



OPEN ACCESS

ORIGINAL ARTICLE

Multidimensional analysis of the effect of occupational exposure to organic solvents on lung cancer risk: the ICARE study

Francesca Mattei,¹ Silvia Liverani,^{2,3} Florence Guida,⁴ Mireille Matrat,^{1,5} Sylvie Cenée,¹ Lamiae Azizi,⁶ Gwenn Menvielle,^{7,8} Marie Sanchez,¹ Corinne Pilorget,^{9,10} Bénédicte Lapôtre-Ledoux,¹¹ Danièle Luce,^{12,13} Sylvia Richardson,³ Isabelle Stücker,¹ ICARE Study Group

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/oemed-2015-103177>).

For numbered affiliations see end of article.

Correspondence to

Francesca Mattei, CESP Centre for Research in Epidemiology and Population Health, 16 avenue Paul Vaillant Couturier, Villejuif Cedex 94807, France; francesca.mattei@inserm.fr

SR and IS equally contributed.

Received 9 July 2015

Revised 22 January 2016

Accepted 5 February 2016

Published Online First

24 February 2016



Open Access
Scan to access more
free content

ABSTRACT

Background The association between lung cancer and occupational exposure to organic solvents is discussed. Since different solvents are often used *simultaneously*, it is difficult to assess the role of individual substances.

Objectives The present study is focused on an in-depth investigation of the potential association between lung cancer risk and occupational exposure to a large group of organic solvents, taking into account the well-known risk factors for lung cancer, tobacco smoking and occupational exposure to asbestos.

Methods We analysed data from the Investigation of occupational and environmental causes of respiratory cancers (ICARE) study, a large French population-based case-control study, set up between 2001 and 2007. A total of 2276 male cases and 2780 male controls were interviewed, and long-life occupational history was collected. In order to overcome the analytical difficulties created by multiple correlated exposures, we carried out a novel type of analysis based on Bayesian profile regression.

Results After analysis with conventional logistic regression methods, none of the 11 solvents examined were associated with lung cancer risk. Through a profile regression approach, we did not observe any significant association between solvent exposure and lung cancer. However, we identified clusters at high risk that are related to occupations known to be at risk of developing lung cancer, such as painters.

Conclusions Organic solvents do not appear to be substantial contributors to the occupational risk of lung cancer for the occupations known to be at risk.

INTRODUCTION

Lung cancer has been and remains the most frequent cancer and the most common cause of death from cancer worldwide, especially in men. The French Institute for Public Health Surveillance (InVS) estimated around 40 000 new lung cancer cases in 2012 in France, representing the first cause of cancer mortality in men and the second in women.¹ Although tobacco smoking is recognised as the leading cause of lung cancer,² occupational exposures contribute to the burden of the disease, with an attributable fraction to these multiple risk factors estimated to be in the range 13–29% in French men.³

What this paper adds

- The role of some organic solvents in some cancers has been established, but in lung cancer it is still the object of discussion.
- The multiple highly correlated exposures in our data made difficult the interpretation of traditional analyses results.
- We proposed a new Bayesian approach allowing us to take into account the whole set of solvents as well as confounding factors, for example, tobacco smoking and occupational exposure to asbestos.
- Organic solvents did not appear to be substantial contributors to the occupational risk of lung cancer.

The presence of multiple exposure patterns is commonly encountered in occupational epidemiological studies investigating disease aetiology. The large number of substances which are typically present in working environments, and their interdependence do not easily allow for isolation of the causative agent contributing to the development of pathologies, such as cancer. A typical case is that of painters and rubber production industry workers who are exposed to multiple agents. Working in these occupations is considered to be a risk factor for lung cancer.⁴ However, the specific agents that potentially contribute to this risk have not been identified yet. For diseases such as lung cancer, an additional source of complexity for teasing out the role of occupational exposures is the need to account for their well-known set of non-occupational risk factors, the most predominant being tobacco smoking.

Millions of workers are exposed to organic solvents due to their wide application in practically all branches of modern industry, such as the manufacture of surface coating (paints, varnishes and printing inks), the manufacture of synthetic fibres, the cleaning sector and the construction industry. Among these substances, benzene and trichloroethylene were classified in group 1 by the



CrossMark

To cite: Mattei F, Liverani S, Guida F, et al. *Occup Environ Med* 2016;**73**:368–377.

International Agency for Research on Cancer (IARC) because sufficient evidence exists of their association with the risk of leukaemia and cancer of kidney, respectively.^{4,5} Contrasting results were found about their association with lung cancer. Indications for a positive association for benzene were shown in a few cohort studies^{6–8} and in studies on animals.^{9–14} Moderate increased risks were also identified in cohorts of female dry-cleaners,¹⁵ and in a few case–control studies on the exposure to tetrachloroethylene,^{15–17} including a previous analysis performed in our population.¹⁸

Since various solvents are often used *simultaneously*, and sometimes as components of solvent mixtures, it is difficult to assess the role of individual substances. For this reason, this study is focused on an in-depth investigation of the potential association between lung cancer and occupational exposure to a large group of organic solvents. In order to overcome the analytical difficulties created by the presence of multiple correlated exposures, we followed an analytic strategy, which complements the traditional analysis approach, based on stepwise logistic regressions, with a Bayesian approach, based on profile regression (PR).¹⁹

METHODS

Study design and population

The Investigation of occupational and environmental causes of respiratory cancers (ICARE) study is a multicentre, population-based, case–control study of respiratory cancers, conducted from 2001 to 2007 in 10 French *départements* (ie, administrative regions) including a cancer registry.²⁰ Eligible cases were participants aged <76 years, who were newly diagnosed with primary cancer of the lung and upper aerodigestive tract during the study period. All histologically confirmed lung cancer types were considered (codes C33 and C34 International Classification of Diseases for Oncology (ICDO) 3rd edition²¹). Controls, with no history of previous respiratory cancer, were randomly selected from the general population through incidence density sampling, in the same departments as cases. Controls were frequently matched to cases by gender and age (<40, 40–54, 55–64, ≥65 years). An additional stratification was used to achieve a distribution by socioeconomic status comparable to that of the general population living in the *départements*.

Of the 4865 eligible cases and 4673 eligible controls identified, 6481 participants participated in the study: 2926 cases (2276 men and 650 women) and 3555 controls (2780 men and 775 women). For more details about study design and participants' selection.²⁰

Ethics approval was obtained from the Institutional Review Board of the French National Institute of Health and Medical Research (IRB-Inserm, number 01-036 and CNIL number 90120).

Data collection

Trained interviewers collected data during face-to-face interviews using standardised questionnaires, which included information about sociodemographic characteristics, residential history, medical history, familial history of cancer, detailed information about tobacco, alcohol and non-alcoholic beverage consumption, and detailed lifetime occupational history (about all jobs held for at least 1 month). Job titles and industrial activities were coded blinded to case–control status according to the International Standard Classification of Occupation (ISCO), 1968 revision,²² and the French Nomenclature of Activities (NAF), 1999 edition.²³

A shortened version of the questionnaire was submitted to proxy respondents and participants too sick or too tired (5% of

men and 3% of women), and mainly to collect smoking data and occupational history.

