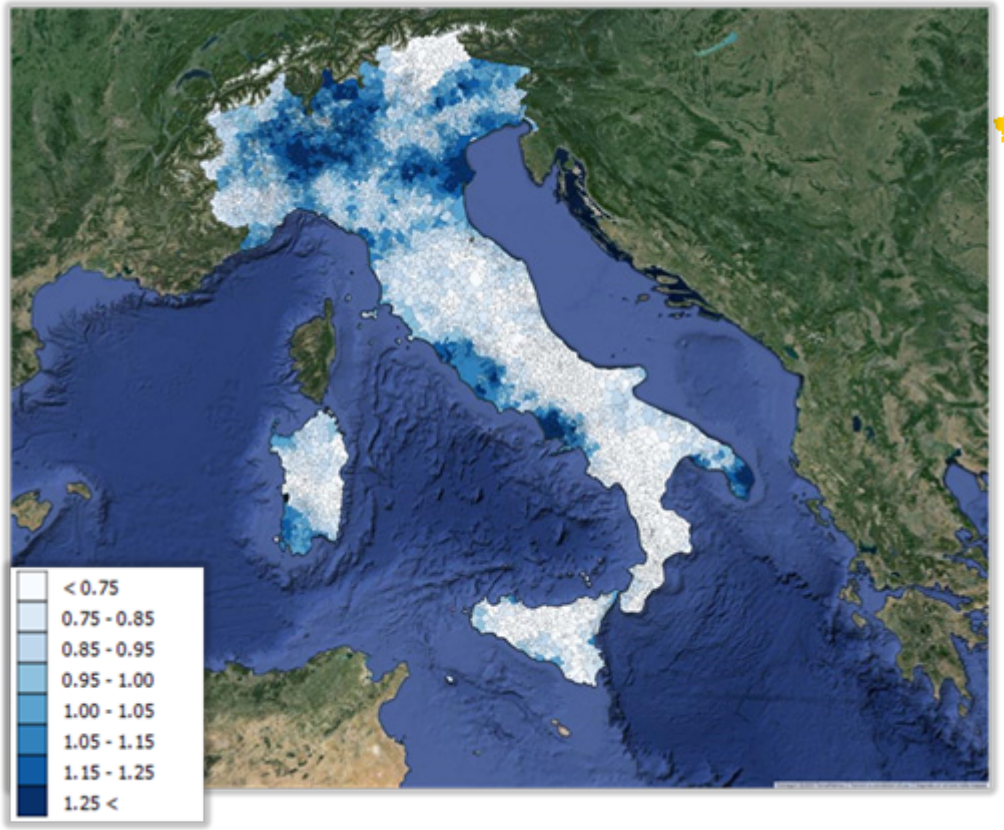
Mortality rates



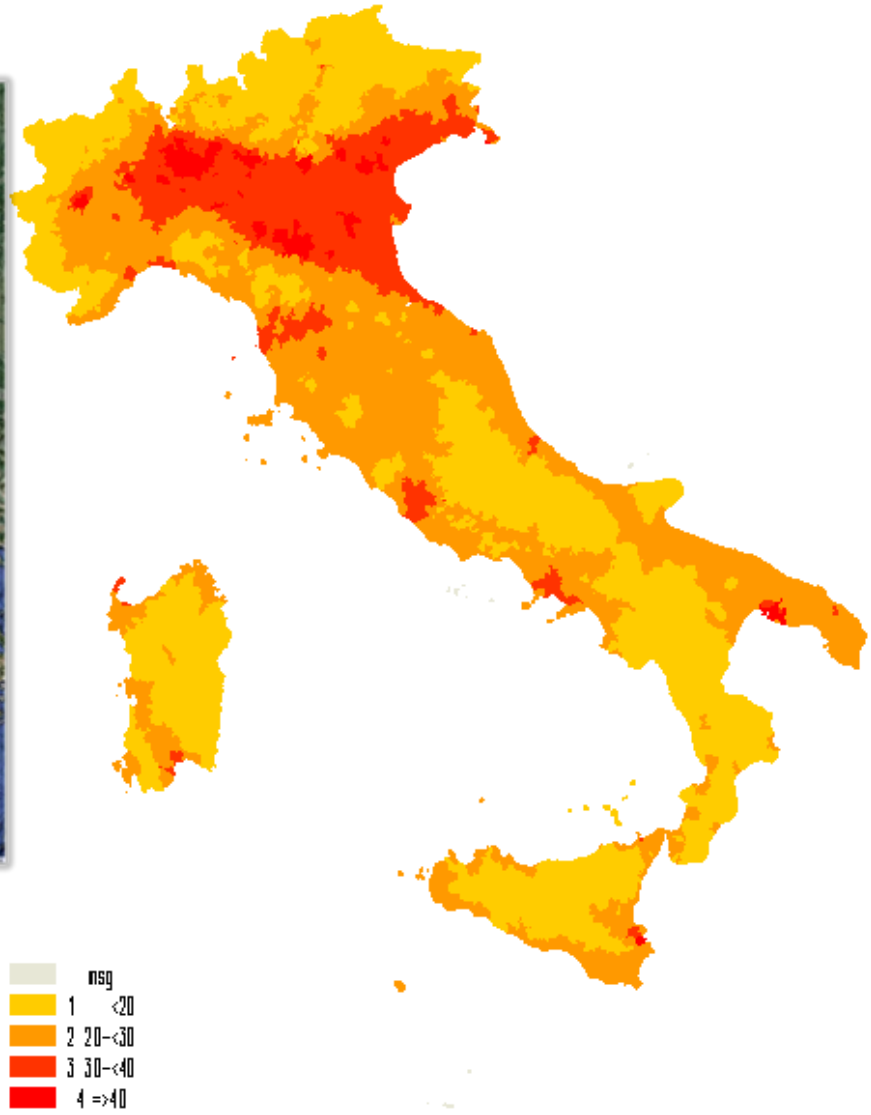
PM10



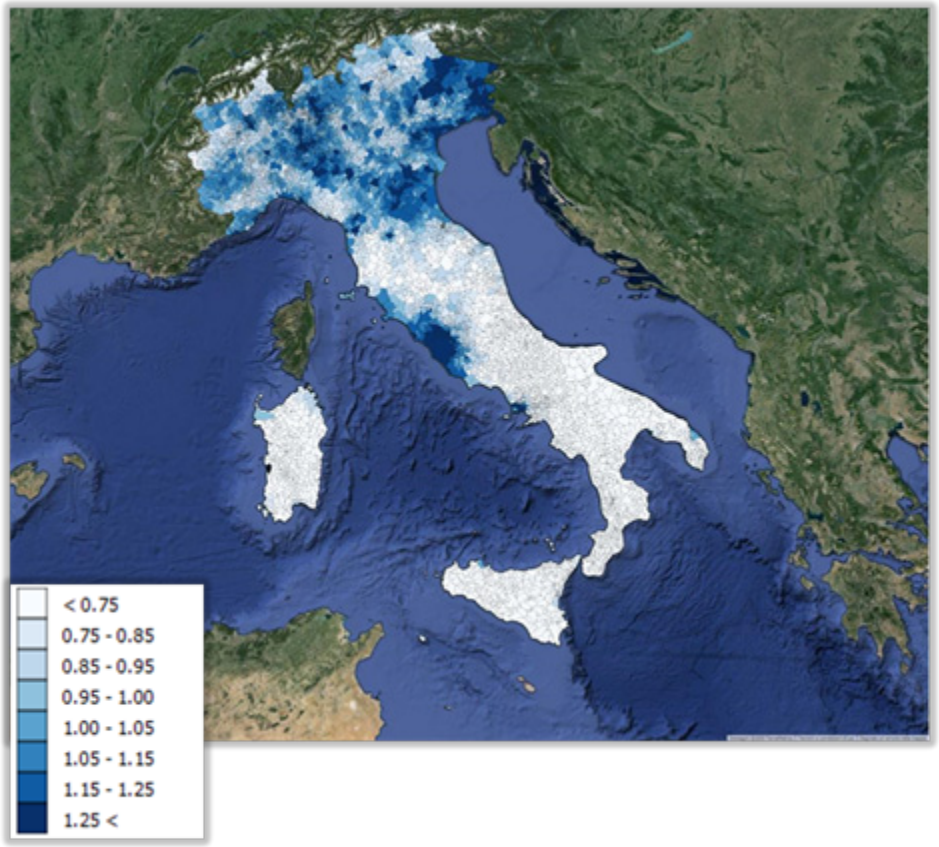
