ON THE LEUKEMIA MORTALITY IN THE PROVINCE OF TARANTO (ITALY): A FIRST HISTORICAL AND GENDER PERSPECTIVE

MARIA ROSA MONTINARI1, GIANFAUSTO SALVADORI2, SIMONA GIARDINA3

1 Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, Lecce - 2 Dipartimento di Matematica e Fisica, Università del Salento, Lecce - 3 Facoltà di Medicina e Chirurgia, Università Cattolica del Sacro Cuore “A. Gemelli”, Roma Italy

ABSTRACT

Introduction: Taranto is the most polluted area of Italy (and, probably, of Europe), characterized by a strong industrial presence since the early sixties. The pollutants involved in this study are carcinogens, mutagens, teratogens and neurotoxins. The target of this study is to look for possible temporal trends of leukemia mortality in this small province (no similar systematic studies are available), and to provide a historical, occupational, and gender interpretation of the results. In addition, a comparison with the overall Italian situation is presented.

Materials and methods: The data represent the number of deaths per year in the province of Taranto, subdivided into Total Deaths and Leukemia’s Deaths. The observations cover a period of 35 years, from 1969 to 2003, and are available for both the female and male sub-populations separately. The analyses are carried out using Least Squares fits and Linear Regressions (Simple, Robust, and Generalized).

Results: A statistically significant increase of leukemia mortality for the whole population is found, in contrast to the trends globally found in Italy.

Conclusions: The study supports the association of larger leukemia mortality with residential proximity to sources of industrial pollutants, and emphasizes the urgency of an actual application of the existing regulations about pollutant emissions.

Key words: [MeSH 2010]: Leukemia [C04.557.337], (Generalized) Linear Models [E05.318.740.500.500], Environmental Pollution [N06.850.460], History of Medicine [K01.400.552], Taranto (Italy) province.

Received November 30, 2014; Accepted May 02, 2015

Introduction

The metropolitan area of Taranto (Apulia region, Southern-Italy), characterized by a strong industrial presence since the early sixties, is an “environmental high-risk area”. Taranto is the most polluted area in Italy, and probably in Europe. In Taranto there is one of the biggest full-cycle steel plants in Europe (the Ilva), a cement factory of national importance, a huge petrochemical refinery, and two thermoelectric plants using the port area for the transit of goods and raw materials. There are also chemical factories producing glue, paint and synthetic enamel, factories that synthesize petrol and coal derivatives, factories producing rubber and plastic, second processing metallurgic industries, and factories producing mechanical, electronic and electro-technical components. Moreover, there are two incinerators and a public disposal plant.

According to the data gathered by the National Inventory of Sources and Emissions (INES)(1), Taranto is at a top place in Italy for industrial emissions: the involved pollutants are carcinogens, mutagens, teratogens and neurotoxins. The INES report highlights extremely high concentrations of Polychlorinated Dibenzo-p-Dioxin (PCDD) and Polychlorinated Dibenzo-p-Dioxin (PCDF) emitted by Ilva: they represent 92% of the total national emissions. Ilva is also a
great producer of Aromatic Polycyclic Hydrocarbon (APH). Taranto has also a record for the emission of other carcinogenic substances, emitting about 95% of the total national production of benzo(a)pyrene, benzo(b,j,k)fluoranthene, benzo(a)anthracene, and dibenzo(a,h)anthracene. This town also generates 42% of the benzene produced on a national scale, and (as remarked by INES) extremely elevated are also the emissions of lead (78%), mercury (57%), cadmium (42%), and chrome (31%).

The terrible environmental situation in Taranto is well documented in a recent book: the title, “15 Steps”, is emblematic, and indicates the distance between the Ilva site and the Saint Brunone cemetery, where ‘many of the victims of the biggest (but silent) Italian environmental disaster are buried’ (the satellite images of the area, freely available on Internet, are indeed impressive).

The target of this study is to identify the presence of possible temporal trends of leukemia mortality within the area of Taranto: a statistical investigation is carried out, in order to understand whether there might exist a significant gender difference in leukemia mortality, possibly due to the role of exposure in occupational environments. In addition, a comparison with the overall Italian situation is presented.

