The use of current data to evaluate the health impact of environmental pollution: the “SENTIERI approach” and the case study of Taranto.

L’uso dei dati correnti per valutare l’impatto sulla salute dell’inquinamento ambientale: “l’approccio SENTIERI” ed il caso di Taranto.

Authors:
Susanna Conti (1), Valerio Manno (1) and Giada Minelli (1) on behalf of the SENTIERI Working Group (the list of the WP has been published in the Reference no. 3)
(1) Ufficio di Statistica dell’Istituto Superiore di Sanità

Abstract in Italiano
Lo studio dell’impatto sulla salute dell’inquinamento ambientale ha ormai assunto un grande rilevo in ambito di Sanità Pubblica e per questo scopo risulta particolarmente utile analizzare i dati correnti (riguardanti la mortalità, le ospedalizzazioni, le caratteristiche demografiche della popolazione) che sono di buona qualità, esaustivi sul territorio nazionale, con dettaglio a livello comunale e disponibili su lunghe serie temporali. In questo ambito, si è sviluppato lo studio SENTIERI, che ha elaborato un approccio originale la cui validità è stata riconosciuta dall’OMS e che viene utilizzato correntemente dall’Istituto Superiore di Sanità per valutare lo stato di salute delle persone che risiedono in aree inquinate. Questo lavoro descrive le principali caratteristiche di tale approccio e ne illustra l’applicazione ad una situazione italiana ormai ben conosciuta e molto dibattuta sia in ambito scientifico che opinione pubblica, nonché oggetto di indagini della magistratura: il sito di Taranto.

Abstract in English
The study of the health impact of pollution is a relevant issue in Public Health and for this purpose is very useful to analyze current data (regarding mortality, hospitalization, demographic characteristics of the populations) that are of good quality, exhaustive for the whole Country at municipality level and available for a long time period. In this framework, the Study SENTIERI has been carried out, developing an original approach, aimed at assessing the health status of people resident in the Italian National Polluted Sites. This approach has been considered valid by WHO and is currently adopted by the Istituto Superiore di Sanità to evaluate the health status of residents
in contaminated sites. This paper regards the main characteristics of such an approach, and presents its application to the site of Taranto, the situation of which has been object of scientific studies, heated debate on the media and also of official inquiries.

**Parole chiave**: Mortalità, Ricoveri Ospedalieri, Siti contaminati, Valutazione a priori, Taranto.
**Key words**: Mortality, Hospitalization, Contaminated Sites, A priori evaluation, Taranto

1. Introduction

The health impact of environmental pollution is an important issue of Public Health; in particular, the health of people living in contaminated site is affected by the legacy of past industrialization and current industrial activities, often in absence of environmental remediation.

The first issue to address when taking into consideration the health impact of pollution is the definition of “contaminated site” (CS). The term CS can have different meanings. A general definition, following the public health perspective, is “areas hosting or having hosted human activities which have produced or might produce environmental contamination of soil, surface or groundwater, air, and food chain, resulting or being able to result in human health impacts” [1]

Contaminated sites are extensively present in Europe (approximately 250,000 sites require clean up, as listed by the European Environment Agency [2]). Several thousands of these sites are located in Italy and a total of 57, defined as National Priority Contaminated Sites (NPCSs), qualify for remediation because of contamination documented in qualitative and/or quantitative terms, and because of a potential health impact (National Priority Contaminated Sites-NPCSs).

SENTIERI Study (Epidemiological study of residents in National Priority Contaminated Sites - NPCSs, formerly funded by the Italian Ministry of Health) evaluated mortality among residents of 44 NPCSs included in the “National Environmental Remediation Program” [3,4].