BMR lung cancer Males



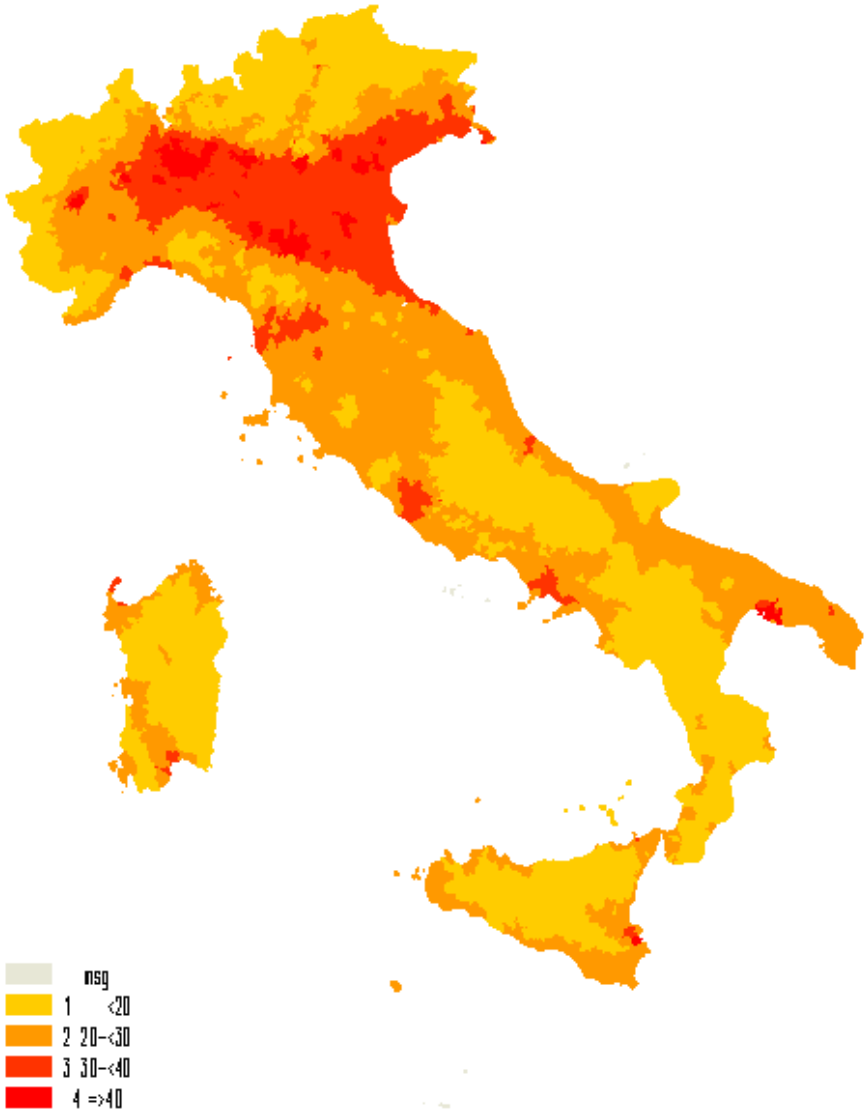
PM10



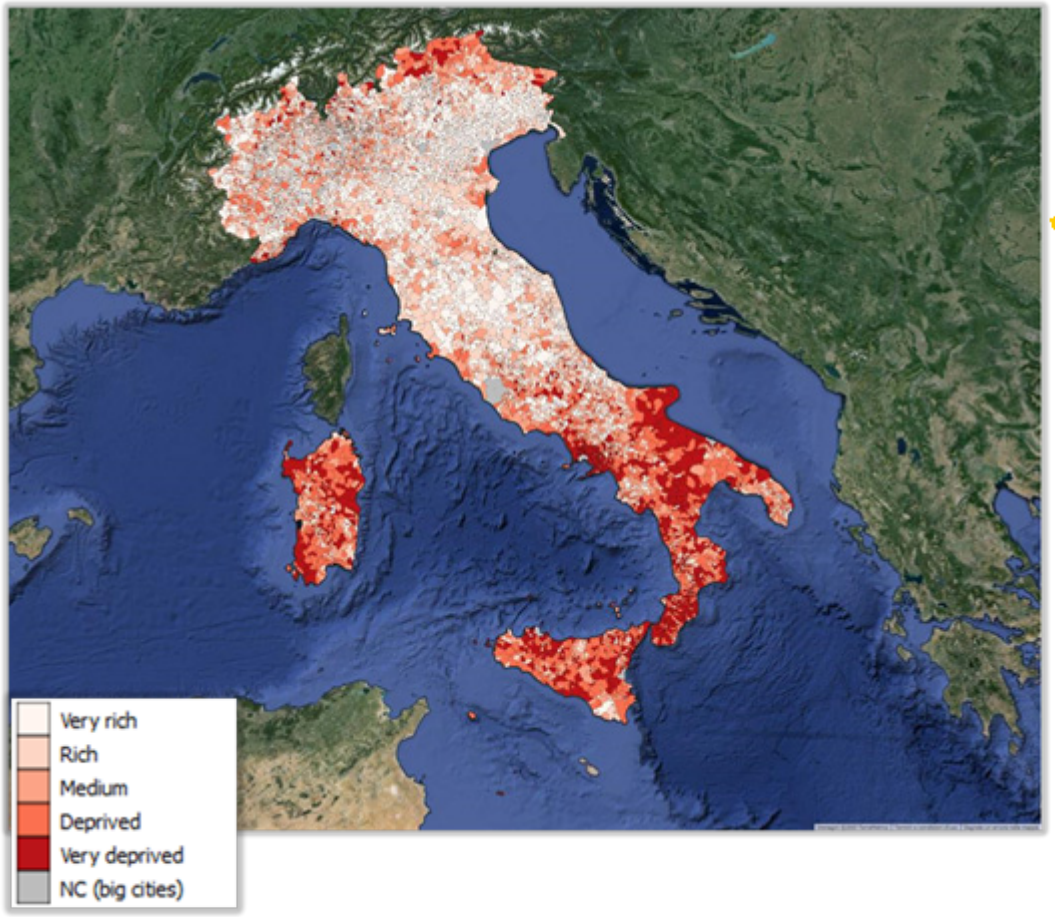