**Leukemia: historical aspects**

Before the nineteenth century, there was nothing in literature indicating a diagnosis of leukemia. The disease might have been observed the first time by Alfred Donné, a French physician and a pioneering microscopist, whose findings were not published until 1855. Instead, in 1845, John Hughes Bennett, Professor of the Institutes of Medicine at the University of Edinburgh, published in the *Edinburgh Medical and Surgical Journal* the first definite description of leukemia. In 1846, Henry Fuller, a physician at St. George’s Hospital in London, examined under the microscope a patient’s blood probably suffering by chronic granulocytic leukemia. This was the first case of leukemia diagnosed, in a living patient, with a blood microscopic examination. In 1847, Virchow first called this disease “Leukhemia” (white blood). He reported again, in 1849, a case of a patient with hypertrophy of the spleen and suggested two forms of the disease, distinguished by the site of origin: the splenic and the lymphatic forms. In 1852, Bennett published a monograph on leucocytahaemia, in which he included the first illustrations of the microscopic appearance of the blood in leukemia.

In 1868, Ernst Neumann, a pathologist at the University of Konigsberg, particularly known for having described a progenitor cell of red blood cells (cell Neumann), was the first to note changes of the bone marrow in leukemia, proposing the term “myelogenous leukemia”.

Regarding therapy, arsenic was the first agent used in the treatment of certain forms of leukemia and in some cases it produced a short remission. Lissauer, a German physician, administered arsenic to a woman with chronic myeloid leukemia. It was the first report of the use of arsenic in a patient with leukemia.

The first transfusion of the blood in the treatment of this emopatia was carried out by George William Callender at St. Bartholomew’s Hospital in London, in 1873. With the discovery of X-rays by Wilhelm Rontgen, in 1895 a new treatment of leukemia began. Initially, the results were similar to those produced by arsenic.

Chemotherapy developed only in 1942, in the treatment of neoplastic diseases. In America, Gilman and Philips made the first clinical observations on nitrogen mustard, published four years later. They demonstrated that methyl-bis-(β-chloroethyl) amine had a significant activity against Hodgkin’s disease.

In 1943, it was shown that preparations of folic acid, a vitamin in the B group, inhibited the growth of experimentally induced tumors in mice and rats. SubbaRow, Chief of the Lederle Laboratories of the American Cyanamid Company, synthesized the first folic acid antagonists which were used for clinical trials. The promising results obtained with folic acid antagonists - pteroylaspartic acid and methylpteroyic acid in the first trial - encouraged further the research and SubbaRow developed the synthetic compound 4-aminopteryglutamic acid, named aminopterin. Shortly after, aminopterin was substituted by methotrexate. In 1949, it was described the therapeutic role of the adrenal corticosteroids, in particular of prednisone, and, a few months later, of 6-mercaptopurine.

**Materials and methods**

The collection of the data required more than one year. No other systematic studies in the same
region are available, and so no reference to existing relevant literature can be made. Four main sources were used to fix the database: (a) the National Archive (branch of Taranto); (b) the Historical Archive of the Town of Taranto (including the Province); (c) the major Hospitals in the province of Taranto (the ones having medical structures suitable for dealing with leukemia); (d) the National Institute of Statistics (ISTAT)\(^{(17)}\).

The study is based on distributions of absolute numbers of deaths occurred in the province of Taranto every year, subdivided into two main classes: “Total deaths” (i.e., due to any cause), and the subset “Leukemia’s deaths”. The historical period considered is 1969-2003, i.e. 35 years of annual occurrences. Note that, although more data are available before 1969, these are not taken into consideration, since they are not yet available in a certified electronic format.

Under the term “leukemia” we consider here all the forms of leukemia as defined by the International Classification of Diseases (ICD). More precisely, the following nosological codes have been used:

- for the period 1969-1979: ICD 8, 2040-2079\(^{(18)}\);
- for the period 1980-2002: ICD 9, 2040-2089\(^{(19)}\);
- for the year 2003: ICD 10, C910-C959\(^{(20)}\).

