Assessment of death risk for asbestos cancers using functional regression models among dockworkers exposed to asbestos in northeastern Italy

P. Girardi[1*], V. Comiati[2], V. Casotto[2], C. Gaetan[1], M.N. Ballarin[3], E. Merler[4], and U. Fedeli[2]

[1] *Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy; paolo.girardi@unive.it, gaetan@unive.it.*
[2] *Epidemiological Department, Azienda Zero, Veneto Region, Padova, Italy; vera.comiati@azero.veneto.it, veronica.casotto@azero.veneto.it, ugo.fedeli@azero.veneto.it.*
[3] *Health and Safety at Work Department (SPISAL), Local Health Authority 3 Serenissima, Venice, Italy; nicoletta.ballarin@aulss3.veneto.it.*
[4] *Formerly: Regional Mesothelioma Register, Occupational Health Unit, Local Health Authority 6 Euganea, Padua, Italy; enzo.linocarlo.merler@gmail.com.*
[*] *Corresponding author*

Abstract. *The retrospective assessment of individual exposure in occupational settings often derives from the association of individual work histories with quantitative and semi-quantitative exposure information. In absence of information on exposure, researchers commonly employed proxy variables, but with strong assumptions and some limitations. In the present work, we propose a new method to estimate the time-varying exposure risk associated to outcomes of interest considering functional regression models and individual work periods. As an illustrating example to understand the potentiality of the method, we analysed a cohort of dock workers occupationally exposed to asbestos in Italy.*

Keywords. *retrospective exposure assessment; functional regression; asbestos; dockworkers; mesothelioma.*

---

1 Introduction

The strong relationship between malignant mesothelioma risk and the cumulative asbestos dose is confirmed by a wide piece of literature [1, 6, 3]. Unlike mesothelioma, lung cancer (LC) is determined by smoking habit, but asbestos is a recognized risk factor for LC [5], again in a dose-response relationship [8]. An excess LC risk at low cumulative asbestos doses has also been observed in never-smoking workers, and meta-analyses suggest that studies with higher-quality exposure assessment provide evidence of an effect at low doses [4]. The study of the relationship between exposure to asbestos and mortality from asbestos-related diseases often relied on surrogates of exposure, such as work duration and/or prevalent job tasks. However, these surrogates may result in misclassification of individual exposure, affecting risk estimates [11]. The Retrospective Exposure Assessment (REA) is a crucial component for the interpretation of the results derived from epidemiological studies in occupational settings [10]. The REA is based on the reliability of information on exposure levels, which in occupational settings is often weak and insufficient to cover the entire occupational window of interest. In fact, in our motivating example, the concentration of asbestos fibers in the air at work in the Venetian dock is unknown; we only know the number of ships loaded with asbestos and the amount (in kilograms) of asbestos handled each year

in this port between 1961 and 1979. In this study, we consider, in addition to standard methods based on the duration of exposure, a flexible approach based on the functional regression model [9, 7] to estimate a mortality risk function based on the work periods of each worker. The aim of this paper is to estimate the excess risk of death from two asbestos-related cancers, pleural cancer (PC) and LC, in a cohort of dockworkers employed at the Port of Venice, using both traditional mortality methods based on a proxy variable and risk estimates from a functional regression model.

2 Methods

2.1 The CLP Cohort and follow-up

The unloading of asbestos was carried out exclusively by members of a dock company called "*Compagnia Lavoratori Portuali*" (CLP). The cohort includes the dockworkers of the CLP as recorded in its official rosters who i) were employed between January 1, 1960 and December 31, 1992; ii) worked for at least 6 months. The duration of work during 1969-1977 (Figure 1) was used as a proxy for the intensity of asbestos exposure. In the study, vital status was followed from enrollment (date of hire or January 1, 1960, whichever was later) to death or end of follow-up (December 31, 2019) using the regional population registry; underlying cause of death for deceased workers was obtained from death certificates and coded according to the International Classification of Diseases, 9th Revision (ICD-9) until 2007 and ICD-10 since 2007.

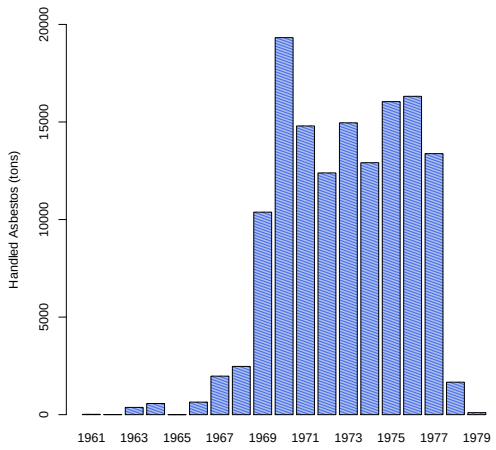

Figure 1: Total quantity (in tons) of asbestos handled in the harbor of Venice by year.