Study agents and exposure assessment

Substances of interest were several organic solvents or families of solvents, which can be grouped into three main categories: chlorinated solvents, fuels and petroleum-based solvents, and oxygenated solvents.

We applied different job-exposure matrices (JEM),^{24–27} developed by the InVS, to assess separately the exposure to each substance (or family or group of substances): tetrachloroethylene, trichloroethylene, methylene chloride, chloroform, carbon tetrachloride, benzene, kerosene/diesel oil/fuel oils, mineral spirits and other light aromatic fractions (white spirits), special boiling point spirits and other aliphatic petroleum-based solvents (SBPs), gasoline, alcohol, ketones/esters, diethyl ether, ethylene glycol and tetrahydrofuran.

For each combination of ISCO and NAF code, the JEM assigned three indices of exposure: (1) probability, expressed by the percentage of exposed workers, (2) intensity, and (3) frequency, given by the proportion of exposed working time (see eTable 1). Information on the duration of each job was collected in the questionnaire.

Asbestos exposure was assessed using a specific JEM,²⁴ following the same methodological procedure described for solvents exposure. In this case, JEM provided two indices for frequency and intensity of exposure according to whether the exposure was due to the working environment or to the specific job tasks.

To account for changes in exposure over time, indices were provided for different calendar periods from 1950 to 2007 for chlorinated solvents, from 1947 to 2005 for fuels and petroleum-based solvents, from 1950 to 2012 for oxygenated solvents, and from 1945 to 2007 for asbestos. We used the indices corresponding to the earliest period of each JEM for job periods starting or holding before these dates (3.3%).

In order to improve specificity of JEMs, we assigned exposure to solvents and asbestos only to those jobs which had probability of exposure greater than a cut-off, chosen according to the categories of JEMs (30% for asbestos, chlorinated and oxygenated solvents, and 50% for fuels and petroleum-based solvents, eTable 1). The jobs below the cut-off were considered as non-exposed and not included in further exposure classification. For all the exposed jobs, we computed the Cumulative Exposure Index (CEI) by summing the product of the duration of each job period, over the entire work history, and weights for exposure probability, frequency and intensity. For asbestos exposure, CEI was calculated taking into account frequency and intensity either linked to working environment and job tasks.²⁸

Statistical analysis

Results are presented for men only, because the small number of exposed women precluded analyses for most solvents. We restricted the study population to participants with a known working history (11 men excluded). Analyses were performed on substances to which at least 15 participants were exposed; thus, carbon tetrachloride, chloroform, tetrachloroethylene and methylene chloride were excluded. CEIs were transformed into categorical variables according to tertiles of controls distribution. This allowed for the definition of low, medium and high levels of exposure for each substance.

We used Spearman correlations (r , correlation coefficient) to assess correlation among substances, including asbestos. We estimated ORs of lung cancer and their corresponding 95% CIs

using multivariable unconditional logistic regression models. These were performed including both all exposure substances and those selected by stepwise forward and backward procedure. In order to take into account the degree of multicollinearity present in our data and to produce regression coefficients which are more stable and reliable, a ridge regression model was also performed.²⁹

Analyses were systematically adjusted for age at interview (<50, 50–60, 61–70, >70 years), *départements*, exposure to asbestos³⁰ (CEI in tertiles) and cigarette smoking history, which was summarised using the Comprehensive Smoking Index (CSI).³¹ This index combines duration of smoking, time since cessation and intensity. It is null in never smokers.

We also performed PR model, a Bayesian clustering method, which partitions observations into clusters with respect to risk factor characteristics and to outcome status. Both risk factors and outcome are used jointly to form the clusters and each cluster is associated via a logistic link to the outcome of interest.³² The number of clusters is not fixed in advance but explored throughout the algorithm. Postprocessing of the output produces a 'representative clustering', which is an appropriate summary of all clustering explored.³³ In particular, the representative clusters together with 95% credibility intervals for the associated logOR with respect to a baseline cluster are given. An estimation of the cluster-specific proportions for each category of exposure, the 'exposure profile of the cluster', is also given, and it is useful for understanding the main characteristics of each cluster. For example, some clusters may be characterised by a high probability of exposure to benzene and gasoline and a much smaller probability of other exposures, while for a different cluster other solvent combinations are predominant. Additionally, the PR model and the Bayesian output allow the carrying out of predictive inference and quantification of the range of risk associated to prespecified combinations of covariates of interest. In our data, clustering was based on CEI of solvents (4 categories) and CEI of asbestos (4 categories). LogORs were adjusted for age at interview (in classes), *départements* and CSI (as continuous variable). To quantify the role of a group of covariates, we specified a number of simulated predictive scenarios, called pseudoprofiles, which are used to compare and contrast risk predictions under different patterns of exposure.³⁴

To study the robustness to the chosen cut-off for the probability of exposure to solvents, sensitivity analyses were performed on different cut-off points (20% and 40% for chlorinated and oxygenated solvents, and 10% for fuels and petroleum-based solvents). The categories of probability of exposure to asbestos did not allow sensitivity analysis to be performed.

All analyses were implemented in R (V3.1.2). We used package 'Hmisc' (V3.15–0) for Spearman correlations; package 'MASS' (V7.3–33) for stepwise logistic regressions, based on Akaike information criterion (AIC); package 'ridge' (V2.1–3) and package 'PRemiuM'³³ (V3.0.32) for PR.

All p values were two-sided, and a p value ≤ 0.05 was the threshold for statistical significance. For the Bayesian analyses, we report 95% credibility intervals.

RESULTS

Population and substances description

The main sociodemographic characteristics of the study population and the histological subtypes of lung cancer cases are presented in table 1. Mean age was slightly higher for cases than for controls. As expected, a statistically significantly higher

proportion of blue-collar workers was found among cases as compared with controls.

Squamous cell carcinomas and adenocarcinoma were the most frequently diagnosed cancer types (35% and 34%, respectively), and small cell carcinomas represented 15% of cases.

Smoking habits data were known for 5012 men (2241 cases), among whom 2.6% were non-smokers' cases (59 cases). As expected, a strong association emerged between lung cancer and CSI, with a clear dose–response effect (p value trend <0.001) (see eTable 2).

Occupational exposures to solvents and asbestos were correlated (see eFigure 1). Particularly strong and statistically significant ($p < 0.001$) were correlations between exposures to ethylene glycol and gasoline ($r = 0.78$), benzene and gasoline ($r = 0.62$), benzene and ethylene glycol ($r = 0.60$), and benzene and white spirits ($r = 0.58$).

Logistic regression analysis

First, we incorporated into the logistic regression models the exposure to one solvent at a time (table 2). Statistically and borderline statistically significant associations emerged only for medium and high levels of white spirits exposure ($OR_{\text{medium}} = 1.66$, 95% CI 1.12 to 2.46 and $OR_{\text{high}} = 1.36$, 95% CI 0.92 to 2.02) but no dose–response relationship was observed. When we included all solvents in the model, the regression did not show any association with lung cancer. We also applied stepwise forward and backward approaches. These selected the same solvents for the final model with 95% CI that were very similar to those obtained from the univariate analysis, namely: (1) white spirits, which are statistically significantly associated with lung cancer at medium level of exposure ($OR_{\text{medium}} = 1.67$, 95% CI 1.13 to 2.48) and borderline significantly associated at high level ($OR_{\text{high}} = 1.40$, 95% CI 0.94 to 2.09); (2) tetrahydrofuran, which was not significantly associated. We finally estimated ORs using ridge regression, which typically produces ORs that are shrunk as compared with the ordinary least squares estimates. For the model including all solvents, the highest ORs were observed for medium and high exposure to white spirits, and medium exposure to SBPs and tetrahydrofuran (table 2).

