

Evaluation of the quality of coding of job episodes collected by self questionnaires among French retired men for use in a job-exposure matrix

C Pilorget, E Imbernon, M Goldberg, S Bonenfant, Y Spycykerelle, B Fournier, J Steinmetz, A Schmaus

Occup Environ Med 2003;**60**:438-443

Background and Aims: The ESPACES study was intended to identify retirees who may have been, according to their job descriptions, exposed to asbestos during their working lives. As part of this study, we analysed the quality of the occupation and activity sector coding as well as its effect on the subjects' exposure status.

Methods: The occupation and activity sector for a sample of 450 retired men were coded twice (with the second coder blinded to the first result), according to the international codes for industries (ISIC-1975) and occupations (ISCO-1968). For each series, linking the information about a job episode (dates, ISIC code, ISCO code) with the matrix allowed attribution of a probability of asbestos exposure to each episode and each subject. The asbestos exposure in the two series was compared by the kappa reproducibility coefficient.

Results: The analysis concerned 425 questionnaires. There was at least one difference in the code for either activity sector (ISIC) or occupation (ISCO) in half the episodes (50.2%). The exposure status estimated by the job-exposure matrix did not change between the series for 84.7% of the subjects. The kappa coefficient was 0.64 for all questionnaires, 0.70 when the questionnaire was coded twice by the same coder, and 0.62 when coded by two different coders.

Conclusions: Despite intra- and inter-differences between coders, the coding of job episodes for the ESPACES study appears satisfactory and hence indicates that the assessment of the subjects' asbestos exposure was assessed without major distortions. This study underlines the usefulness of employing coders specifically trained for this technique.

See end of article for authors' affiliations

Correspondence to: C Pilorget, INSERM Unité 88, 14 rue du Val d'Osne, 94415 Saint-Maurice, France; Corinne.Pilorget@st-maurice.inserm.fr

Accepted 11 October 2002

Exposure assessment is an important stage in epidemiological studies, especially those of diseases with long latency periods when exposures that may have occurred in the distant past have to be taken into account.¹ The objective of the ESPACES (Evaluation of Post-retirement Asbestos Follow-up in Health Examination Centres) project^{2,3} was to identify among the retired men in the French population those who had been exposed to asbestos during their working life, to inform them of their entitlement to post-retirement follow up, and to assist them with the necessary procedures, in accordance with the 1995 French regulation instituting such medical follow up for those exposed to carcinogenic substances while employed. The ESPACES project, which offers a tool for identifying these populations, could help increase the hitherto sparse application of this regulation.

In this view, asbestos exposure was assessed by a job-exposure matrix (JEM) specific for asbestos⁴ and structured according to standard international classifications, ISCO (International Standard Classification of Occupation)⁵ for occupations and ISIC (International Standard Industrial Classification of all Economic Activities)⁶ for activity sectors. Individual exposure was assessed by linking each job episode, obtained through a questionnaire completed by the retired subjects, with the matrix by the means of occupation and activity sector codes.

The use of a JEM to assess individual exposure has long been known to involve some misclassification engendered by the imprecision of the job exposure assessments included in the matrix.⁷⁻⁹ Much less attention has been paid to the frequency and consequences of the misclassification which can also stem from the process of ISCO-ISIC coding of the job titles and activity sectors reported by the subjects.

In this study, a subgroup of the questionnaires of subjects included in the ESPACES project were coded twice. Our objective was to assess the quality of the coding of the "working life" portion of the questionnaires and its impact on the subjects' asbestos exposure classification.

METHODS

Six thousand recently retired men (in 1994, 1995, and 1996) were randomly selected from the National Health Insurance files in six French "départements" (administrative units) in 1998. They received a self administered questionnaire through the mail, and 59.5% responded (n = 3572).

The questionnaire included a section about their work history; it asked them to list their successive job episodes, with the period (start and end dates), employer's name, activity sector, and job, all as free response questions. Each questionnaire was then coded manually by specially trained staff, following the international classifications of industries (ISIC) and of occupations (ISCO). ISIC uses four digits, and ISCO five.

We used a randomly selected sample of 450 questionnaires from the ESPACES study to assess the quality of the coding procedure. These 450 questionnaires were first distributed among three coders for the initial coding. Three weeks later, each questionnaire was coded again, with the coder blinded to

Abbreviations: ISCO, International Standard Classification of Occupation; ISIC, International Standard Industrial Classification of all Economic Activities; JEM, job-exposure matrix

Table 1 Asbestos exposure in job episodes (n=2099) and subjects (n=425) in the two coding series after linkage with the job-exposure matrix

	Exposed* n (%)	Not exposed† n (%)	Not classified‡ n (%)
<i>Episodes</i>			
1st series	1044 (49.7)	784 (37.4)	271 (12.9)
2nd series	1039 (49.5)	819 (39)	241 (11.5)
1st and 2nd series	918 (43.7)	666 (31.7)	132 (6.3)
<i>Subjects</i>			
1st series	303 (71.3)	68 (16)	54 (12.7)
2nd series	317 (74.6)	60 (14.1)	48 (11.3)
1st and 2nd series	288 (67.8)	44 (10.3)	28 (6.6)

*Exposed episodes: exposure probability >0; exposed subjects: at least one episode was classified as exposed.

†Not exposed episodes: exposure probability = 0; not exposed subjects: all the episodes were not exposed.

‡Not classified episodes: the linkage with the matrix did not succeed; not classified subjects: all the episodes were not classified, or the episodes were either not exposed or not classified.

the first result. For the second coding, the questionnaires were distributed so that each coder worked on 50 questionnaires she had coded the first time and on 100 questionnaires coded by the other two coders, so that we could analyse the intra- and inter-coder variability.