2.2 Standard analysis with proxy variables

Based on the death records and work histories, we conducted a follow-up cohort study, computing for each worker's age, period, and person-years from 1960 to 2018, with a 1-year resolution. The effect on the mortality of PC and LC was modeled by means of Poisson regressions with offset constant as

follows:

$$\frac{g(\mu_i)}{PY_i} = \beta_0 + \beta_1 A_i + \beta_2 P_i + \gamma I(D_i > 5) \quad (1)$$

with the death occurrence $Y_i \sim \text{Poi}(\mu_i)$, while $g() = \log()$ was the canonical link function and PY_i indicated the individual person-years of the follow-up duration to model the mortality rate, as an offset variable. We considered the covariate age A_i and period P_i as confounders in continuous form, while D_i is the duration in the period 1969-1977 (time window with a peak in the amount of asbestos handled at the port). The parameter γ estimated the adjusted risk associated with a work duration longer than five years. Because of the presence of time-varying duration and the occurrence of LC and PC cases during the follow-up period, we recalculate the duration variable D_i according to the period P_i and the temporal window of interest (a,b) with $D_i(P_i) = \sum_{k=1}^{N_w} \int_a^{P_i} I_{\left(s_{i,k},e_{i,k}\right)}(t)dt$ where, each work period k is defined by a starting time $s_{i,k}$ and an ending time $e_{i,k}$. For the worker w, N_w represents the total number of working periods. The model with a time-dependent duration variable becomes as follows

$$\frac{g(\mu_i)}{PY_i} = \beta_0 + \beta_1 A_i + \beta_2 P_i + \gamma I(D_i(P_i) > 5) \quad (2)$$

The results were presented in terms of Mortality Rate Ratios (MRR) by the exponential transformation of γ coefficients with its 95% Confidence Interval (95% CI). The computation of the follow-up data and the multiple splitting for time-dependent variables was performed using the `popEpi` package of the R software.

2.3 A functional regression model

The previous approach has some limitations: the cut-off of 5 years was arbitrary, and it does not take into account possible differences in exposure intensity over time. In our proposal, we substituted the covariate $I(D_i > 5)$ in (1) with a functional covariate $D_i(t)$ in the following way,

$$\frac{g(\mu_i)}{PY_i} = \beta_0 + \beta_1 A_i + \beta_2 P_i + \int_a^b \gamma(t)\, D_i(t)dt$$

where $D_i(t) = \sum_{k=1}^{N_w} I_{\left(s_{i,k},e_{i,k}\right)}(t)$ is a rectangular pulse function that identifies N_w work periods for the w-th worker during the temporal window between (a, b). As previously set, each work period k is defined by a starting time $s_{i,k}$ and an ending time $e_{i,k}$, with $a \leq s_{i,k} < e_{i,k} \leq b$. The function $\gamma(t)$ modulates the unknown risk exposure function (see Figure 2).

We adopt a flexible specification of $\gamma(t)$ by means of a cubic B-spline basis $\{\phi_m(t), m = 1,\ldots,M\}$, so that the functional terms can be simplified into a weighted sum of terms, namely

$$\int_a^b \gamma(t)\, D_i(t)dt = \sum_{k=1}^{N_w} \int_{s_{i,k}}^{e_{i,k}} \gamma(t)dt = \sum_{k=1}^{N_w} \int_{s_{i,k}}^{e_{i,k}} \sum_{m=1}^{M} \gamma_m \phi_m(t)dt =$$

$$= \sum_{k=1}^{N_w} \sum_{m=1}^{M} \gamma_m \int_{s_{i,k}}^{e_{i,k}} \phi_m(t)dt = \sum_{k=1}^{N_w} \sum_{m=1}^{M} \gamma_m \phi_{i,k,m}$$

where $\phi_{i,k,m} = \int_{s_{i,k}}^{e_{i,k}} \phi_m(t)dt$. The model is defined in the following way

$$\frac{g(\mu_i)}{PY_i} = \beta_0 + \beta_1 A_i + \beta_2 P_i + \sum_{k=1}^{N_w} \sum_{m=1}^{M} \gamma_m \phi_{i,k,m}.$$

