



Original article

Scand J Work Environ Health 2016;42(2):162-169

doi:10.5271/sjweh.3547

The 2-phase case-control design: an efficient way to use expert-time

by [Wild P](#), [Schill W](#), [Bourgard E](#), [Drescher K](#), [Gonzalez M](#), [Paris C](#)

The paper presents an optimized way to use expert exposure assessment in a subsample of cases and controls, chosen to maximize its information content in the context of a 2-phase lung cancer case-control. This study showed dose-response relationships with exposure to diesel motor exhaust, asbestos, PAH and crystalline silica.

Affiliation: INRS direction scientifique, Rue du Morvan, CS 60027, 54519 Vandoeuvre Les Nancy Cedex, France. Pascal.wild@inrs.fr

Key terms: [2-phase case-control design](#); [attributable fraction](#); [cancer](#); [case-control design](#); [case-control study](#); [dose-response relationship](#); [epidemiological method](#); [expert-based exposure assessment](#); [exposure assessment](#); [occupational epidemiology](#); [occupational exposure](#); [occupational exposure](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/26791138



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

The 2-phase case–control design: an efficient way to use expert time

by Pascal Wild, PhD,¹ Walter Schill, PhD,² Eve Bourgkard, PhD,³ Karsten Drescher, PhD,⁴ Maria Gonzalez, PhD,⁵ Christophe Paris, PhD⁶

Wild P, Schill W, Bourgkard E, Drescher K, Gonzales M, Paris C. The 2-phase case–control design: an efficient way to use expert time. *Scand J Work Environ Health*. 2016;42(2):162–169. doi:10.5271/sjweh.3547

Objectives The objective of this paper is to show the benefits of using a 2-phase case–control (2PCC) design in identifying dose–response relationships between cumulative occupational exposure as assessed by experts and lung cancer incidence in an actual study.

Methods A population-based case–control study including 246 cases and 531 controls was conducted in an area with high lung cancer rates in Northeast France. Detailed occupational and personal risk factors were obtained in face-to-face interviews. Cumulative expert-based exposure scores were obtained from a subset of 215 cases and 269 controls stratified on smoking and a prior algorithmic exposure score for asbestos, crystalline silica, and polycyclic aromatic hydrocarbons (PAH) in the framework of a 2PCC design. This subset deliberately under-sampled large strata among controls but not among cases. Logistic regression models adapted to 2PCC studies were applied and corresponding computations of attributable fractions and their confidence intervals developed.

Results Based on this 2PCC design, statistically significant dose–response relationships were obtained for asbestos, crystalline silica, PAH, and diesel motor exhaust. Simulations within this study showed that 2PCC studies were always more powerful than random samples.

Conclusion The 2PCC design may be the design of choice when resources allow only a limited number of subjects with a full expert-based exposure assessment.

Key terms attributable fraction; cancer; dose–response relationship; epidemiological method; expert-based exposure assessment; exposure assessment; occupational epidemiology; occupational exposure.

The main source of evidence concerning the carcinogenicity of occupational hazards in humans is obtained via epidemiological studies among workers, and one of the main criteria for establishing a causal link is the existence of dose–response relationships. Historically, most of dose–response relationships have been obtained in cohort studies of workers who are, if not solely, usually predominantly exposed to a single carcinogen. For instance, in the recent IARC classification of diesel motor exhaust (DME) as a group 1 carcinogen, the cohort studies (1, 2) – in which exposure had been accurately determined and a dose–relationship observed – have carried great weight. Population-based case–

control studies have been less successful in identifying dose–response relationships. This might be due to the fact that exposure cannot rely on measurements carried out in the actual workplace and must be estimated indirectly. Most population-based case–control studies that managed to identify dose–response relationships relied on expert-based exposure assessments [see eg, (3, 4)].

This possibly reflects the fact that experienced experts are able to synthesize the different types of available information (job, industry, time-period, task descriptions, use of protective devices, self-reported exposure, etc.) into meaningful quantitative exposure levels. However, expert time is rare and costly, thus

¹ INRS, Scientific Management, Vandoeuvre Les Nancy, France.