BMR lung cancer Females



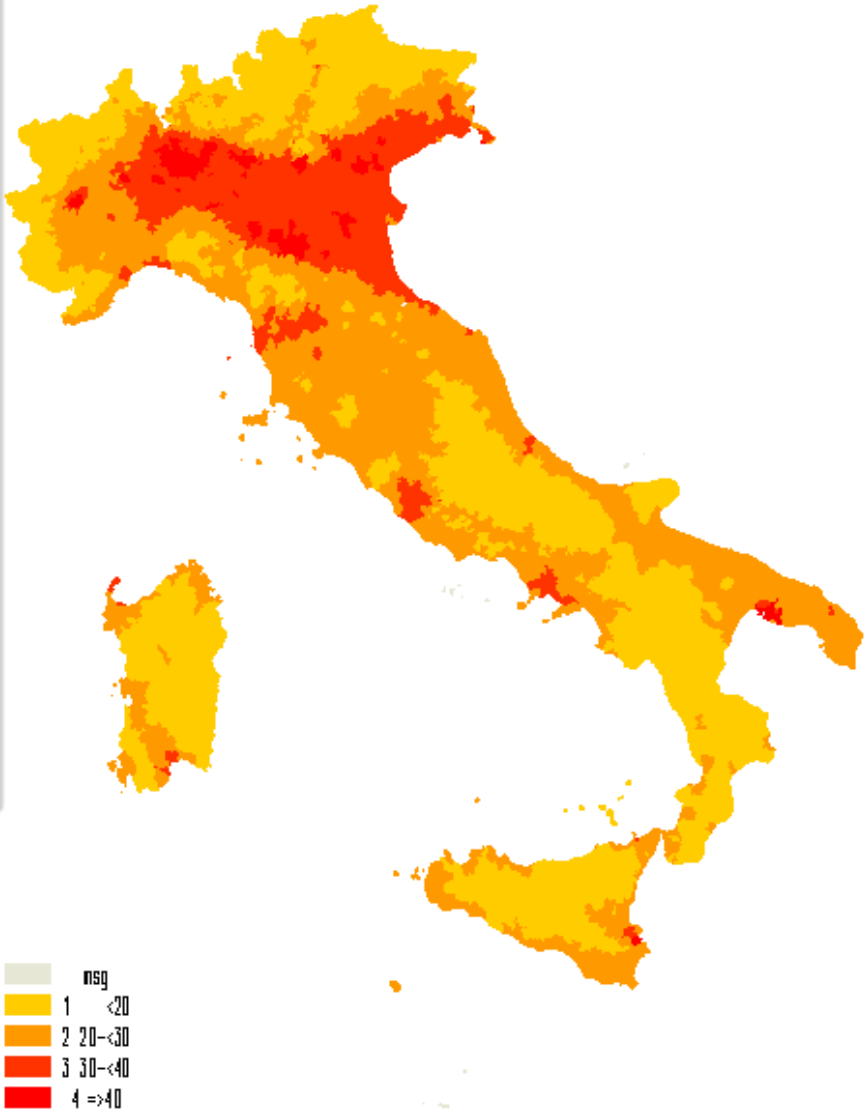
PM10



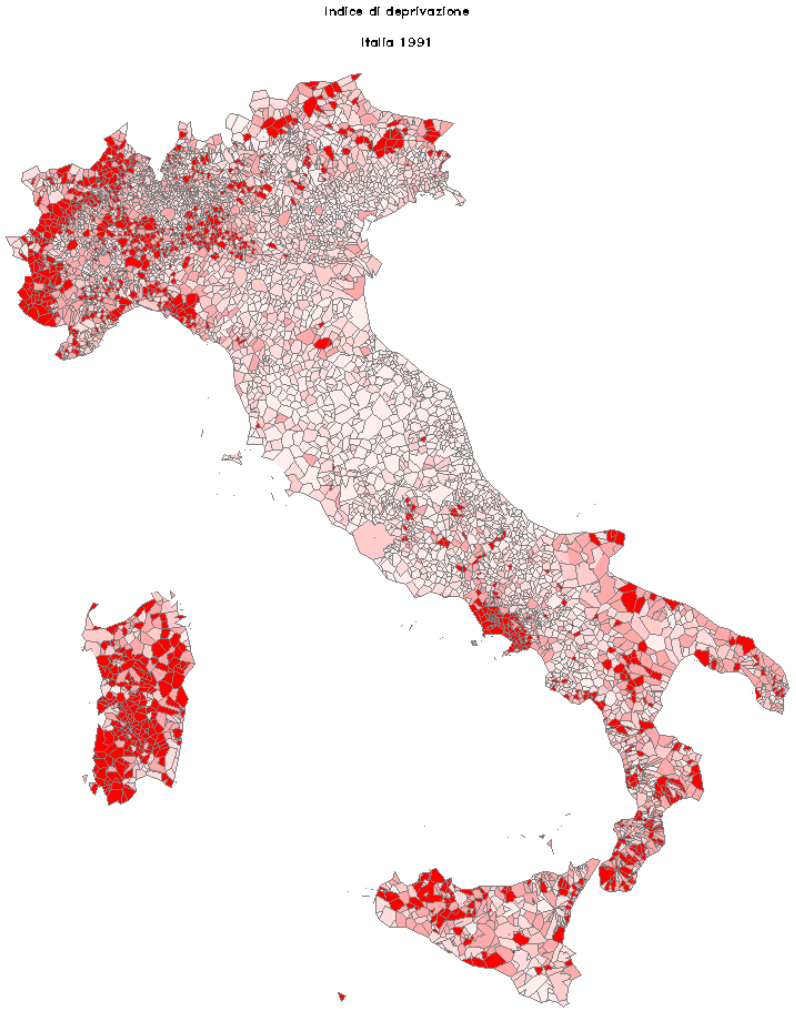