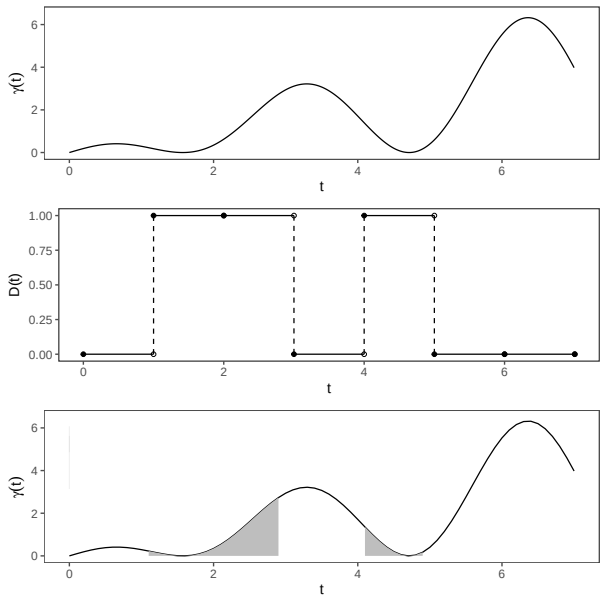

Figure 2: Graphical representation of the exposure risk function $\gamma(t)$ (top), the pulse rectangular function $D(t)$ indicating the working periods (middle), and their combination (bottom) with the grey area that reports the individual exposure (bottom).

The functional term $\int_a^b \gamma(t)\ D_i(t)dt = \sum_{k=1}^{N_w}\sum_{m=1}^{M}\gamma_m\phi_{i,k,m}$ can vary according to the exposure working period; its exponential transformation can be expressed as a functional MRR.

As presented in (2), we considered the time-dependent variant of the functional regressor

$$\sum_{k=1}^{N_w}\sum_{m=1}^{M}\gamma_m\phi^*_{i,k,m},$$

recomputing the integral $\phi^*_{i,k,m} = \int_{s_{i,k}}^{e^*_{i,k}}\phi_m(t)dt$, with $e^*_{i,k} = \max(\min(e_{i,k}, P_i), s_{i,k})$.

In order to regularize model parameters by shrinking the regression coefficients, we implemented in the previous model a Ridge penalization procedure [2].

The vector of parameters $\theta = (\beta_0, \beta_1,\ \beta_2,\ \gamma_1, \ldots, \gamma_M)$ were estimated by minimizing the penalized log-likelihood

$$\frac{1}{N}\sum_{i=1}^{N} l\left(y_i, \mu_i\right) + \lambda||\theta||_2,$$

where $l\left(y_i, \mu_i\right)$ is the negative Poisson log-likelihood contribution for the *i-th* observation, identified by the individual age and period split, while the parameter λ is the penalization parameter applied to the L^2 norm of the vector of coefficients θ. The value of λ was chosen to minimize the Generalized Cross Validation (GCV) index [12]. The number of splines was set to six in order to cover the entire temporal temporal window from 1960 to 1990 with equally spaced nodes. 95% CI was obtained by nonparametric bootstrap resampling of 5000 data sets. We used the R package *cubicSplines* to compute basis function integrals, while the package *glmnet* implemented the minimization.

3 Results

The inclusion criteria were met by 1,892 dockworkers, 22.7% of whom were already employed by the company at the time of enrollment in 1960; the majority were hired in the years just before (63% in 1960-1970) or during the period of peak asbestos unloading in the port of Venice (11% in 1970-1980), with very few starting employment after (2.5%). A minority of workers stopped working at the port before 1970 (14%) or during 1970-1980 (11%), while most workers stopped working for the company after 1980 (75%). Cohort members accumulated a total of 75,082 person-years of follow-up through December 31, 2019. 1,189 subjects (62.8%) had died by the end of follow-up; the cause of death was unknown for 44 of these workers (3.7%). The median follow-up was 46 years (IQR: 36-51).

		Lung cancer		
Work duration in 1969-1977	**Deaths, n**	**MRR**	**95% CI**	***P*-value**
No work	24	Ref.		
(0, 5) years	26	1.46	0.82-2.62	0.200
$\geq$ 5 years	136	1.76	1.05-3.07	0.038
		Pleural cancer		
Work duration in 1969-1977	**Deaths, n**	**MRR**	**95% CI**	***P*-value**
< 5 years	3	Ref.		
$\geq$ 5 years	28	2.50	0.87-10.6	0.136

Table 1: Number of deaths for lung and pleural cancer and adjusted mortality rate ratio (MRR) with 95% confidence intervals (CI) by length of employment in the period of peak asbestos exposure (1969-1977).

LC mortality showed increased risks with increasing exposure (Table 1). Among PC cases, with a reference category of <5 years of work in the period 1969-1977, a longer duration was associated with an increased risk of death from PC, although not statistically significant (MRR: 2.50, 95%CI: 0.87-10.6).