PR analysis

Postprocessing of PR output produced a representative clustering dividing the entire male sample into 13 clusters. Each of them was characterised by a specific exposure profile summarised in table 3, which shows the posterior mean of the associated exposure probabilities.

For each cluster, we highlighted in bold the modal probability for every substance, for example, in cluster 6, 46% of participants were exposed to a high level of alcohol, 74% to a high level of ketones/esters and no other solvents; or in cluster 9, 49% were exposed to a low level of benzene, 48% to a medium level of white spirits, 52% to a low level of alcohol and 49% to a medium level of ketones/esters.

We assumed cluster 5 as the baseline for logORs estimation, as it is made up of non-exposed participants. All clusters have been presented in table 3 ranked in ascending order of logOR.

Figure 1 shows box plots of the distribution for logORs for each cluster (relative to the non-exposed cluster 5), with their corresponding 95% credible intervals (CI*). The horizontal line represents the average logOR in the whole population, obtained as a weighted average among clusters. We call it the logOR mean. We chose to display this as we want to use the clustering as a guide to uncover subgroups of participants who are clearly

Table 1 Sociodemographic characteristics among cases and controls and distribution of histological subtypes of lung cancer cases, men, France, 2001–2007

Variables	Cases (n=2260*) n (%)	Controls (n=2780) n (%)	OR†	95%CI
Age				
<50	279 (12.3)	636 (22.9)	Ref.	–
50–60	837 (37.0)	937 (33.7)	2.04	1.72 to 2.42
61–70	814 (36.0)	911 (32.8)	2.04	1.72 to 2.42
>70	330 (14.6)	296 (10.6)	2.55	2.06 to 3.15
Mean (SD)	60 (9.1)	58 (9.9)	p<0.001‡	
Socioeconomic status (the longest duration)				
Farmers	69 (3.1)	168 (6.0)	0.75	0.55 to 1.03
Self-employed workers	139 (6.2)	152 (5.5)	1.65	1.26 to 2.15
Managers	238 (10.5)	544 (19.6)	0.78	0.64 to 0.96
Intermediate white-collar workers	314 (13.9)	564 (20.3)	Ref.	–
Office and sales employees	218 (9.6)	297 (10.7)	1.32	1.06 to 1.65
Blue-collar workers	1267 (56.1)	1053 (37.9)	2.18	1.86 to 2.56
Missing	15 (0.7)	2 (0.1)		
Lung cancer histology				
Squamous	787 (34.8)	–		
Small cell	333 (14.7)	–		
Adenocarcinoma	778 (34.4)	–		
Non-small cell	199 (8.8)	–		
Other	147 (6.5)	–		
Multiple phenotype§	16 (0.7)	–		

*Participants with at least one known exposure were kept (5 cases were excluded).

†OR adjusted for age at interview and *département*.

‡p Value derived from Student t test (two sided).

§Multiple phenotype corresponds to tumours with several histological entities.

different both in terms of their exposure and their risk from the rest of the sample. Hence, we have coloured in grey what we call ‘baseline subgroups’, baseline in so far as their 95% CI* logORs contained the logOR mean representing the average logOR of the sample. We do not interpret in detail the profiles of baseline clusters.

Only clusters 11 and 13 (displayed in white in figure 1) have logORs significantly higher than the logOR mean, meaning that the association with lung cancer was significantly stronger in these clusters than in the others. The highest risk in cluster 13 was composed of participants characterised by exposure to high levels of asbestos and white spirits, and a low level of benzene. Cluster 11 was the largest cluster, made up of participants with a high level of asbestos exposure and low level of ketones/esters. The low number of participants composing cluster 12 did not allow a stable logOR estimate to be obtained (table 3).

We also cross-classified clusters according to their participants’ jobs. The job distribution of cluster 13 encompasses a large proportion of painters in the construction field (35%). In cluster 11, plumbers represented 14%, and carpenters and joinery 13% and sheet-metal workers 4% of all jobs grouped in that cluster (see eFigure 2).

Since the exposure to white spirits and benzene was common to high-risk clusters, we were interested to understand if and how the exposure to these substances affected the risk. With this purpose, we set up several pseudoprofiles of covariate patterns.

Figure 2 shows the posterior predictive density of logORs (relative to pseudoprofile 1) for different combinations of CEI of white spirits and benzene levels. CEI of asbestos and other

solvents reflected the covariate patterns present in the main sample (for technical details, Ref. 34).

These plots allow one to visually understand how logOR changes as we altered these covariates from low to high exposures. We can observe a small non-significant shift to the right of logOR for the predicted risks, with wide credible intervals.

DISCUSSION

ICARE is one of the largest case–control studies investigating the role of occupational risk factors in respiratory cancer that collects detailed information about lifelong working history. Participants are exposed to many substances during their working life, and in this study we focused on investigating the role of several solvents on lung cancer.

The study was carried out in collaboration with the French network of cancer registries, allowing us to recruit cancer cases in almost all healthcare facilities in the included *départements*. The frequency matching strategy allowed for selection of controls comparable to cases with respect to gender distribution. Whereas the unique control group assumed for the two groups of cancer cases (lung and upper aerodigestive tract) encompassed a slight difference in terms of age, the large number of participants in each age group allowed for satisfactory adjustment with respect to age.

Thanks to the frequency matching, controls were comparable to cases with respect to age and gender distribution. Controls were randomly selected in the same *département* as cases, through incidence density sampling, and according to the socio-economic distribution of the corresponding general population.

Table 2 Association between lung cancer and exposure to organic solvents*, men, France, 2001–2007