² Leibniz Institute for Prevention Research and Epidemiology -BIPS, Bremen, Germany.

³ INRS, Department of Epidemiology, Vandoeuvre Les Nancy, France.

⁴ Statistical Office, Bremen, Germany.

⁵ Strasbourg University Hospital, Department of Occupational Diseases, Strasbourg, France.

⁶ Lorraine University, team INGRES, Vandoeuvre Les Nancy, France.

Correspondence to: Pascal Wild, INRS direction scientifique, Rue du Morvan, CS 60027, 54519 Vandoeuvre Les Nancy Cedex, France. [E-mail: Pascal.wild@inrs.fr]

expert-assessment is often obtained only for subsets of subjects. So-called 2-phase designs (5) allow an efficient choice of such subsets as was repeatedly shown from a theoretical as well as applied point of view (6–8). Briefly, in 2-phase studies, the full dataset (the phase-1 dataset) is stratified on the case–control status and exposure estimates available at this stage. Within these strata, possibly unequal fractions of cases and controls are selected for further data collection (phase-2 dataset). In the present case, this second data collection is a detailed expert assessment. The statistical analysis makes use of these sampling fractions and thus uses not only the phase-2 data but also the phase-1 information, thereby adding additional power, see for example (9). Moreover, this design allows oversampling of the most informative subjects, ie, those subjects with the rarest disease-exposure combinations.

Despite these properties, this design is quite uncommon in the epidemiological literature and in the field of occupational epidemiology, we know only of one previous paper (9). In the present paper, we aim to show the power and the feasibility of the 2-phase case–control (2PCC) design in making efficient use of expert time in the context of occupational epidemiology and on the basis of an actual lung cancer case–control study. Moreover, we present some new developments in computing the attributable fractions for this design.

Methods

Phase-1 case–control study

The study, which is the basis of our paper and which constitutes phase 1 of our 2PCC, is a population-based lung cancer case–control study in Northeast France. Details of the data collection and the population have been presented elsewhere (10). The ethics committee of the French data protection agency approved the study. Briefly, all the hospitals located in this area were contacted and agreed to declare their incident lung cancer cases. Cases were eligible if: (i) male and between 40–80 years old, (ii) resident in the study region, (iii) lung cancer was histologically confirmed, and (iv) written statements of informed consent were provided. Of 548 identified lung cancer cases, 246 were interviewed between 2006–2010.

Eight hundred male controls, aged 40–80 years, stratified on 10-year age classes and broad socioeconomic class and selected by a random digit dialing procedure in the study area, agreed to be contacted. Of these, 531 finally accepted to be interviewed.

All the cases and controls were contacted by telephone and interviewed at home in a face-to-face interview

mainly focused on their occupational exposure, smoking, and disease history. Occupational exposure was first assessed by obtaining a lifelong list of all jobs held for three or more months. For each job, a general exposure questionnaire and, if relevant, one or more detailed exposure questionnaire specific to 20 jobs or activities were applied. Finally, a list of 47 task-specific questions was applied to the whole job history independently of any job code. Bourgard et al (11) gives further details on this approach and its validity. For each job, a rule-based exposure level in four categories was assigned to each job for asbestos and polycyclic aromatic hydrocarbons (PAH) based on the task questionnaire and for crystalline silica based on the general and specific questionnaires. Cumulative exposure scores were computed multiplying this level by job duration and exposure frequency and summing them up over the job history.

Setting up the two-phase study design

The 2-phase study was designed following the approach described in Schill et al (12). First, a stratification summarizing the exposure variables of interest was set up, cross-tabulating smoking and the three rule-based exposure scores. The first stratum was defined by never smokers whatever the exposure scores. Smokers were cross-classified by the algorithmic cumulative exposure scores of asbestos, PAH, and crystalline silica, each in turn divided into non-exposed, lower than the median score, and higher than this median. Finally, some of these strata were collapsed owing to small numbers, resulting finally to 21 strata.