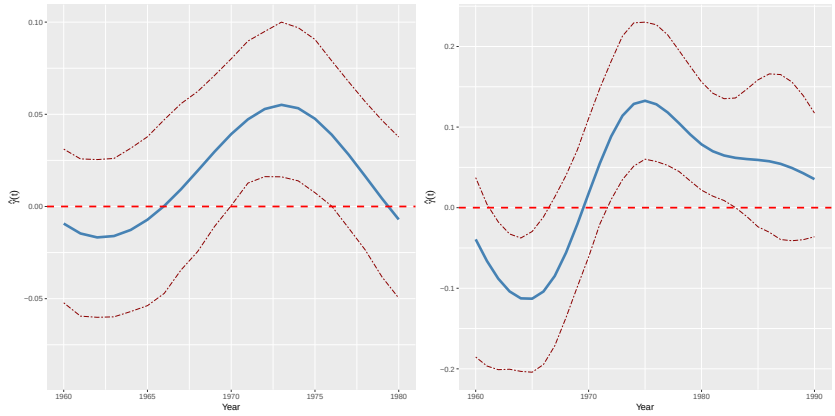

Figure 3: Estimated functional component $\gamma(t)$ for lung cancer (left) and pleural cancer (right) by year of employment (t) (blue) and its relative 95% IC (dotted dark red).

Considering the regression functional model results, the estimated risk function for lung cancer mortality exhibited a peak in the year 1973 (Figure 3 - left): looking at the bootstrap 95% IC, the risk became statistically significant in the period 1971-1976. Considering the estimated mortality function risk for

lung cancer, the functional MRR estimated by the model in the time window 1969-1977 was equal to 1.44. A similar pattern was reported by the mortality for pleural cancer (Figure 3 - right), but with higher risk estimates: the highest MM mortality risk was estimated in 1974 with a 95%CI higher than the null effect for a period more extended than the LC one which started from 1972 and ended on 1982. The estimated functional risk function produced an MRR equal to 2.05 for the period 1969-1977.

4 Discussion

Our approach allowed us to estimate the mortality risk of pleural and lung cancer without any a priori assumptions. We found that the risk exposure curve follows the amount of asbestos handled at the shipyard for both asbestos-related diseases considered. Further developments are needed to account for the latency between cumulative exposure and death and to add a spatial layer in the context of spatio-temporal exposures.

References

[1] John T Hodgson and Andrew Darnton. The quantitative risks of mesothelioma and lung cancer in relation to asbestos exposure. *Annals of occupational hygiene*, 44(8):565–601, 2000.

[2] AE Horel. Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58:54–59, 1962.

[3] A Lacourt, C Gramond, P Rolland, S Ducamp, S Audignon, P Astoul, A Gilg Soit Ilg, M Rinaldo, C Raherison, F Galateau-Salle, et al. Occupational and non-occupational attributable risk of asbestos exposure for malignant pleural mesothelioma. *Thorax*, 69(6):532–539, 2014.

[4] Virissa Lenters, Roel Vermeulen, Sies Dogger, Leslie Stayner, Lützen Portengen, Alex Burdorf, and Dick Heederik. A meta-analysis of asbestos and lung cancer: is better quality exposure assessment associated with steeper slopes of the exposure–response relationships? *Environmental Health Perspectives*, 119(11):1547–1555, 2011.

[5] Dana Loomis, Neela Guha, Amy L Hall, and Kurt Straif. Identifying occupational carcinogens: an update from the iarc monographs. *Occupational and environmental medicine*, 75(8):593–603, 2018.

[6] Corrado Magnani, Bice Fubini, Dario Mirabelli, Pier Alberto Bertazzi, Claudio Bianchi, Elisabetta Chellini, Valerio Gennaro, Alessandro Marinaccio, Massimo Menegozzo, Enzo Merler, et al. Pleural mesothelioma: epidemiological and public health issues. report from the second italian consensus conference on pleural mesothelioma. *La Medicina del lavoro*, 104(3):191–202, 2013.

[7] Jeffrey S Morris. Functional regression. *Annual Review of Statistics and Its Application*, 2:321–359, 2015.

[8] Lene Snabe Nielsen, Jesper Baelum, Jesper Rasmussen, Søren Dahl, Karen Ege Olsen, Maria Albin, Niels Christian Hansen, and David Sherson. Occupational asbestos exposure and lung cancer—a systematic review of the literature. *Archives of environmental & occupational health*, 69(4):191–206, 2014.

[9] James O Ramsay and Bernard W Silverman. *Fitting differential equations to functional data: Principal differential analysis*. Springer, 2005.

[10] Patricia Stewart. Challenges to retrospective exposure assessment. *Scandinavian journal of work, environment & health*, pages 505–510, 1999.

[11] Patricia A Stewart and Robert F Herrick. Issues in performing retrospective exposure assessment. *Applied Occupational and Environmental Hygiene*, 6(6):421–427, 1991.

[12] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.