The distinguishing feature of SENTIERI Project is the “a priori evaluation” of the epidemiological evidence of the causal association for each combination environmental exposure/cause of death selected for the analysis. This approach is interesting, since when performing epidemiologic studies, there is a risk for researchers to become data-driven. This can be the case when commenting results for causes showing an increase, possibly on the basis of statistical significance. In the SENTIERI Project, a standardized procedure was set up to collect the available epidemiological literature, which was reviewed and led to classify each cause of death (for 63 causes) and environmental exposures combination (chemicals, petrochemicals and refineries, steel plants, power plants, mines and/or quarries, harbor areas, asbestos or other mineral fibers,
landfills and incinerators - labelled on the basis of the decrees defining the sites’ boundaries) in terms of strength of causal association: Sufficient to infer the presence of a causal association (S), Limited to infer the presence of a causal association (L), and Inadequate to infer the presence or the absence of a causal association (I). The procedures and results of the evidence evaluation have been published [3].

SENTIERI analyzed mortality at municipal level in the period 1995-2002. In further Projects, the SENTIERI approach has been used to update data regarding CSs analyzing the most recent current data available (nowadays, the most recent official mortality data refer to the year 2011).

Moreover, the SENTIERI approach has been used to describe also the hospitalization among persons residing in CSs.

2. Material and methods

2.1 Mortality

The data source for the mortality analysis is the Italian Mortality Database, based on official data at municipality level, from the Italian National Institute of Statistics (ISTAT), run by the Statistics Unit of Italian National Health Institute (Ufficio di Statistica dell’Istituto Superiore di Sanità).

Mortality data are codified according to the International Classification of Disease (ICD); it has been revised approximately every 10 years; the purpose of the revision is to stay abreast of medical advances in terms of disease nomenclature and etiology. In Italy, deaths have been codified according to the Ninth Revision (ICD-9) until 2002 (5); since 2003, the Tenth Revision (ICD-10) has been adopted (6). ICD-10 differs from ICD-9 in several respects: ICD-10 is far more detailed than ICD-9, with about 12,000 categories compared with about 5,000 categories; ICD-10 uses alphanumeric codes compared with numeric codes in ICD-9 (7). Data referred to the years 2004 and 2005 are not available from ISTAT.

For each NPCS the causes of death selected by the above mentioned procedure of a priori evaluation are examined. In order to have a general description of the residents’ health profile, main broad groups of causes of death were also considered. The complete list of the selected 63 causes has been published (4) and is also shown in Table 1.

Standardized mortality ratios (SMR) adjusted for an ad hoc deprivation index together with 90% Confidence Intervals (90% CIs) are computed using regional rates for comparison [4].
SMR$_i$, referred to residents in a selected geographic area $i$, is defined as the ratio between the number of observed and expected deaths, $\text{SMR}_i = \frac{\sum_j e_j}{\sum_j T_j n_j} \times 100$

Where, for each age-group $j$: $e_j$ are the observed events (deaths), $n_j$ is the population and $T_j$ is the death rate of the reference population (regional population).

As well as the point estimation of SMR, also its confidence interval was calculated (CI 90%); if the observed deaths were less than 100, CI was estimated based on Poisson’s distribution, otherwise, on the Byar method.

### 2.2 Hospitalization

The source of data regarding hospitalizations in Italy is the Hospital Discharges Records Data Base (HDF); in Italian, “Schede di Dimissione Ospedaliera” (SDO). Information on hospitalizations contained in HDRs are collected by all Italian public and private hospitals and are then transmitted to the Ministry of Health. The data collected include information such as demographics (gender, date and place of birth, place of residence and so on), admission and discharge data, principal diagnosis and up to five secondary diagnoses. To code diagnoses, ICD 9-CM (International Classification of Diseases-Clinical Modification, 9th Revision) is used.

The use of HDF for the purpose of describing the health status of residents in CSs in Italy has been discussed [8].

In the HDR Data base that the Ministry of Health transmits to the Unit of Statistics of the Italy’s Institute of Public Health, each discharge form contains also an anonymous code, unique for each individual; therefore, it is possible to analyze the information regarding persons that are hospitalized, and not only the hospitalizations. Only the main diagnosis reported in the HDF was taken into consideration and if, during the study period, an individual has been hospitalized for the same diagnosis more than once, only the first hospitalization occurred during the study period was considered [8,19].