Second, the stratum-specific numbers of cases and controls were chosen (table 1) for inclusion in phase 2. We wanted to be able to assess simultaneously the four main occupational carcinogens and some confounders so that virtually all cases and half of the controls were selected with approximately equal numbers in each stratum as suggested by (6). It should be noted that equal numbers correspond to very different sampling fractions in cases and controls. Phase 2 is thus deliberately non representative of phase 1, but representative within each stratum.

Phase-2 data collection: expert exposure assessment

For the cases and controls selected, two industrial hygiene experts, blind to the case–control status and the algorithmic exposure evaluation, coded the exposure to a pre-defined list of potential occupational carcinogens: asbestos, man-made mineral fibers, painting, PAH, crystalline silica, mild steel welding fumes, stainless steel welding fumes, iron mining, painting, strong acids, and DME. These exposure codes were based on all the available information in the questionnaires. The experts sum-

Table 1. Stratification of the two-phase study and numbers of subjects included in phase one and phase two according to stratum and case/control status. [PAH=polycyclic aromatic hydrocarbons.]

	Strata	Asbestos ^a	PAH ^a	Crystalline silica ^a	Phase 1		Phase II	
					Controls	Cases	Controls	Cases
Smokers and ex-smokers	1	No	No	No	94	31	19	27
	2	Low	No	No	13	10	10	9
	3	No	No	Low	23	8	8	8
	4	Low	No	Low	25	4	14	4
	5	High	No	No or Low	15	9	14	6
	6	No	Low	No or Low	29	8	16	8
	7	Low	Low	No or Low	31	19	18	17
	8	High	Low	No or Low	14	6	12	5
	9	No	High	No or Low	10	3	8	3
	10	Low	High	No or Low	22	13	21	11
	11	High	High	No or Low	15	21	15	18
	12	No	No	High	19	5	10	5
	13	Low	No	High	7	6	7	6
	14	High	No	High	13	5	10	4
	15	No	Low	High	11	8	6	8
	16	Low	Low	High	16	11	13	10
	17	High	Low	High	11	14	11	13
	18	No	High	High	9	6	7	4
	19	Low	High	High	8	15	8	11
	20	High	High	High	25	36	22	30
Non-smokers	21	All	All	All	121	8	20	8
Total					531	246	269	215

^a Asbestos, PAH, and crystalline silica were classified as No, Low or High according to the algorithmic exposure assessment, the cut-point between the Low and High categories was based on a median exposure.

marized the exposure at individual level to each of the carcinogens as a series of homogeneous exposure periods with an exposure frequency, a semi-quantitative exposure level in four increasing categories and duration in years. As in the algorithmic step, a cumulative exposure score was computed for each exposure period. These cumulative exposures were log-transformed after adding one so that non-exposed subjects had a zero log-transformed exposure. Discordant cumulative scores were identified and a second panel of three senior experts provided a final expertise obtained in consensus. Finally, for each potential carcinogen, we computed a final exposure estimate. If the initial expert assessments were consistent, this estimate was the mean of the log-transformed cumulative exposures of these two experts. If not, the exposure estimate was the consensus estimate of the second panel.

Statistical analysis

The statistical analysis was based on logistic regression adapted to 2-phase data. We chose the maximum likelihood approach as it is the more powerful. An alternative would have been the weighted likelihood approach, which is less sensitive to misclassification. Basically, these methods consist of modelling the detailed phase-2 data in a logistic regression weighted by stratum-specific ratios of numbers of subjects in phases 1 and 2. When using maximum likelihood, these weights are adjusted according to the model used (13). We used the SAS/

IML® package (14). We also developed attributable fraction computations adapted to 2-phase studies, extending work by others (15, 16), in order to be able to estimate within this design the overall public health burden due to these occupational exposures. Mathematical details and a freely available add-on to the above-mentioned SAS software are available at <http://tinyurl.com/schill-af2p>.