CEI	Cases	Controls	OR†	95% CI	Cases	Controls	OR‡	95% CI	OR§	95% CI	Ridge regression OR¶
TCE											
Not exposed	2183	2717	Ref.	–	2165	2711	Ref.	–	Not selected	–	Ref.
Low	14	17	1.04	0.45 to 2.39	14	17	0.83	0.34 to 2.02	–	–	0.94
Medium	14	16	1.53	0.64 to 3.66	14	16	0.95	0.37 to 2.48	–	–	1.08
High	22	19	1.15	0.55 to 2.41	21	19	0.91	0.43 to 1.94	–	–	1.01
BZ											
Not exposed	1892	2454	Ref.	–	1856	2444	Ref.	–	Not selected	–	Ref.
Low	108	103	1.29	0.92 to 1.80	108	102	1.01	0.59 to 1.71	–	–	1.11
Medium	106	103	1.14	0.80 to 1.60	105	102	0.86	0.45 to 1.65	–	–	1.03
High	129	106	0.97	0.70 to 1.35	126	106	0.95	0.44 to 2.04	–	–	1.06
WS											
Not exposed	1994	2577	Ref.	–	1953	2563	Ref.	–	Ref.	–	Ref.
Low	52	65	0.92	0.58 to 1.45	51	65	0.89	0.52 to 1.51	0.94	0.59 to 1.48	0.95
Medium	87	62	1.66	1.12 to 2.46	86	62	1.53	0.87 to 2.67	1.67	1.13 to 2.48	1.29
High	106	66	1.36	0.92 to 2.02	105	64	1.52	0.77 to 3.01	1.40	0.94 to 2.09	1.25
KDF											
Not exposed	1925	2421	Ref.	–	1894	2410	Ref.	–	Not selected	–	Ref.
Low	104	116	0.97	0.69 to 1.35	103	116	0.98	0.69 to 1.38	–	–	0.97
Medium	99	111	1.14	0.81 to 1.60	98	111	1.16	0.80 to 1.67	–	–	1.07
High	104	117	0.77	0.55 to 1.07	100	117	0.79	0.48 to 1.29	–	–	0.89
SBPs											
Not exposed	2168	2708	Ref.	–	2129	2695	Ref.	–	Not selected	–	Ref.
Low	19	20	1.16	0.54 to 2.49	19	20	1.24	0.53 to 2.92	–	–	1.04
Medium	25	19	1.78	0.90 to 3.51	25	19	1.55	0.70 to 3.46	–	–	1.24
High	22	20	1.33	0.61 to 2.89	22	20	1.13	0.46 to 2.76	–	–	1.07
GL											
Not exposed	2043	2554	Ref.	–	2012	2543	Ref.	–	Not selected	–	Ref.
Low	57	71	0.75	0.49 to 1.17	57	71	0.74	0.45 to 1.21	–	–	0.87
Medium	61	68	1.11	0.73 to 1.70	61	68	1.23	0.55 to 2.75	–	–	1.06
High	67	72	0.68	0.45 to 1.03	65	72	1.10	0.43 to 2.82	–	–	0.94
Alc											
Not exposed	1895	2380	Ref.	–	1866	2369	Ref.	–	Not selected	–	Ref.
Low	121	128	1.10	0.80 to 1.51	118	128	1.00	0.71 to 1.40	–	–	1.02
Medium	91	127	0.90	0.63 to 1.27	90	126	0.91	0.60 to 1.37	–	–	0.93
High	122	131	1.23	0.90 to 1.68	121	131	1.26	0.85 to 1.88	–	–	1.09
KetEst											
Not exposed	1786	2313	Ref.	–	1758	2301	Ref.	–	Not selected	–	Ref.
Low	163	150	1.21	0.91 to 1.61	161	150	1.28	0.91 to 1.80	–	–	1.16
Medium	149	150	1.16	0.86 to 1.56	146	150	1.32	0.92 to 1.90	–	–	1.16
High	132	154	0.89	0.66 to 1.20	130	153	0.81	0.55 to 1.20	–	–	0.93
Et₂O											
Not exposed	2208	2715	Ref.	–	2172	2706	Ref.	–	Not selected	–	Ref.
Low	11	16	1.02	0.40 to 2.61	10	16	1.04	0.38 to 2.85	–	–	1.00
Medium	8	16	0.80	0.29 to 2.18	8	16	0.73	0.26 to 2.06	–	–	0.85
High	5	17	0.38	0.12 to 1.28	5	16	0.36	0.11 to 1.25	–	–	0.59
Glycol											
Not exposed	2079	2586	Ref.	–	2046	2576	Ref.	–	Not selected	–	Ref.
Low	55	59	1.12	0.71 to 1.76	55	59	0.98	0.49 to 1.97	–	–	1.04
Medium	54	58	0.91	0.57 to 1.44	53	58	0.74	0.38 to 1.46	–	–	0.94
High	43	61	0.57	0.35 to 0.92	41	61	0.63	0.27 to 1.48	–	–	0.82
THF											
Not exposed	2202	2735	Ref.	–	2167	2725	Ref.	–	Ref.	–	Ref.
Low	3	10	0.22	0.05 to 0.94	3	10	0.19	0.04 to 0.79	0.24	0.06 to 1.00	0.44
Medium	13	9	2.07	0.71 to 5.99	13	9	1.71	0.56 to 5.20	2.22	0.76 to 6.44	1.60
High	12	10	1.37	0.44 to 4.21	12	10	1.14	0.36 to 3.65	1.47	0.47 to 4.53	1.26

*All models were systematically adjusted for age, départements, CSI and CEI of asbestos.

†Logistic regression model performed separately for each substance.

‡Logistic regression model performed on all substances at a time.

§Logistic regression model performed respectively with stepwise backward and stepwise forward variable selection.

¶Ridge regression estimates, corresponding to the optimal λ equal to 598.97 obtained using cross-validation.³⁵ We did not penalise the adjusting variable coefficients.

Alc, alcohol; BZ, benzene; CEI, Cumulative Exposure Index; CSI, Comprehensive Smoking Index; Et₂O, diethyl ether; GL, gasoline; Glycol, ethylene glycol; High, CEI > 2nd tertile; KDF, kerosene/diesel oil/fuel oils; KetEst, ketones/esters; Low, CEI ≤ 1st tertile; Medium, 1st tertile < CEI ≤ 2nd tertile; SBPs, special boiling point spirits and other aliphatic petroleum-based solvents; TCE, trichloroethylene; THF, tetrahydrofuran; WS, white spirits.

Table 3 Results of profile regression analysis*,† showing the covariate profile of each cluster and the resulting logOR associated with each cluster, men, France, 2001–2007