Deprivation Index 2001



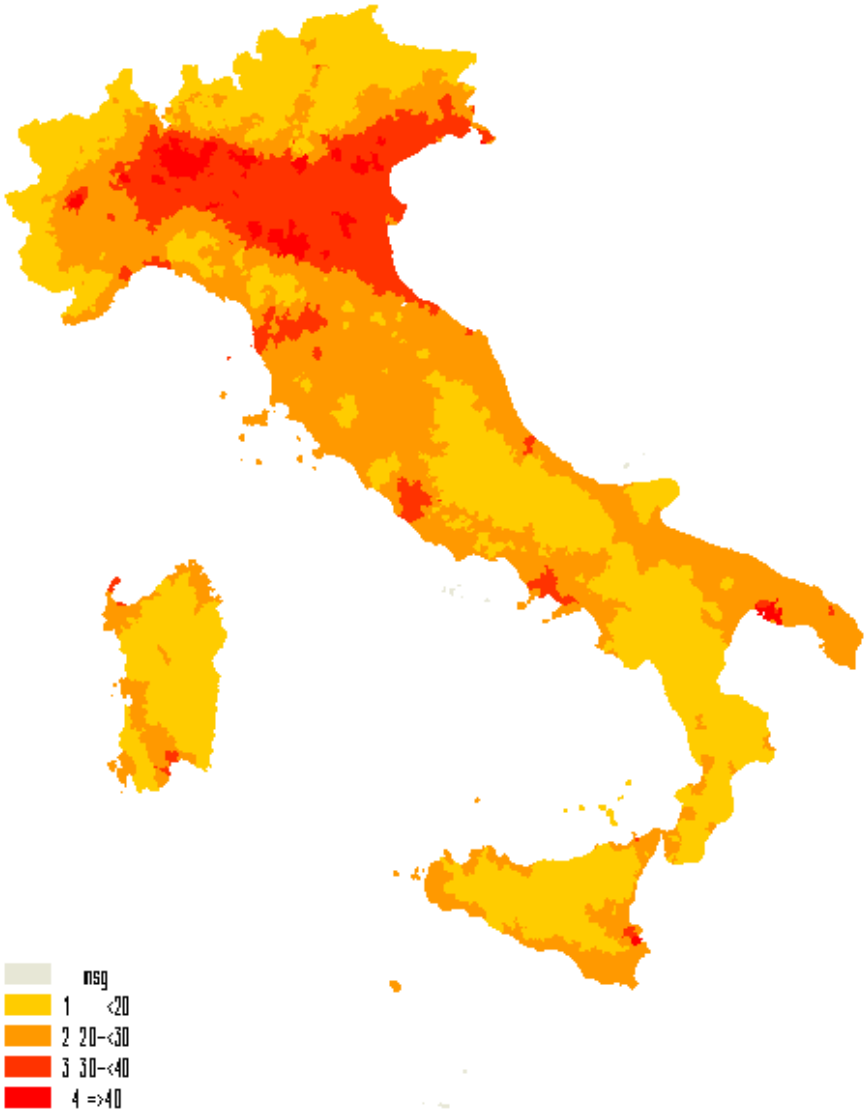
PM10



Deprivation Index 1991



PM10



Results (Italy)

	NO2	Pm10	Pm25