The logistic regression model included as independent variables the confounders selected in the prior analysis of the full phase-1 data, with the exception of wine drinking whose borderline protective effect was not considered worth keeping. In order to assess the associations with occupational exposures, we included all the expert-based quantitative phase-2 exposure scores (asbestos, man-made mineral fibers, PAH, crystalline silica, mild steel welding fumes, stainless steel welding fumes, iron mining, painting, strong acids, and DME) and performed a backward selection procedure, keeping all the exposure variables with a P-value for inclusion of <10%. For the quantitative exposure variables thus selected, we then fitted a model including the indicator variables of the exposure tertiles compared to the non-exposed in order to check for non-monotonous patterns. Quartiles as used in the previous paper (10) led to a non-identifiable model.

Simulations

In order to document the added power due to the 2PCC design as compared to the standard analysis of a random

subsample, we sampled 1000 times the phase-2 exposure for all subjects not belonging to phase 2: separately for cases and controls, we assigned the expert-assessed exposure of a phase-2 subject randomly chosen from the same stratum to each subject not in phase 2.

Firstly, from each such completed phase-1 dataset, we randomly drew 215 cases, and 269 controls were analyzed using standard logistic regression for comparison with the results in table 3.

Secondly, in order to document the power of the 2PCC design with less subjects in phase 2, we sampled 1000 times five cases and controls (or all subjects if <5 had been subjected to an expertise) from each stratum of phase 2, that is 105 controls and 100 cases, and analyzed them using 2P logistic regression.

In parallel, we randomly sampled – without 2P stratification – 105 controls and 100 cases from each of the 1000 completed datasets described earlier and analyzed them using standard logistic regression.

All analyses used the same list of independent variables for the analysis of the three series of simulations as outlined in table 3. This list comprised the confounders and the four quantitative occupational exposure scores.

Results

Table 2 shows that the age-structure of cases and controls in our original data was similar in phases 1 and 2. On the other hand, the number of smokers and the number of exposed subjects was by design overrepresented in phase-2 controls. Among cases there were few differences between phases 1 and 2. A second striking feature is that the experts identified more subjects exposed to PAH and much fewer for silica than the algorithm. However, in the phase-2 dataset, the algorithm-based exposure indices were all higher than the corresponding expert-based estimates.

Table 3 provides the results of the 2-phase analysis. When running the backwards selection procedure, only asbestos, PAH, crystalline silica, and DME remained in the model (data not shown). Statistically significant dose–response slopes were observed for asbestos, crystalline silica, and DME and borderline for PAH adjusting for quantitative smoking and time since cessation, age, and familial history of lung cancer by age of onset.

Moreover, all the odds ratio (OR) corresponding to the highest exposure tertiles were statistically significant and all either >2 (asbestos, silica and DME) or very nearly so (PAH).

The estimated fractions attributable to occupational exposure were high (>50% for all the selected occupational carcinogens) but of same order of magnitude than in the analysis of the full phase-1 data using algorithmic

exposure scores given in Wild et al (10): 56% (41–67%) for continuous scores, 52% (32–66%) for discretized scores.

The simulations of the 1000 unstratified random subsets of the same size as our phase-2 study, led to statistical significance in 100% of the cases for asbestos, but only 39% for silica, 13% for PAH and 20% for DME. This should be interpreted bearing in mind the P-values in our 2PCC analysis presented in table 3 (asbestos 0.0001, silica 0.009, PAH 0.08, DME 0.02). Concerning the simulation study of the smaller samples of 105 controls and 100 cases, the number of statistically significant results from the subsets of the 2PCC, respectively for the simple random samples again without the 2-phase-specific stratification, were 85% versus 70% for asbestos, 26% versus 20% for silica, 19% versus 10% for PAH, and 62% versus 11% for DME. One can note that for DME, the 2PCC sample of 205, was more often significant than the unstandardized random sample of 484. The results from these simulations document the added power of the 2PCC design.

Discussion

In this paper, we showed the power of the 2-phase study design. Despite a limited sample size, a statistically significant dose–response relationships could be identified for asbestos, PAH, crystalline silica, and DME simultaneously when using expert-based exposure assessments in such a design. This efficiency gain is further substantiated by the comparison of simulations of unstratified random subsets of cases and controls with 2PCC studies of the same size. Moreover, we devised maximum-likelihood estimates and confidence intervals of attributable fractions within 2PCC studies which were as yet unavailable.