Cluster		1	2	3	4	5	6	7	8	9	10	11	12	13
Number of participants		162	97	46	124	3760	136	149	89	56	38	194	22	139
LogOR		−0.20	−0.16	−0.09	−0.08	0.00	0.05	0.17	0.32	0.51	0.51	0.55	0.69	0.74
95% Credible interval		(−0.56 to 0.16)	(−0.62 to 0.32)	(−0.78 to 0.64)	(−0.5 to 0.33)	(0.00 to 0.00)	(−0.36 to 0.46)	(−0.22 to 0.56)	(−0.19 to 0.83)	(−0.06 to 1.11)	(−0.26 to 1.28)	(0.2 to 0.91)	(−0.35 to 1.71)	(0.32 to 1.16)
Asb	Not exposed	0.74	0.34	0.66	0.03	0.69	0.50	0.06	0.37	0.32	0.83	0.14	0.52	0.04
	Low	0.10	0.26	0.11	0.06	0.10	0.14	0.50	0.12	0.25	0.11	0.11	0.22	0.09
	Medium	0.09	0.19	0.09	0.62	0.10	0.21	0.27	0.22	0.23	0.03	0.22	0.09	0.20
	High	0.07	0.21	0.14	0.28	0.12	0.15	0.17	0.29	0.20	0.03	0.53	0.17	0.67
TCE	Not exposed	0.95	0.97	0.73	0.97	0.99	0.94	0.95	0.89	0.87	0.42	0.97	0.90	0.97
	Low	0.03	0.01	0.22	0.01	0.00	0.01	0.02	0.03	0.04	0.08	0.01	0.03	0.01
	Medium	0.01	0.01	0.03	0.01	0.00	0.03	0.01	0.02	0.06	0.33	0.01	0.03	0.01
	High	0.01	0.01	0.03	0.01	0.01	0.02	0.02	0.05	0.03	0.16	0.01	0.04	0.01
BZ	Not exposed	0.98	0.87	0.15	0.01	1.00	0.95	0.03	0.03	0.30	0.02	0.97	0.06	0.08
	Low	0.01	0.09	0.66	0.02	0.00	0.03	0.18	0.37	0.49	0.36	0.02	0.09	0.39
	Medium	0.01	0.02	0.14	0.14	0.00	0.01	0.51	0.23	0.18	0.47	0.01	0.31	0.32
	High	0.00	0.02	0.05	0.83	0.00	0.01	0.28	0.36	0.03	0.14	0.01	0.54	0.21
GL	Not exposed	0.98	0.71	0.91	0.01	0.99	0.86	0.02	0.87	0.94	0.92	0.91	0.07	0.96
	Low	0.01	0.24	0.04	0.02	0.01	0.13	0.21	0.07	0.03	0.02	0.07	0.08	0.03
	Medium	0.00	0.03	0.02	0.08	0.00	0.01	0.71	0.03	0.02	0.03	0.01	0.33	0.01
	High	0.00	0.02	0.02	0.89	0.00	0.01	0.06	0.02	0.02	0.03	0.01	0.53	0.01
SBPs	Not exposed	0.97	0.97	0.11	0.97	1.00	0.96	0.96	0.75	0.92	0.03	0.98	0.74	0.97
	Low	0.01	0.01	0.50	0.02	0.00	0.01	0.02	0.03	0.04	0.03	0.01	0.04	0.01
	Medium	0.01	0.01	0.27	0.01	0.00	0.02	0.01	0.05	0.03	0.50	0.01	0.11	0.01
	High	0.01	0.01	0.12	0.01	0.00	0.01	0.01	0.17	0.02	0.43	0.01	0.11	0.01
KDF	Not exposed	0.95	0.18	0.60	0.21	0.93	0.92	0.48	0.84	0.86	0.19	0.92	0.12	0.85
	Low	0.03	0.11	0.05	0.05	0.04	0.05	0.13	0.07	0.06	0.02	0.04	0.13	0.05
	Medium	0.02	0.38	0.30	0.07	0.03	0.02	0.18	0.04	0.04	0.16	0.03	0.12	0.04
	High	0.01	0.32	0.04	0.68	0.00	0.01	0.21	0.05	0.05	0.62	0.01	0.63	0.05
WS	Not exposed	0.96	0.93	0.45	0.96	0.99	0.65	0.92	0.03	0.12	0.04	0.93	0.76	0.03
	Low	0.02	0.03	0.36	0.02	0.00	0.26	0.02	0.04	0.31	0.05	0.04	0.04	0.13
	Medium	0.01	0.02	0.15	0.01	0.00	0.07	0.05	0.28	0.48	0.55	0.02	0.11	0.26
	High	0.01	0.01	0.04	0.02	0.00	0.02	0.01	0.65	0.09	0.37	0.01	0.10	0.58
Alc	Not exposed	0.16	0.91	0.44	0.91	0.95	0.10	0.89	0.08	0.31	0.11	0.79	0.83	0.91
	Low	0.27	0.04	0.19	0.04	0.03	0.10	0.06	0.07	0.52	0.07	0.13	0.08	0.05
	Medium	0.33	0.04	0.28	0.03	0.01	0.34	0.03	0.36	0.13	0.11	0.05	0.04	0.03
	High	0.23	0.01	0.09	0.01	0.02	0.46	0.01	0.49	0.05	0.71	0.02	0.04	0.02
KetEst	Not exposed	0.71	0.76	0.54	0.06	0.96	0.08	0.16	0.05	0.20	0.86	0.16	0.72	0.90
	Low	0.11	0.14	0.08	0.17	0.02	0.06	0.51	0.02	0.21	0.05	0.49	0.09	0.03
	Medium	0.11	0.08	0.03	0.53	0.01	0.12	0.30	0.03	0.49	0.05	0.31	0.13	0.05
	High	0.07	0.01	0.36	0.24	0.01	0.74	0.03	0.89	0.10	0.05	0.03	0.06	0.02
Et₂O	Not exposed	0.71	0.97	0.94	0.97	0.99	0.95	0.98	0.96	0.92	0.91	0.98	0.86	0.98
	Low	0.09	0.02	0.02	0.01	0.00	0.02	0.01	0.01	0.04	0.02	0.01	0.07	0.01

Continued

Table 3 Continued

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of participants	162	97	46	124	3760	136	149	89	56	38	194	22	139
LogOR	-0.20	-0.16	-0.09	-0.08	0.00	0.05	0.17	0.32	0.51	0.51	0.55	0.69	0.74
95% Credible interval	(-0.56 to 0.16)	(-0.62 to 0.32)	(-0.78 to 0.64)	(-0.5 to 0.33)	(0.00 to 0.00)	(-0.36 to 0.46)	(-0.22 to 0.56)	(-0.19 to 0.83)	(-0.06 to 1.11)	(-0.26 to 1.28)	(0.2 to 0.91)	(-0.35 to 1.71)	(0.32 to 1.16)
Glycol	Medium 0.10 High 0.09 Not exposed 0.98	0.01 0.01 0.62	0.02 0.02 0.93	0.01 0.01 0.02	0.00 0.00 1.00	0.01 0.01 0.97	0.01 0.01 0.04	0.01 0.01 0.04	0.01 0.01 0.94	0.02 0.02 0.95	0.05 0.02 0.92	0.03 0.03 0.96	0.01 0.01 0.97
THF	Low 0.01 Medium 0.01 High 0.01 Not exposed 0.98	0.10 0.17 0.10 0.96	0.02 0.03 0.02 0.94	0.03 0.32 0.64 0.98	0.00 0.00 1.00	0.01 0.01 0.96	0.01 0.01 0.98	0.01 0.05 0.97	0.02 0.02 0.94	0.02 0.02 0.92	0.02 0.03 0.93	0.02 0.01 0.75	0.01 0.01 0.88
	Low 0.01 Medium 0.01 High 0.01	0.02 0.01 0.01	0.02 0.02 0.02	0.01 0.01 0.01	0.00 0.00 0.00	0.01 0.01 0.01	0.01 0.01 0.01	0.01 0.01 0.01	0.02 0.02 0.03	0.02 0.02 0.02	0.02 0.05 0.02	0.03 0.10 0.03	0.01 0.01 0.01

* All substances CEI in classes according to tertile distribution: Low, CEI ≤ 1st tertile; Medium, 1st tertile < CEI ≤ 2nd tertile; High, CEI > 2nd tertile.
 † The mode for every substance in each cluster is highlighted in bold.
 Alc, alcohol; Asb, asbestos; BEI, Benzene Exposure Index; EtO, diethyl ether; GL, gasoline; Glycol, ethylene glycol; KDF, kerosene/diesel oil/fuel oils; KetEst, ketones/esters; SBPs, special boiling point spirits and other aliphatic petroleum-based solvents; TCE, trichloroethylene; THF, tetrahydrofuran; WS, white spirits.