For the sake of the analysis, the observations have been grouped into age-classes of 10 years each. These classes (or their union) broadly individuate peculiar ages of the standard human life: \(\{0-9\}\) years: childhood; \(\{10-19\}\) years: adolescence; \(\{20-29\}\) & \(\{30-39\}\) years: adulthood; \(\{40-49\}\) & \(\{50-59\}\) years: maturity; \(\{60-69\}\) & \(\{70-79\}\) years: oldness; \(\{80+\}\) years: senescence. Thus, a total of nine age-classes is used in this work, the last one including elder people with an age of 80 years or more.

The investigation is carried out using Least Squares fits\(^{(21)}\), Linear Regressions\(^{(22, 23, 24)}\), and Generalized Regressions\(^{(25, 26)}\). In particular, we are interested in testing whether leukemia mortality could be considered as statistically constant, or otherwise increasing (respectively, decreasing). For this purpose, both Simple, Robust, and Binomial Regression algorithms are used. The analyses are performed using the software MATLAB (The MathWorks, Inc., © 1994-2015). The standard linear regression model used in this work is as follows:

\[
Y_i = \theta_0 + \theta_1 x_i + \epsilon_i \quad (1)
\]

with \(i=1,\ldots,n\) (here \(n = 35\)). The explanatory variables \(x_i\)'s represent the years (from 1969 to 2003), the \(Y_i\)'s are the corresponding random responses (i.e., the percentage of people died of leukemia during the year \(x_i\) with respect to the total number of deaths in the same year), the \(\epsilon_i\)'s are the random fluctuations (assumed to be independent identically distributed, with Gaussian law \(N(0,\sigma^2)\)), and \(\theta_0, \theta_1\) are the linear regression parameters (respectively, intercept and slope). Our main target is to investigate whether or not \(\theta_1\) is significantly different from zero (via the associated p-Values): this would mean that leukemia mortality is statistically non-constant in time.

The generalized regression used here is the standard Logit model, since it is most appropriate to fit the process of death counting:

\[
E\left(\frac{L_i}{T_i}\right) = p_i = \frac{e^{\theta_0 + \theta_1 x_i}}{1 + e^{\theta_0 + \theta_1 x_i}} \quad (2)
\]

with \(i=1,\ldots,n\). The variables \(L_i\) and \(T_i\) represent, respectively, the number of deaths due to leukemia and the total number of deaths during the year \(x_i\), the \(p_i\)'s are the probabilities of “success” (i.e., death due to leukemia), and \(\theta_0, \theta_1\) are the generalized regression parameters. Our main target is to investigate whether or not \(\theta_1\) is significantly different from zero (via the associated p-Values): this would mean that the probability of dying of leukemia is statistically non-constant in time.

**Results**

Hereinafter, we shall investigate the occurrence of leukemia in each age-class, as well as the temporal trends of leukemia mortality in both Females (\(F\)), Males (\(M\)), and the whole Population (\(P\)), obtained by merging the sub-populations “\(F\)” and “\(M\)”\(^{(27)}\). The outcomes will be commented later in the Discussion section.

**Age-class occurrences**

The first result of interest concerns the occurrence of leukemia in each age-class. In figure 1 we show the mean annual number of deaths of Leukemia during the period 1969-2003, for each age-class: this provides some information about which part of the population is (on average) numerically more affected by such a disease. The error bars represent the standard error. The two horizon-
tal dashed lines indicate the value of the mean calculated over aggregation of age-classes (only the whole Population is considered here): the thick dashed line corresponds to the range \{0-49\} years, and takes on the value \(\approx 1.78\); the thin dashed line corresponds to the range \{50+\} years, and takes on the value \(\approx 6.19\).

**Age-class percentages**

A further result of interest concerns the percentage of deaths of leukemia with respect to the total number of deaths in each age-class. In figure 2 we show the percentage of deaths due to leukemia during the period 1969-2003, for each age-class: this provides some information about which range of ages is more “risky”. The error bars represent the standard error.