First, we discuss the specifics of the 2-phase design. This design is based on a stratification on an exposure estimate that is available before any expert assessment. In the case of our study, we used an algorithmic exposure assessment but a job-exposure matrix (JEM) could have been an alternative source for such phase-1 exposure estimate. A striking feature of the phase-2 dataset is its design-based difference with the total sample. As shown in table 2, 68% of the phase-2 controls were found to be exposed to asbestos compared to 52% of phase-1 controls. This biased selection, which is, moreover, different between cases and controls, might seem counterintuitive to epidemiologists. However, this is completely taken into account by the weighting performed in the dedicated statistical analyses for which packages in SAS (14) as well as in R (17) are now available. These weights do not only correct for this lack of

Table 2. General and exposure characteristics in phases one and two by case status [DME=diesel motor exhaust; IQR=interquartile range; M=median]

	Phase I								Phase II							
	Controls				Cases				Controls				Cases			
	N	%	M	IQR	N	%	M	IQR	N	%	M	IQR	N	%	M	IQR
Age class																
40-49	53	10			14	6			30	11			14	7		
50-59	140	26			67	27			69	26			59	27		
60-69	177	33			86	35			93	35			75	35		
70-79	161	30			79	32			77	29			67	31		
Smoking																
Never smokers	121	23			8	3			20	7			8	4		
Pack-years			26	13-40			37	25-51			27	15-41			37	26-51
Familial lung cancer history																
None	497	94			218	89			252	94			190	89		
Early onset	8	2			11	5			3	1			11	5		
Late onset	26	5			17	7			14	5			14	7		
Asbestos by algorithm																
Exposed	275	52			175	71			184	68			150	70		
Cumulative exposure (years f/mL)			3.5	0.3-22.7			9.1	0.9-70.9			6.1	0.4-29.5			9.2	0.6-71.4
Asbestos by expertise																
Exposed									179	67			151	70		
Cumulative exposure (years f/mL)											1.9	0.6-7.5			3.3	0.8-15.1
PAH by algorithm																
Exposed	120	23			92	37			79	29			78	36		
Cumulative exposure (years ng/mL)			3.2	0.4-17.8			9.6	1.5-92.0			3.2	0.4-16.7			11.9	1.4-125
PAH by expertise																
Exposed									143	53			130	61		
Cumulative exposure (years ng/mL)											0.4	0.1-1.1		0.6	0.2-3.2	
Crystalline silica by algorithm																
Exposed	337	63			193	78			206	77			173	80		
Cumulative exposure (years u/mL)			500	100-2095			1390	300-4854			640	106-2094			1303	232-4122
Crystalline silica by expertise																
Exposed									125	47			115	54		
Cumulative exposure (years u/mL)											8.5	0.9-94.7			45	3.3-214
DME by algorithm																
Exposed	121	23			91	37			71	26			75	35		
DME by expertise																
Exposed									88	33			81	38		
Cumulative exposure (years u/mL)											2.7	0.7-22.7			14.7	1.2-59.0

representativeness but also add some information from phase 1, ie, some power. This added information is more important if the stratification is finer, but for statistical reasons, the strata must not be too small. This led to our choice of the median exposure as the cut-point between categories and our decision to collapse some strata resulting from the cross-classification of the phase-1 exposure variables.

What did we gain with respect to a full assessment? In the present study, we included 484 subjects in the second phase for expert assessment out of 777 study participants; thus, while saving nearly 300 expert

assessments, the effort was still substantial. However for the simultaneous detection of the four dose-response relationships, the number of subjects included in phase 2 seems adequate as this showed quite clear-cut associations for the most frequent occupational carcinogens. When comparing with the simulations of a random subsample of the same size, it becomes clear that the chance of having such simultaneous associations would have been highly unlikely, again documenting the gain in power through the 2PCC design.