All natural	0.976 (0.973-0.978)	0.978 (0.974-0.983)	0.970 (0.958-.986)
Circulatory system	0.965 (0.958-0.971)	0.963 (0.957-0.970)	0.966 (0.958-.0976)

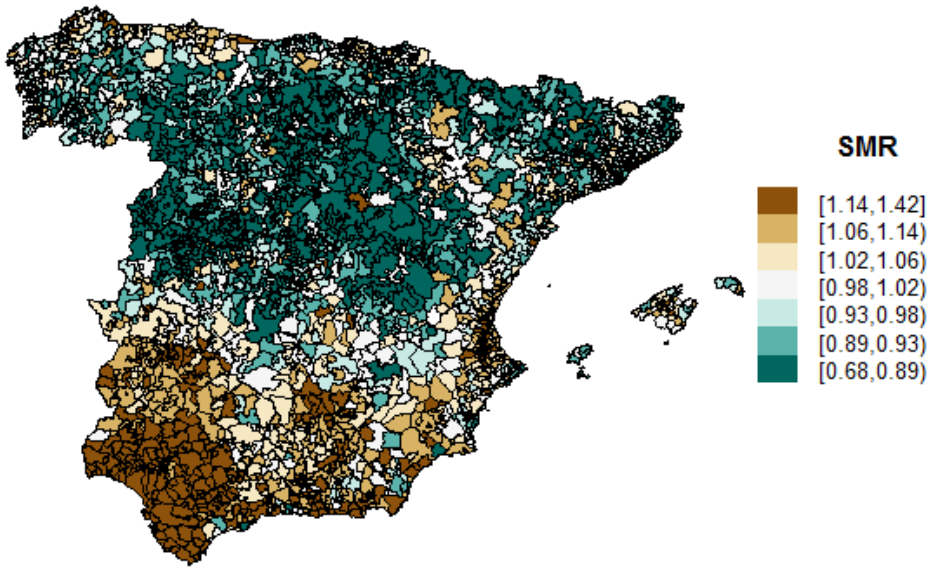
- Results not affected by inclusion of spatial random effects