As an empirical attempt, we fit (via a Least Squares technique) the following Lognormal-like function \(f\) to the percentages of deaths of leukemia shown in figure 2:

\[
f(t) = \alpha \exp\left\{\ln(t)-\beta\right\}/\gamma
\]

with \(t>0\), \(\alpha>0\), \(\beta\in\mathbb{R}>0\), and \(\gamma>0\). As abscissas of the available data to be used in the fit, we take the median age of each age-class (i.e., \(t_1=5\) years, \(t_2=15\) years, ...). The results are shown in figure 3: the error bars represent the standard error. The estimates of the parameters are reported in table 1. Visually, in all cases the function \(f\) provides acceptable fits: actually, the corresponding values of \(R^2\) are always larger than 0.9.

**Temporal trends of leukemia mortality**

The search for possible temporal trends of leukemia mortality (here, percentage of leukemia’s

---

![Figure 1: Mean annual number of deaths of leukemia for each age-class.](image1)

![Figure 2: Percentage of deaths of leukemia for each age-class.](image2)

![Figure 3: Fits of the percentage of deaths of leukemia for the Female sub-population (A), the Male sub-population (B), the whole Population (C), and comparison of the fits (D).](image3)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Female (F)</th>
<th>Male (M)</th>
<th>Population (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)</td>
<td>6.27</td>
<td>2.99</td>
<td>3.89</td>
</tr>
<tr>
<td>[95% C.I.]</td>
<td>[4.93, 7.60]</td>
<td>[2.35, 3.64]</td>
<td>[3.12, 4.66]</td>
</tr>
<tr>
<td>(\beta)</td>
<td>2.71</td>
<td>2.68</td>
<td>2.71</td>
</tr>
<tr>
<td>[95% C.I.]</td>
<td>[2.52, 2.90]</td>
<td>[2.46, 2.90]</td>
<td>[2.52, 2.90]</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>0.88</td>
<td>1.52</td>
<td>1.21</td>
</tr>
<tr>
<td>[95% C.I.]</td>
<td>[0.51, 1.24]</td>
<td>[0.86, 2.17]</td>
<td>[0.75, 1.68]</td>
</tr>
</tbody>
</table>

**Table 1**: Estimates and 95% Confidence Intervals of the parameters of the function \(f\) given by Eq. (3).
deaths with respect to the annual number of total deaths) is shown in figure 4, considering the period 1969-2003. The study is carried out via both Simple (S.R.) and Robust Regressions (R.R.); the dotted lines are 95% confidence bounds. The analysis of the residuals (see the Supplementary Materials) supports the validity of the regression models adopted.

The study is carried out via Binomial Regressions, and the dotted lines are 95% confidence bounds.

Discussion

In this Section the results presented in previous sub-sections will be commented.

Age-class occurrences

It is interesting to note in Figure 1 that all the “younger” classes (say, ages less than 50 years) are affected by leukemia essentially in the same way, since the corresponding averages calculated over different age-classes are comparable. Furthermore, they are significantly less prone to leukemia than “older” classes: in fact, the average of the mean annual number of deaths for people younger than 50 years is about 1/3 of the one of older people.

Apparently, the Male population is slightly more affected by the disease than the Female one: the largest values are observed in “older” age-classes, with a peak at {70-79} years. This suggests that environmental, and especially occupational exposures, may have an important role in the development of the disease, as remarked in (27), where chemicals and/or other occupational hazards are identified as risk factors. Risk from cancer among workers in the petroleum industry exposed to benzene and other hydrocarbons has been the subject of an extensive literature (28). Benzene is a widely recognized cause of leukemia (29). Case-control studies have found from two-to 4.5-times increases in leukemia risk associated with benzene; risk estimates from occupational cohorts have ranged from less than one to more than seven-times, based on small numbers of observed deaths (30, 31). Indeed, from a sociological point of view, industrial jobs are generally an exclusive of men in Taranto’s province.