It must however be stressed that the full benefit of 2-phase studies is more apparent when the exposure

Table 3. Results of the multiple two-phase logistic models [AF=attributable fractions in %; DME=diesel motor exhaust; OR=odds ratio; PAH=polycyclic aromatic hydrocarbons; 95%CI=95% confidence interval;]

	Model with continuous exposure variables ^a				Model with exposure variables in tertiles ^a			
	OR	95% CI	AF	95% CI	OR	95% CI	AF	95% CI
Asbestos								
Ln (years f/mL+1)	2.097	1.463–3.006	28	19–35				
Non-exposed					1		34	18–45
T1 <1.01 years f/mL					1.69	0.98–2.92		
T2 1.01–6.72 years f/mL					1.38	0.79–2.40		
T3 >6.72 years f/mL					2.96	1.70–5.15		
PAH								
Ln(years ng/mL+1)	1.605	0.953–2.703	10	0–20				
Non-exposed					1		19	0–35
T1 <0.22 years ng/mL					1.12	0.63–2.00		
T2 0.22–0.94 years ng/mL					1.38	0.78–2.44		
T3 >0.94 years ng/mL					1.96	1.11–3.46		
Crystalline silica								
Ln(years u/mL+1) ^b	1.379	1.094–1.738	19	8–28				
Non-exposed					1		18	4–30
T1 <3.0 years u/mL					0.81	0.45–1.45		
T2 3.0–71.5 years u/mL					1.76	0.98–3.13		
T3 >71.5 years u/mL					2.18	1.20–3.97		
DME								
Ln(years u/mL+1) ^b	1.205	1.035–1.402	13	4–21				
Non-exposed					1		8	–6–21
T1 <1.28 years u/mL					0.94	0.49–1.82		
T2 1.28–18.7 years u/mL					0.96	0.50–1.83		
T3 >18.7 years u/mL					2.13	1.07–4.22		
All occupational exposures			50	42–56			56	45–64

^a Adjusted on cumulative smoking, time since cessation, age and familial history of lung cancer by age of onset.

^b Arbitrary units for exposure level based on assumed multiplicative scale between semi-quantitative exposure levels.

prevalence is low. In such a case, a random sample of cases and controls would include too small a number of exposed study participants to lead to any clear dose–response relationship. The power of 2PCC studies is mostly driven by the size of the phase-2 sample and the exposure distribution therein, although as mentioned before, phase 1 contributes some information. The size of phase 2 would thus not have to be increased in the case of a phase 1 with several thousand cases and controls in a low-exposure prevalence area. This assumes however that the phase-1 exposure assessment methods, and hence the stratification, identify the exposed subjects with sufficient accuracy. If the phase-1 stratification is less efficient in identifying these rare exposures, the result will be a reduction of the overall efficiency of the 2-phase design.

As an example, one could consider a study focused on the carcinogenic potential of man-made mineral fibers (MMMF) in the absence of asbestos – a rare occurrence, see for example (18) – it would be possible to apply a stratification on an algorithmic exposure or a JEM cross-classifying asbestos and MMMF. The 2PCC study would include all cases and controls instead of a random sample, for which phase 1 would suggest an MMMF exposure without asbestos. The expert assessment in phase 2 would then confirm this pattern of exposure (or not) and have greater power in detecting any specific association with MMMF.

Next, we discuss the use of experts. First, while we acknowledge that no expertise can be perfect, the data collection was, to our knowledge, reasonable in the context of a retrospective study and allowed for as-good-as-possible evaluation. When considering continuous cumulative exposure variables, we observed much higher dose–response slopes than in our previous paper, although when considering the categorical classes, the OR are quite similar. This is partly due to the fact that the cumulative exposure indices were systematically larger when assessed by the algorithms compared to the expert-based indices. This might indicate that the algorithmic scores associated with some tasks were tuned to the worst cases whereas the expert estimation could be more differentiated in assessing lower exposure when warranted. However, in order to explore whether this might also be due to exposure misclassifications, we divided the individual algorithmic cumulative exposure scores (not log transformed) by a single constant so that the mean (log-transformed) algorithmic and expert-based cumulative exposure scores were approximately equal. The aim of this transformation was to try to homogenize the units between the two exposure assessments in order to be able to compare the slopes. For asbestos as well as for PAH and crystalline silica (data not shown), the expert-based dose–response slopes were still steeper than those based on the algorithmic expo-