Owing to the large number of participants, we could not use an expert assessment based on job title history and free job description, known to be the gold standard technique, to determine past occupational exposure. Instead, we applied specific JEMs, developed at the InVS. JEMs are conceived to assign exposure in a reproducible and automatic way so that the assessment can be considered objective and independent of case-control status. Being aware of the risk of non-differential misclassification bias, we increased specificity by establishing a cut-off of 30% (or 50% as applicable) on the probability of exposure to consider a job title as 'exposed'.

As often occurs in environmental epidemiological studies, our data were characterised by multiple highly correlated exposures, which led to known problems for statistical analyses.³⁶ Fitting multiple logistic regressions models incorporating one exposure variable at a time does not allow for control of other exposure variables, which could be confounding factors.³⁷ In addition, when a study generates a large number of tests, as in our case, multiple comparison issues arise and should be addressed.³⁸

Alternatively, incorporating all exposure variables in the same regression model can achieve results difficult to interpret. In our data, no statistically significant associations emerged when all exposure variables were included in the model. When we performed a variable selection based on stepwise methods, the final model returned two substances, white spirits and tetrahydrofuran, which were respectively weakly and not significantly associated with lung cancer. It is known that in the presence of multicollinearity maximum likelihood estimates may be unstable.³⁹ Implementing Ridge regression produced lower estimates overall for most substances, highlighting again sensitivity to multicollinearity and the difficulty of interpreting logistic regression results in this case.

PR is instead a Bayesian statistical approach conceived to examine the effect of combinations of variables that structure the variability of the data, allowing one to overcome several limitations of traditional regression methods.^{19 36} An important difference between the two approaches arises in the main unit of inference: logistic regression models the risk of individual participants, while PR infers the risk of groups of participants.³²

PR divided the sample into 13 clusters, including one composed by participants never exposed to any solvent. All the other clusters were characterised by different profiles of exposure to solvents and asbestos. Only two were statistically associated with an increased lung cancer risk, clusters 11 and 13. The cluster at highest risk was composed of participants exposed to white spirits and benzene, in addition to asbestos. A pseudoprofile performed on benzene and white spirits revealed a non-significant small shift of predicted logORs with large 95% CIs*. We also reran pseudoprofiles on exposure to white spirits only. Overall, predicted logORs were higher but not statistically significant and the 95% CIs* were still large. The pseudoprofile aspect of analysis led us to conclude that neither white spirits nor benzene may have a role in lung cancer development.

In order to ensure that failure to observe any significant role of solvents was not due to our choice of cut-off, we performed a sensitivity analysis with different cut-points for the probability of exposure to each solvent. Results were similar. Overall, we did not find any evidence in support of the carcinogenicity of any of the organic solvents investigated.

To finalise our analysis, it was important to see how job titles were distributed among clusters. Although we did not observe any significant association between solvent exposure and lung cancer, clusters identified by the method were related to occupations that are known to be at risk of developing lung cancer. We

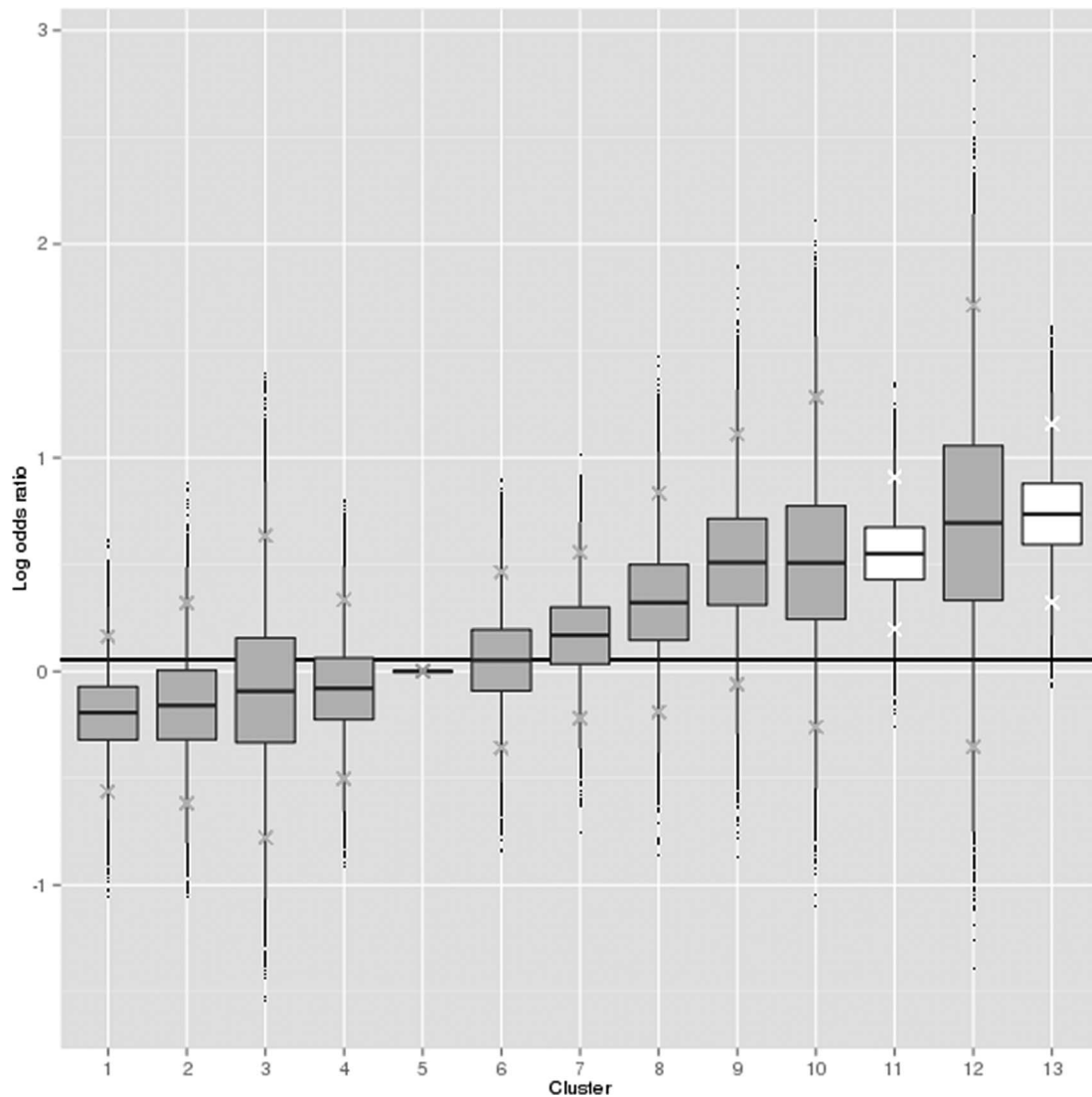


Figure 1 LogORs of clusters relative to the non-exposed cluster, men, France, 2001–2007. The baseline cluster is cluster 5. The horizontal line represents the logOR mean for each analysis. The clusters whose 95% credible interval includes the logOR mean are coloured in grey and the rest in white.