The list of diagnoses taken into consideration is presented in Table 2.

Standardized hospitalization ratios (SHR) adjusted for an ad hoc deprivation index together with 90% Confidence Intervals (90% CIs) were computed using regional rates for comparison.

### 3. DISCUSSION
Some relevant issues regarding the above mentioned approach are discussed: reference population, social and economic potential confounding, the choice of 90% probability level and, finally, main limitations and strengths.

### 3.1 Reference Population

For the reference population the same data of the area units under study are needed: cases and populations stratified by gender and age categories.

The reference population should be selected considering two different needs: i) it should be comparable to the studied populations for factors that can affect the health profile with the exception of the contamination at study—the differences in the health profile between the compared populations should be ideally due only to the differences in environmental exposures, namely, to the contamination; ii) it should be sufficiently numerous to obtain stable reference rates also for rare diseases. These two needs have opposite requirements, as the first one is usually negatively correlated with the dimension of the population, while the second one is positively correlated with the dimension of the population. The reference populations should be selected balancing these two needs. Usually one or two populations among the following are selected as reference population: national, and regional, local (i.e., a population composed of populations residing in the neighborhood of the contaminated area). In the SENTIERI approach, the Regional Population has been used as reference.

### 3.2 Social and economic potential confounding

In geographical studies of environment and health, confounding from social and economic factors may occur. To control such confounding effect, standardization techniques have been extensively used since the mid-1990s, taking into consideration the “deprivation”; deprivation can be defined as “a state of observable and demonstrable disadvantage relative to the local community or the wider society or nation to which an individual, family, or groups belong” [9]. Deprivation indices are area-based measures of material and social disadvantageous circumstances, that is, indicators of relative deprivation at population level.

To account for possible confounding from socioeconomic factors in SENTIERI project, an ad hoc Deprivation Index was built and applied to the SMR estimates, (SENTIERI DI). The deprivation index was constructed using the 2001 national census variables representing the
following socioeconomic domains: education, unemployment, dwelling ownership, and overcrowding [4, 10].

3.3 The choice of 90% probability level

In SENTIERI Project 90% Confidence Intervals were computed for risk estimators to present the range of uncertainty. Estimators were not accompanied by a hypothesis test to discriminate statistical significant from non statistical significant results. The choice of 90% level was made to minimize the a critical use of CI as surrogate of hypothesis testing; such use could lead to consider relevant only those estimators for which the CI exclude the null value, i.e. the ones customarily defined as “statistical significant” [11,12]. The discriminating use of statistical significance in the evaluation of causal associations in epidemiology has been discussed since 1965 [13] and recently reproposed [14]. This is particularly the case in SENTIERI Project were an a priori evaluation of the epidemiological evidence was completed to identify those diseases for which the causal association with the environmental exposures was either ascertained or probable. In SENTIERI the primary interest is on excesses or defects, i.e. direction of the risk estimators, and after on their size and precision, the latter inferred from the range of CI.

3.3 Main limitations of the SENTIERI approach

In environmental epidemiology exposure ascertainment is a key phase because the exposure/s affecting the study population should ideally be described in detail, while in most instances the available exposure information is indirect and qualitative. In ecological investigations the exposure/s can be a single event from a point emission source of some contaminants, more often the contaminants are a heterogeneous mixture progressively polluting different matrices in the area. For example, in SENTIERI Project the sources of environmental exposures were abstracted from the legislative Decrees defining sites’ boundaries and fixed on the basis of the possible sources of contamination (e.g. chemical industry, steel plants, landfills). A further limitation lies in the implicit assumption that all residents in the area under investigation experience the same exposures, while exposure variability is likely to be substantial, due to many factors (e.g. concentration of contaminants and their diffusion to soil and water, distance of residence from polluting sources).

The possible consequences of such non differential exposure misclassification are complex and direction of the resulting bias is not predictable [15]. In addition, information exposure source with possible health impact, such as concurrent air pollution from road traffic and exposures in the occupational setting, often are not available. Finally, vital statistics are accessible for a given
administrative area whose boundaries hardly correspond to the distribution of environmental pollutants, so that the misclassification of exposure (and loss of statistical power) is common. A more detailed description of these limitations of ecological study design is available [15,16,17].