Results (Italy)

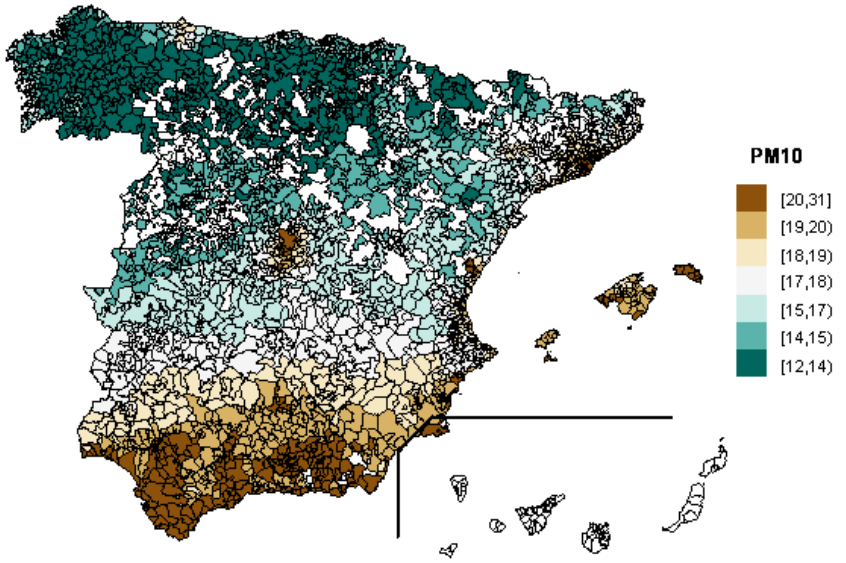
- Totally opposite results than the analysis with individual data.
- It shows how wrong an ecological analysis can be.

Spain

SMR



PM10

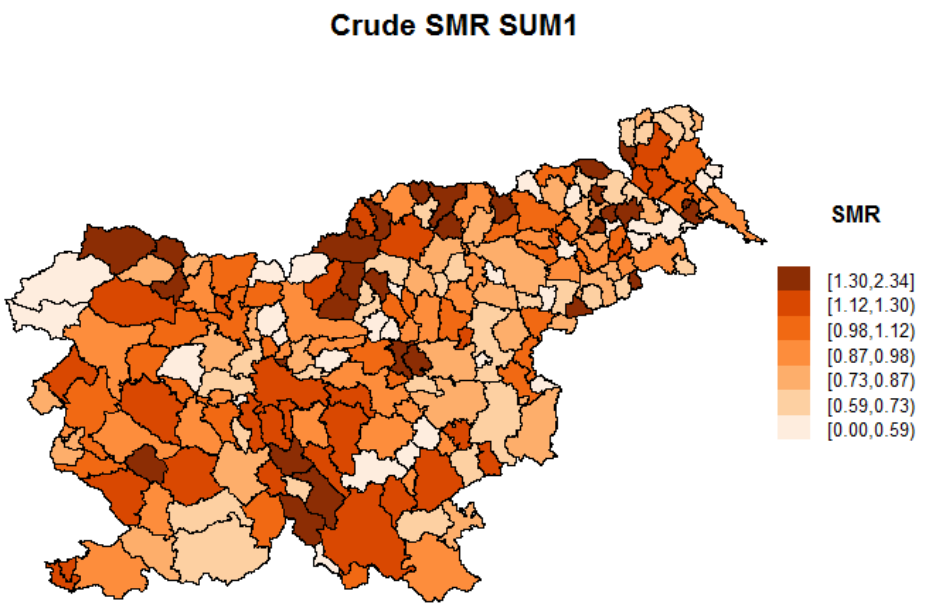


Results (Spain)