Age-class percentages

According to the results reported in (32), children living in areas characterized by high levels of petrochemical air pollution have a significant larger risk of developing leukemia. Furthermore, a recent study (33) associates some genetic polymorphisms with a different susceptibility to leukemia. Indeed, genetic polymorphisms at multiple loci of CYP2E1 (a well-known gene for its capacity to bioactivate many procarcinogens, including benzene and N-nitrosodimethylamine) are likely to be associated with the risk of developing childhood Acute
Lymphoblastic/Lymphocytic leukemia (ALL). Apparently, the results shown in figure 2 are consistent with those of the cited works: in fact, the “adolescence” class {10-19} years is the most “unsafe” one, and then the risk monotonically decreases for older classes.

In addition, as pointed out in[34], patient sex has been identified as a risk factor for numerous long-term adverse outcomes, such as exposure to alkylating agents, anthracyclines, radiotherapy, with female sex more commonly associated with larger risk. Literature data have reported, among women, ALL to be associated with trichloroethylene contamination of drinking water[35]. No association was found for men, and Acute Myeloid leukemia (AML) was unrelated to drinking water contamination. Accordingly, our data show that the percentage of deaths due to leukemia is generally larger for the Female population: in particular, the peak of deaths is maximum for female teenagers ({10-19} years).

The apparently different mortality risk between females and males in Taranto could be interpreted as a greater susceptibility of female sex to local environmental pollution. This point is particularly interesting since, as already mentioned above, from a sociological point of view, women are not expected to suffer from direct occupational exposure, but only from indirect environmental one. Incidentally, in a small district next to the province of Taranto (namely, the area of Salento), a historical investigation of leukemia mortality[36] has shown that female population is more affected than the male one.

The fits shown in figure 3 indicate different behaviors of the percentages of deaths of leukemia. Most importantly, the function f provides an estimate of the mortality risk due to leukemia for any (or all) gender(s), and any age of interest: this may serve as a guideline for further studies. The lower-right plot (D) emphasizes the remarkable difference between the behaviors of the Female and Male sub-populations, at least considering the “younger” age-classes (say, up to “adult-hood”), in agreement with the considerations expressed above. Instead, there are no significant differences between the decay rates of the mortality percentages for “older” age-classes.

**Temporal trends of leukemia mortality**

The statistical results are reported in table 2. The p-Values associated with the estimates of the slope parameter $\theta_1$ in Eq.(1) suggest the following conclusions concerning the test of the null hypothesis $H_0: \theta_1=0$.

[F]: $H_0$ should not be rejected for the Female sub-population, since the corresponding p-Values are larger than 11%.

[M]: $H_0$ should be rejected for the Male sub-population, with a standard 5% Type I error.

[P]: $H_0$ can be rejected for the whole Population, since the corresponding p-Values are about (or smaller than) 1%.

Statistically speaking, we may conclude that a significant increasing trend of leukemia mortality is present considering the whole Population, and is likely if considering the Male sub-population alone. A comparison of the estimated Robust Regression lines is shown in the lower-right plot (D) of Figure 4: in particular, an increase from $\approx$0.72% in 1969 to $\approx$0.92% in 2003 can be estimated for the whole Population.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Female (F)</th>
<th>Male (M)</th>
<th>Population (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>0.0057</td>
<td>0.0075</td>
<td>0.0067</td>
</tr>
<tr>
<td>p-Value</td>
<td>0.1174</td>
<td>0.0336</td>
<td>0.0476</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.0441</td>
<td>0.0512</td>
<td>0.0404</td>
</tr>
</tbody>
</table>

Table 2: Statistical results of the linear regressions (Eq. (1)) shown in Figure 4.

**Temporal trends of the probability of dying of leukemia**

The statistical results are reported in table 3. The statistical significance of the temporal trends is based on the analysis of the p-Values associated with the estimates of the parameter $\beta_1$ in Eq.(2): this suggests the following conclusions concerning the test of the null hypothesis $H_0: \beta_1=0$.

[F]: $H_0$ should not be rejected for the Female sub-population, since the corresponding p-Value is about 7%.

[M]: $H_0$ should be rejected for the Male sub-population, with about a 3% Type I error.

[P]: $H_0$ can be rejected for the whole Population, since the corresponding p-Value is smaller than 1%.