sure estimates. As the units are now assumed to be the same, this tentatively supports the use of expert-based exposure assessments as compared to our algorithmic scores. This is in contrast to recent results (19, 20) showing that, if based on closed form questions, algorithmic exposure estimates perform as well as expert assessments. In that case, the more transparent and reproducible algorithmic assessment should be preferred. On the other hand, when not sufficiently detailed closed form questions are available or when it is necessary to interpret open questions, as was the case for DME exposure in our study, an expert assessment is the only achievable way to obtain quantitative exposure estimates.

Finally we discuss the substantive aspects of the present paper. Our expert exposure assessment confirmed the dose-response relationships based on algorithmic exposure estimations that we had observed in an earlier analysis of the algorithmic part of this case-control study for asbestos, crystalline silica and, partially, for PAH (10). What is new is the significant dose-response for DME based on the individual expert assessments. This is in agreement with the recent classification of the IARC of diesel as a group 1 carcinogen, but we know of no population case-control study having shown such a dose-response relationship. It is to be noted that the test for the dose-response somewhat arbitrarily relies on the linear fit of the log-transformed scores. This trend roughly fits the presented analysis by tertiles if one allows for the large confidence intervals in the OR estimation. There is however a hint at a possible non-linearity for DME, as there is no excess risk in the two lowest exposure tertiles.

These results must be interpreted bearing in mind the overall strengths and weaknesses of the case-control study discussed extensively in our previous paper (10). Briefly, the results of this paper show a prevalence of occupational exposure in the study area, which is higher than in most published papers. This high exposure prevalence might be partly due to methodological differences with other studies in the exposure assessment, but the fact that the overall lung cancer and mesothelioma rates in the study area are very high is consistent with the present findings. The controls were selected based on a random digit dialing procedure. Although this may have led to selection bias, this is unlikely as the same procedure was applied to the cases who were not interviewed if they could not be reached by telephone.

Concluding remarks

When expert assessment is the only way to obtain quantitative estimates of exposure in case-control studies and a prior exposure estimation is available, the 2-phase design provides an adequate framework for the efficient use of expert time.

Acknowledgments

We acknowledge Dr Sigrid Behr for her helpful comments on an earlier draft.

The study was cofinanced by a grant from the French ANSES funding agency (grant number EST 07-17), the “Programme Hospitalier de Recherche Clinique – PHRC 2004”), and the core funding of the INRS.

The authors declare no conflicts of interest.

References

1. Attfield MD, Schleiff PL, Lubin JH, Blair A, Stewart PA, Vermeulen R et al. The diesel exhaust in miners study: A cohort mortality study with emphasis on lung cancer. *J Natl Cancer Inst.* 2012;104:869–83. <http://dx.doi.org/10.1093/jnci/djs035>.
2. Silverman DT, Samanic CM, Lubin JH, Blair AE, Stewart PA, Vermeulen R, Coble JB et al. The diesel exhaust in miners study: A nested case-control study of lung cancer and diesel exhaust. *J Natl Cancer Inst.* 2012;104:855–68. <http://dx.doi.org/10.1093/jnci/djs034>.
3. Iwatsubo Y, Pairon JC, Boutin C, Menard O, Massin N, Caillaud D et al. Pleural mesothelioma: Dose-response relation at low levels of asbestos exposure in a french population-based case-control study. *Am J Epidemiol.* 1998;148:133–42. <http://dx.doi.org/10.1093/oxfordjournals.aje.a009616>.
4. Vida S, Pintos J, Parent ME, Lavoue J, Siemiatycki J. Occupational exposure to sil-ica and lung cancer: Pooled analysis of two case-control studies in montreal, Canada. *Cancer Epidemiol Biomarkers Prev.* 2010;19:1602–11.
5. Breslow NE. Case-control study, two-phase. In: Armitage P, Colton, T., editor. *Encyclopedia of Biostatistics*. New York: Wiley; 2000. p. 532–40. <http://dx.doi.org/10.1002/0470011815.b2a03029>.
6. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. *Am J Epidemiol.* 1988;128:1198–206.
7. Wild P, Andrieu N, Goldstein AM, Schill W. Flexible two-phase studies for rare exposures: Feasibility, planning and efficiency issues of a new variant. *Epidemiol Perspect Innov.* 2008;5:4. <http://dx.doi.org/10.1186/1742-5573-5-4>.
8. Haneuse S, Hedt-Gauthier B, Chimbwandira F, Makombe S, Tenthani L, Jahn A. Strategies for monitoring and evaluation of resource-limited national antiretroviral therapy programs: The two-phase design. *BMC Med Res Methodol.* 2015;15:31. <http://dx.doi.org/10.1186/s12874-015-0027-9>.
9. Pohlabein H, Wild P, Schill W, Ahrens W, Jahn I, Bolm-Audorff U et al. Asbestos fibreyears and lung cancer: A two phase case-control study with expert exposure assessment. *Occup Environ Med* 2002;59:410–4. <http://dx.doi.org/10.1136/oem.59.6.410>.