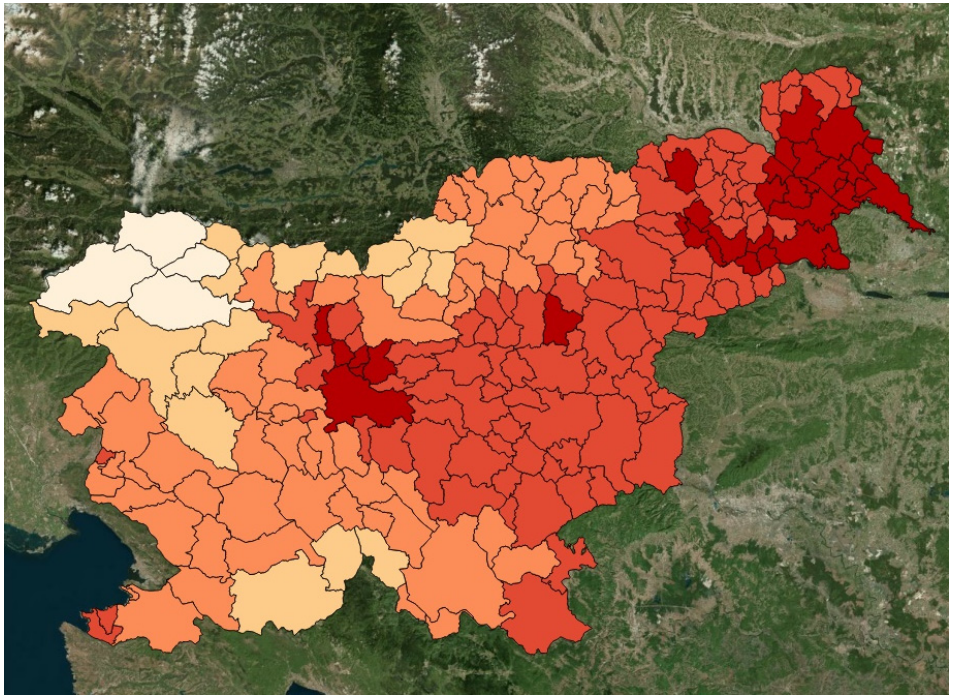
	Unadjusted RR (95% CI)		Adjusted RR (95% CI)		Adjusted + spatial term RR (95% CI)	
PM_{10} (5 µg/m^3)	1.106	(1.095, 1.117)*	1.077	(1.068, 1.088)*	1.016	(0.992, 1.041)

Slovenia

SMR



PM10



Analisi ecologica Slovenia

Main Mortality cause	PM10 RR for 10 μg/m^3 increase	
	without spatial term	with spatial term
All natural causes	**1.009 (1.004-1.015)**	**1.009 (1.004-1.015)**
Cardiovascular diseases	**1.012 (1.005-1.018)**	1.008 (0.996-1.018)
Respiratory Diseases	**1.014 (1.000-1.028)**	**1.014 (1.000-1.028)**
Neoplasms	1.003 (0.998-1.007)	1.003 (0.998-1.007)
Lung cancer	0.989 (0.981-0.997)	0.989 (0.981-0.997)

Main Morbidity cause	PM10 RR for 10 μg/m^3 increase	
	without spatial term	with spatial term
All natural causes	1.000 (0.996-1.005)	1.006 (0.999-1.014)
Cardiovascular diseases	0.996 (0.991-1.002)	1.004 (0.994-1.014)
Respiratory Diseases	1.005 (0.998-1.012)	**1.014 (1.002-1.027**)
Neoplasms	0.998 (0.992-1.003)	1.005 (0.994-1.015)
Lung cancer	0.976 (0.960-0.992)	1.013 (0.984-1.045)

Conclusions

- Ecologic analyses are known to have several potential pitfalls.
- They can provide true results, but there is no way to know unless we have individual data.
- Ecologic analysis always should be interpreted with caution, even when producing expected results.
- Efforts should be made to collect individual data. The MEDHISS approach offers a cost-efficient solution using existing datasets.

Conclusions (2) - addendum

- In air pollution epidemiology the results of an ecologic analysis between mortality (morbidity) and pollutants depend from the spatial structure of other confounders (deprivation, smoking prevalence, obesity prevalence) and its availability
- Given a good set of confounders at the same ecological level the ecologic analysis can obtain an estimate of the true risk even if lower than the true one due to ecological fallacy

Conclusions

- MEDHISS approach: semi-individual analysis.
- Much more valid than ecologic studies.
- If Health Surveys are able to provide geographical coordinates of addresses, other air pollution models that capture within-city variations can be applied.