3.2 Major strengths of the SENTIERI approach

Its major strengths are the standardization of the mortality analysis and NPCSs classification in terms of environmental exposure which allow the study of all NPCSs in one country; the a priori evidence evaluation to comment and interpret study results is a key characterizing element of the project. Additional assets are that the mortality analysis can be updated and other vital statistics data can be analyzed; also the a priori evidence evaluation can be brought up to date following the established criteria and procedures.

The SENTIERI approach is currently used by the Istituto Superiore di Sanità (ISS) to describe the health profile of people resident in polluted areas, to give answers to questions asked by members of Italian Parliament (“Interrogazioni Parlamentari”) or of Government.

The SENTIERI Project is also part of the Italian Official Statistics Plan (“Programma Statistico Nazionale”) [18].

The application of the SENTIERI approach to the site of Taranto, the situation of which has been object of many scientific studies, heated debate on the media and also of official inquiries is briefly presented.

4. The case of Taranto

The NPCS of Taranto is of major interest because of the presence of several polluting sources, such as a large steel plant (ILVA), a refinery, the harbor, and both controlled and illegal waste dumps.

The site is located in Apulia Region (Southern Italy), includes two municipalities (Taranto and Statte) and 214,348 inhabitants at the 2011 Census. The demographic background of residents is shown in Figures 1 (NPCS of Taranto) and 2 (Apulia Region). To compare the age structures, the Kolmogorov-Smirnov test was performed: the result was significant, both for male and female, indicating a different age structure. The age pyramids shows that the population of Taranto NPCS is older than Apulia population.
In this paper the results regarding the analysis, according to the SENTIERI approach, referring to mortality (2003-2010) and hospitalization (2005-2010) are presented.

4.1 Mortality

Number of observed deaths, standardized mortality ratios adjusted for deprivation and confidence intervals are shown in table 1.

Among both genders there was an excess in overall mortality, and in mortality from all neoplasms and from diseases of circulatory, respiratory and digestive systems. Mortality from many specific causes shows excesses in both genders; among neoplasms: liver cancer, non Hodgkin lymphoma; among non neoplastic pathologies: infectious disease, dementias, hypertension, ischaemic heart disease, cirrhosis. Among men only, there are excesses in mortality from pancreas cancer, melanoma and myeloid leukemia; among women only, there are deaths in excess from neoplasms of lympho-haematopoietic system and multiple myeloma.

Taking into considerations the causes of death associated with the environmental exposure in the Site of Taranto, in both genders excesses of death were observed for lung cancer and respiratory system diseases (overall, both acute and chronic); mortality from pleural mesothelioma shows an excess among men.

4.2 Hospitalizations

Number of observed cases, standardized hospitalization ratios adjusted for deprivation and confidence intervals are shown in table 2.

Many diagnoses present excesses in hospitalization: all natural causes examined, among both genders, all malignant neoplasms, many specific neoplasms and many important diseases.

4.3 Conclusions

In Taranto NPCS many epidemiological studies has been conducted (cohort studies regarding mortality and morbidity, based also on data from Cancer Registries) and their results have been published [19,20]; in both genders, excess risks for a number of causes of death and hospitalization have been observed; for these causes, an etiologic role of environmental exposure present in Taranto NPCS can be supported on the basis of a priori evaluation of the epidemiological evidence completed in SENTIERI.
References

1) WHO, Contaminated Sites and Health. Report of Two WHO Workshops: Syracuse, Italy, 18 November 2011; Catania, Italy, 21-22 June 2012, WHO Regional Office for Europe, Copenhagen, Denmark.


8) Biggeri et al. Environment and health in high risk areas of Sardinia. Epidemiol Prev 2006, Suppl.1


17) Savitz DA. A niche for ecologic studies in environmental epidemiology. Epidemiology 2012; 23; 53-54.