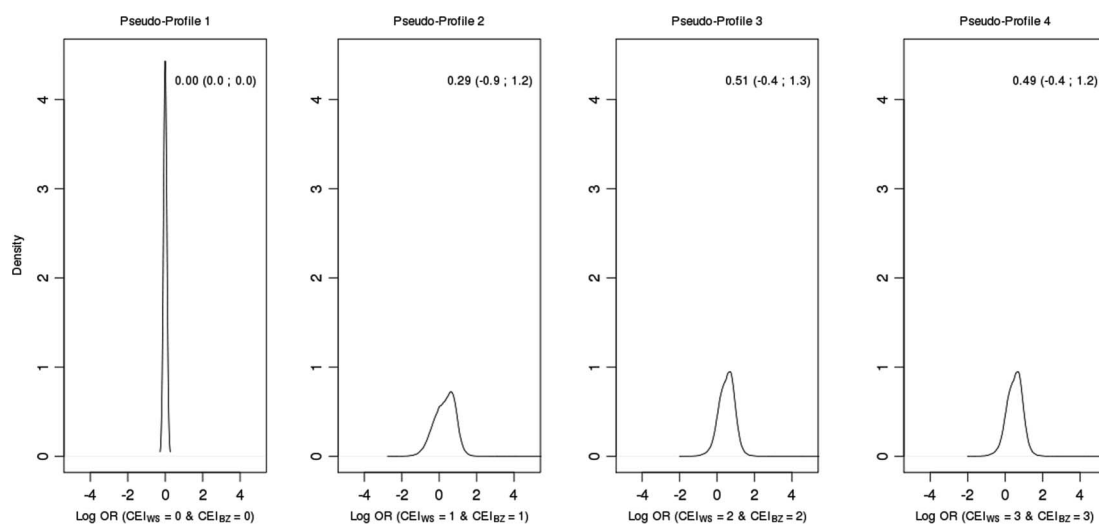


Figure 2 Density estimates of predicted logOR for Cumulative Exposure Index (CEI) of white spirits (WS) and benzene (BZ) combinations (logORs relative to pseudoprofile 1), men, France, 2001–2007. In each plot, mean logOR and corresponding credible interval at 95% are reported.

found an interesting concentration of painters,⁴⁰ construction workers, plumbers and pipe fitters, carpenters, joinery and parquet workers in the high-risk clusters.

These results are in agreement with previous analyses of occupations performed on the same study population, where statistically significant associations had been obtained for painters (in the construction field) with an OR estimated at 2.68; for plumbers and pipe fitters with an OR of 2.27 and for carpenters and joiners with an OR of 1.45.⁴¹ What our new analysis shows is that organic solvents do not appear to be substantial contributors to the occupational risk of lung cancer for these occupations.

However, these occupations are exposed to other substances which are known or suspected carcinogens for the lung, such as cadmium and chromium compounds for painters,³⁰ silica dust³⁰ and wood dust for wood industries,³⁰ and welding fumes⁴² and asbestos for plumbers and pipe fitters.^{40 42} The question of which of these agents contribute to increasing the risk of lung cancer for these occupations is still open.

Conclusions

With this multidimensional comprehensive analysis, we dissected a complex pattern of exposure to a large group of solvents using appropriate methodology. We did not detect any strong effect of solvents on risk of lung cancer. The large sample size and the careful study design aimed at reducing the different sources of bias give additional weight to our findings.

Author affiliations

¹Université Paris Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, France

²Department of Mathematics, Brunel University London, Uxbridge, UK

³MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge Biomedical Campus, Cambridge, UK

⁴Department of Epidemiology and Biostatistics, Imperial College London, MRC-PHE Centre for Environment and Health, School of Public Health, London, UK

⁵Faculty of Medicine IFR 10, University Paris Est-Créteil, Créteil, France

⁶University of Sydney, Sydney School of Public Health, Screening an Evaluation Test Program, Sydney, New South Wales, Australia

⁷Department of social epidemiology, INSERM, UMR_S 1136, Pierre Louis Institute of Epidemiology and Public Health, Villejuif, France

⁸Department of social epidemiology, Sorbonne University, UPMC University of Paris 06, UMR_S 1136, Pierre Louis Institute of Epidemiology and Public Health, Villejuif, France

⁹French Institute for Public Health Surveillance, Saint-Maurice, France

¹⁰Epidemiological research and surveillance unit in transport, occupation and environment, Claude Bernard Lyon1 University, Lyon, France

¹¹Somme Cancer Registry, Amiens, France

¹²INSERM, U 1085_IRSET, Pointe-à-Pitre, France

¹³University of Rennes 1, Rennes, France

Acknowledgements The authors thank Ms Joëlle Fevotte for designing occupational questionnaires; all members of the MatGéné working group from the Institut de Veille Sanitaire and, in particular, Ms Brigitte Dananché for providing job-exposure matrices; and Mr Gwenael GR Leday for his support in performing statistical analyses.

Collaborators ICARE Study Group: Anne-Valérie Guizard; Arlette Danzon; Anne-Sophie Woronoff; Velten Michel; Antoine Buemi; Émilie Marrer; Brigitte Tretarre; Marc Colonna; Patricia Delafosse; Paolo Bercelli; Florence Molinie; Simona Bara; Benedicte Lapotre-Ledoux; Nicole Raverdy; Oumar Gaye; Farida Lamkarkach; Mireille Matrat; Florence Guida; Sylvie Cène; Matthieu Carton; Diane Cyr; Gwen Menvielle; Sophie Paget-Bailly; Loredana Radoi; Annie Schamus; Alexandra Papadopoulos; Danièle Luce; Isabelle Stücker; Corinne Pilorget; Joëlle Fevotte.

Contributors IS and DL are the co-principal investigators of the ICARE Study. They designed the study, directed its implementation, and oversaw all aspects of the study, including patients and controls recruitment, funding, and quality control of data. SR developed the methodological Bayesian approach, provided critical comments on results interpretation and revised the manuscript critically for intellectual content. FM has performed the statistical analyses and interpreted the results. She wrote the manuscript and reviewed the whole document. SL made a substantial contribution to the analysis and interpretation of data. LA contributed to the statistical analysis. FG contributed to the interpretation of the results. SC and MS performed the data management, applied job-exposure matrices, and prepared data

sets for statistical analyses. MM and GM have been involved in the conception of the different variables included in the analysis and in the strategy of analysis. They participated in the writing of the manuscript. CP was involved in exposure assessment. BL-L is responsible for a cancer register and was particularly involved in the coding of the histology of the lung cancer cases. The ICARE Study Group implemented the ICARE Study.

Funding This analysis was supported by La Fondation ARC pour la Recherche sur le Cancer. The ICARE study was funded by the French National Research Agency (ANR); French National Cancer Institute (INCA); French Agency for Food, Environmental and Occupational Health and Safety (ANSES); French Institute for Public Health Surveillance (InVS); Fondation pour la Recherche Médicale (FRM); La Fondation de France; Ministry of Labour (Direction Générale du Travail); Ministry of Health (Direction Générale de la Santé).

Competing interests None declared.

Patient consent Obtained.

Ethics approval The Institutional Review Board of the French National Institute of Health and Medical Research (IRB-Inserm, number 01-036 and CNIL number 90120).