10. Wild P, Gonzalez M, Bourgkard E, Courouble N, Clement-Duchene C, Martinet Y et al. Occupational risk factors have to be considered in the definition of high-risk lung cancer populations. *Br J Cancer*. 2012;106:1346–52. <http://dx.doi.org/10.1038/bjc.2012.75>.
11. Bourgkard E, Wild P, Gonzalez M, Fevotte J, Penven E, Paris C. Comparison of exposure assessment methods in a lung cancer case–control study: Performance of a lifelong task-based questionnaire for asbestos and pahs. *Occup Environ Med*. 2013;70:884–91. <http://dx.doi.org/10.1136/oemed-2013-101467>.
12. Schill W, Wild P, Pigeot I. A planning tool for two-phase case–control studies. *Comput Methods Programs Biomed*. 2007;88:175–81. <http://dx.doi.org/10.1016/j.cmpb.2007.08.002>.
13. Breslow NE, Holubkov R. Weighted likelihood, pseudo-likelihood and maximum likelihood methods for logistic regression analysis of two-stage data. *Stat Med*. 1997;16:103–16. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19970115\)16:1<103::AID-SIM474>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1097-0258(19970115)16:1<103::AID-SIM474>3.0.CO;2-P).
14. Schill W, Enders D, Drescher K. A sas-package for logistic two-phase studies. *J Statist Soft*. 2014;59:1–13.
15. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population at-tributable risk for multiple risk factors using case–control data. *Am J Epidemiol*. 1985;122:904–14.
16. Greenland S, Drescher K. Maximum likelihood estimation of the attributable frac-tion from logistic models. *Biometrics* 1993;49:865–72. <http://dx.doi.org/10.2307/2532206>.
17. Lumley T. *Complex surveys: A guide to analysis using r*. Wiley: New York; 2010. <http://dx.doi.org/10.1002/9780470580066>.
18. Guida F, Paget-Bailly S, Lamkarkach F, Gaye O, Ducamp S, Menvielle G 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. <http://dx.doi.org/10.1097/JOM.0b013e318289ee8b>.
19. Peters S, Glass DC, Milne E, Fritschi L. Rule-based exposure assessment versus case-by-case expert assessment using the same information in a community-based study. *Occup Environ Med*. 2014;71(3):215–9. <http://dx.doi.org/10.1136/oemed-2013-101699>.
20. Friesen MC, Pronk A, Wheeler DC, Chen YC, Locke SJ, Zaebs DD et al. Compari-son of algorithm-based estimates of occupational diesel exhaust exposure to those of multiple independent raters in a population-based case–control study. *Ann Occup Hyg*. 2013;57:470–81. <http://dx.doi.org/10.1093/annhyg/mes082>.

Received for publication: 10 June 2015