Statistically speaking, we may conclude that a significant increasing trend of the probability of dying of leukemia is present considering the whole Population, and is likely if considering the Male sub-population alone. A comparison of the estimated Binomial Regression lines is shown in the lower-
right plot (D): in particular, an increase from $p \approx 0.0070$ in 1969 to $p \approx 0.0093$ in 2003 can be estimated for the whole Population.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Female (F)</th>
<th>Male (M)</th>
<th>Population (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.0078</td>
<td>0.0088</td>
<td>0.0083</td>
</tr>
<tr>
<td>p-Value</td>
<td>0.0648</td>
<td>0.0252</td>
<td>0.0041</td>
</tr>
<tr>
<td>SE</td>
<td>0.0043</td>
<td>0.0039</td>
<td>0.0029</td>
</tr>
</tbody>
</table>

**Table 3:** Statistical results of the binomial regressions (Eq. (2)) shown in Figure 5.

**Conclusion**

The target of this study is to analyze the leukemia mortality in the province of Taranto, and to try and provide a historical, occupational, and gender interpretation of the results. The literature concerning the overall Italian situation reports the following results. A recent study\(^{[35]}\) has analyzed temporal trends of leukemia mortality in adulthood in the database of the Italian Network of Cancer Registries during the period 1986-1997. Starting from the late 80’s, mortality has decreased both in Female and Male populations, although a slight increase is evident in older age-classes $\{75+\}$ years. Another study\(^{[30]}\), involving the whole Italian population and the period 1970-1999, indicates decreasing temporal trends of leukemia mortality in the age-class $\{0-54\}$ years, and increasing ones in older age-classes $\{75+\}$ years.

Apparently, our results point out that the situation in Taranto is peculiar, and yields temporal trends of leukemia mortality in contrast to the ones globally found in Italy. As an explanation, we refer to the outcomes of several studies (see, among others,\(^{[39]}\) and\(^{[40]}\)), indicating that acute leukemia may be associated with residential proximity to sources of industrial pollutants.

Note that on December 16\(^{th}\), 2008, the Apulia’s Regional Council approved a bylaw\(^{[41]}\) that specified pollutant emissions monitoring systems, and fixed dioxin emissions top limit at 0.4 ng I-TEQ/Nm3, in agreement with the European regulations (Aarhus Protocol), ratified and made executive by the law of March 6\(^{th}\), 2006\(^{[42]}\). However, the date of the complete fulfillment of the regional law, initially fixed on December 31\(^{st}\), 2009, was then postponed on December 31\(^{st}\), 2010, essentially because of the pressure of the industries operating within the Taranto area, which continuously threaten massive dismissals in this economically depressed province.

Only on July 20\(^{th}\), 2012, the Apulia Region approved the Law n. 21/2012 concerning “Rules to protect health, environment, and land from polluting industrial emissions for Apulia districts already declared at large environmental risk”.

We believe that this study emphasizes the urgency of an actual application of the existing regulations about pollutant emissions: as a matter of facts, leukemia is only one, among many others\(^{[2]}\), of the pathologies that affect the population in the highly polluted metropolitan area of Taranto.

**References**


4) Bennett JH. Case of hypertrophy of the spleen and liver in which death took place from suppression of the blood. Edinb Med Surg J 1845; 64: 413-423.

5) Fuller HW. Particulars of a case in which enormous enlargement of the spleen and liver, together with dilatation of all the blood vessels of the body were found coincident with a peculiarly altered condition of the blood. Lancet 1846; 2: 43-44.


7) Bennett JH. Leucocytocemia or White Cell Blood. Edinburgh 1852; 7-82.


Acknowledgments
The Authors wish to express their gratitude to Dr. A. Ciccarese (Direzione centrale per le esigenze degli utilizzatori, integrazione e territorio; ISTAT, Rome) for her collaboration concerning data provision, and to Dr. S. Bruzzone (Direzione centrale per le statistiche e le indagini sulle istituzioni sociali; ISTAT, Rome) for helpful comments and suggestions.

Corresponding author
MARIA ROSA MONTINARI
Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento
Lecce (Italy)