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Binder-Foucard F, Belot A, Delafosse P, et al. *Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Partie 1—Tumeurs solides*. Saint-Maurice, France: Institut de Veille Sanitaire, 2013.
- IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Tobacco Smoke and Involuntary Smoking. Vol 83. Lyon, France: International Agency for Research on Cancer, 2004.
- Imbernon E. *Estimation du nombre de cas de certains cancers attribuables à des facteurs professionnels en France*. Saint-Maurice: Institut de Veille Sanitaire, 2003.
- IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Chemical agents and related occupations. Vol 100F. Lyon, France: International Agency for Research on Cancer, 2012.
- IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Trichloroethylene, tetrachloroethylene and some other chlorinated agents. Vol 106. Lyon, France: International Agency for Research on Cancer, 2014.
- Hayes RB, Yin SN, Dosemeci M, et al. Mortality among benzene-exposed workers in China. *Environ Health Perspect* 1996;104(Suppl 6):1349–52.
- Collins JJ, Ireland B, Buckley CF, et al. Lymphohaematopoietic cancer mortality among workers with benzene exposure. *Occup Environ Med* 2003;60:676–9.
- Sorahan T, Kinlen LJ, Doll R. Cancer risks in a historical UK cohort of benzene exposed workers. *Occup Environ Med* 2005;62:231–6.
- Snyder CA, Sellakumar AR, James DJ, et al. The carcinogenicity of discontinuous inhaled benzene exposures in CD-1 and C57Bl/6 mice. *Arch Toxicol* 1988;62:331–5.
- Farris GM, Everitt JJ, Irons RD, et al. Carcinogenicity of inhaled benzene in CBA mice. *Fundam Appl Toxicol* 1993;20:503–7.
- National Toxicology Program. NTP Toxicology and Carcinogenesis Studies of Benzene (CAS No. 71-43-2) in F344/N Rats and B6C3F1 Mice (Gavage Studies). *Natl Toxicol Program Tech Rep Ser* 1986;289:1–277.
- Mehlman MA. Carcinogenic effects of benzene: Cesare Maltoni's contributions. *Ann N Y Acad Sci* 2002;982:137–48.
- Maltoni C, Conti B, Perino G, et al. Further evidence of benzene carcinogenicity. Results on Wistar rats and Swiss mice treated by ingestion. *Ann N Y Acad Sci* 1988;534:412–26.
- Stoner GD, Conran PB, Greisiger EA, et al. Comparison of two routes of chemical administration on the lung adenoma response in strain A/J mice. *Toxicol Appl Pharmacol* 1986;82:19–31.
- Mundt KA, Birk T, Burch MT. Critical review of the epidemiological literature on occupational exposure to perchloroethylene and cancer. *Int Arch Occup Environ Health* 2003;76:473–91.
- Brownson RC, Alavanja MC, Chang JC. Occupational risk factors for lung cancer among nonsmoking women: a case-control study in Missouri (United States). *Cancer Causes Control* 1993;4:449–54.
- Vizcaya D, Christensen KY, Lavoue J, et al. Risk of lung cancer associated with six types of chlorinated solvents: results from two case-control studies in Montreal, Canada. *Occup Environ Med* 2013;70:81–5.
- Mattei F, Guida F, Matrat M, et al. Exposure to chlorinated solvents and lung cancer: results of the ICARE study. *Occup Environ Med* 2014;71:681–9.

- 19 Papathomas M, Molitor J, Richardson S, *et al.* Examining the joint effect of multiple risk factors using exposure risk profiles: lung cancer in nonsmokers. *Environ Health Perspect* 2011;119:84–91.
- 20 Luce D, Stucker I. Investigation of occupational and environmental causes of respiratory cancers (ICARE): a multicenter, population-based case-control study in France. *BMC Public Health* 2011;11:928.
- 21 WHO. *International classification of diseases for oncology*. 3rd edn. Geneva, Switzerland: World Health Organization, 2000.
- 22 International Labour Office. *International Standard Classification of Occupations*. Geneva, Switzerland: International Labour Office, 1968.
- 23 INSEE. *Nomenclature d'activités Françaises*. Institut national de la statistique et des études économiques, 2000.
- 24 Fevotte J, Dananche B, Delabre L, *et al.* Matgene: a program to develop job-exposure matrices in the general population in France. *Ann Occup Hyg* 2011;55:865–78.
- 25 Matgéné Gdt. *Présentation de matrices emplois-expositions à cinq solvants chlorés. Quelques applications à un échantillon de population en France*. Saint-Maurice, France: Institut de Veille Sanitaire, 2009.
- 26 Matgéné Gdt. *Présentation d'une matrice emplois-expositions aux carburants et solvants pétroliers*. Saint-Maurice, France: Institut de Veille Sanitaire, 2007.
- 27 Dananché B, Houot M, Luce D, *et al.* *Estimation de prévalences d'expositions professionnelles en France à des solvants oxygénés aliphatiques. Calcul à partir des matrices emplois-expositions du programme Matgéné*. Saint-Maurice, France: Institut de Veille Sanitaire, 2015.
- 28 Guida F, Paget-Bailly S, Lamkarkach F, *et al.* Risk of lung cancer associated with occupational exposure to mineral wools: updating knowledge from a French population-based case-control study, the ICARE study. *J Occup Environ Med* 2013;55:786–95.
- 29 Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning*, paragraph 3 Linear Methods for Regression, 2013. statweb.stanford.edu/~tibs/ElemStatLearn/
- 30 IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Arsenic, Metals, Fibres, and Dusts. Vol 100C. Lyon, France: International Agency for Research on Cancer, 2012.
- 31 Leffondre K, Abrahamowicz M, Xiao Y, *et al.* Modelling smoking history using a Comprehensive Smoking Index: application to lung cancer. *Stat Med* 2006;25:4132–46.
- 32 Molitor J, Papathomas M, Jerrett M, *et al.* Bayesian profile regression with an application to The National Survey of Children's Health. *Biostatistics* 2010;11:484–98.
- 33 Liverani S, Hastie DI, Azizi L, *et al.* PREMIUM: an R package for Profile Regression Mixture Models using Dirichlet Processes. *J Stat Software* 2015;64:1–30.
- 34 Hastie DI, Liverani S, Azizi L, *et al.* A semi-parametric approach to estimate risk functions associated with multi-dimensional exposure profiles: application to smoking and lung cancer. *BMC Med Res Methodol* 2013;13:129.
- 35 Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. *Biom J* 2010;52:70–84.
- 36 MacLehose RF, Dunson DB, Herring AH, *et al.* Bayesian methods for highly correlated exposure data. *Epidemiology* 2007;18:199–207.
- 37 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10:37–48.
- 38 Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. *Am J Epidemiol* 1995;142:904–8.
- 39 Myers RH. *Classical and Modern Regression with Applications*. PWS-Kent Publishing Company, Boston, Massachusetts, 1990.
- 40 IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Some organic Solvents, Resin Monomers and Related Compounds, Pigments and Occupational Exposures in Paint Manufacture and Painting. Vol 47. Lyon, France: International Agency for Research on Cancer, 1989.
- 41 Guida F, Papadopoulos A, Menvielle G, *et al.* Risk of lung cancer and occupational history: results of a French population-based case-control study, the ICARE Study. *J Occup Environ Med* 2011;53:1068–77.
- 42 IARC Monographs on the Evaluation of the Carcinogenic Risks to Humans. Chromium, Nickel and Welding. Vol 49. Lyon, France: International Agency for Research on Cancer, 1990.