After verifying the consistency of identification numbers, dates of birth, and beginning and end dates of job episodes, we excluded 25 subjects from the analysis (subjects present in only one coding series, identification number error).

In both series, we computed the frequencies of the activity sector (selected on the first digit of the ISIC code), the activity categories (first two digits of ISIC code), main occupational groups (first digit of the ISCO code), and occupational subgroups (first two digits of the ISCO code). For each of these categories, we compared the frequency of occurrence ranking of ISIC and ISCO codes between the two series by means of the Spearman rank correlation coefficient.¹⁰

The job history thus coded was then linked with an asbestos specific JEM to assess asbestos exposure for each job episode of each subject. Its job axis combines ISCO and ISIC codes. This JEM provides for each ISCO/ISIC combination: (1) an index of the probability of asbestos exposure (non-exposed, 0.05, 0.3, 0.5, 0.7, 1); (2) an index of the frequency of exposure; and (3) an index of the intensity of exposure. When significant technical or protective changes occurred that were likely to modify exposure conditions, the corresponding date was included in the matrix, and the exposure indices modified; thus, the JEM might contain several periods for the same ISCO/ISIC combination. This JEM includes 10 625 different ISCO/ISIC combinations, including those for different periods. The JEM was elaborated by experts⁷ and was already used to assess the asbestos exposure in different epidemiological studies or to validate other assessment methods.¹¹⁻¹⁴ The exposure status of each episode was defined as "exposed" if the matrix assigned a non-zero probability, as "non-exposed" if the probability is zero, and as "not classified" if the linkage with the matrix did not succeed. The exposure status for each subject was defined as "exposed" if at least one job episode was classified with a non-zero probability. A lifetime exposure probability was assigned to each subject: it corresponded to the episode with the highest probability.

We also studied the agreement of the exposure assessment results in both series, for each job episode and each subject, with the unweighted kappa reproducibility coefficient (κ), which quantifies the reproducibility of qualitative variable measurements.¹⁵⁻¹⁶ Moreover, the coder effect was studied more specifically by analysing, for individual coders and for pairs of coders, the coding agreement, the resulting exposure, and probability of exposure.

Finally, the asbestos exposure classification by the matrix in each series was compared with the results obtained by the trained physicians who interviewed the subjects.

We used SAS software for the analysis.

RESULTS

The final analysis involved 425 subjects with 2099 job episodes. Overall, the questionnaires of 146 subjects (34.4%), corresponding to 755 episodes (36%), were treated twice by the same person, and the others (279 subjects and 1344 episodes) coded by two different coders.

Occupations and activity sectors

The analysis of the coding differences shows that 1054 episodes (50.2%), corresponding to 354 subjects (83.2%), included at least one coding difference (that is, the codes used in the two series were different, regardless of the digit which differed) for activity sector (ISIC) or occupation (ISCO), while 278 episodes (13.2%) involving 137 subjects (32.2%) differed for both variables. The coding differences for activity involved 30.8% of the episodes (64.9% of subjects), and those for occupations 32.7% of the episodes (65.1% of subjects).

No ISIC code was assigned for 6.8% of the job episodes in the first coding series and 6.4% in the second. Two main industries (first ISIC digit) represented more than half the ISIC codes used in all the questionnaires: these were manufacturing (37.5% in both series) and construction and public works (22.4% in the first series and 23.4% in the second). The frequency rankings for the activity sector codes were very close (Spearman correlation coefficient: 0.98; $p \leq 0.01$). The calculation for the ranking of specific activity category (two first ISIC digits) yielded the same Spearman coefficient ($r = 0.98$).

The main occupational group (first ISCO digit) coded most frequently included plant and machine operators and assemblers (including drivers): they accounted for 71.8% and 71.4% of the ISCO codes used in both series. They were followed by skilled agricultural and fishery workers (6.3% and 6.1%), professionals (5.9% and 5.4%), clerks (4.2% and 4.6%), and service workers and shop and market sales workers (3.8% and 4%). In both series, no code was assigned in 3.1% of the episodes. The frequency rankings for the main occupational groups and for the occupational subgroups (first two ISCO digits) were also very close (respective Spearman correlation coefficients: 0.97, $p \leq 0.01$; and 0.93, $p \leq 0.01$).

Asbestos exposure

The asbestos exposure in job episodes and subjects in the two coding series after linkage with the JEM are shown in table 1 and the kappa reproducibility coefficients for exposure variables according to coder pairs are shown in table 2.

In 81.7% of the job episodes, the asbestos exposure assessment did not change from the first to the second series. The kappa was 0.69 for the episode exposure status. When we

Table 2 Kappa coefficient of exposure variables according to coder pairs

	Episodes n	Subjects n	Kappa			
			Episode exposure status	Episode exposure probability	Subject exposure status	Subject exposure probability
General	2099	425	0.69	0.66	0.64	0.60
<i>Same coder</i>	755	146	0.78	0.77	0.70	0.66
Coder 1 for both coding series	227	49	0.80	0.75	0.86	0.66
Coder 2 for both coding series	277	49	0.79	0.77	0.53	0.60
Coder 3 for both coding series	251	48	0.75	0.77	0.66	0.72
<i>Different coders</i>	1344	279	0.64	0.60	0.62	0.57
Pair 1–2	464	91	0.61	0.57	0.60	0.49
Pair 1–3	446	92	0.61	0.60	0.58	0.60
Pair 2–3	434	96	0.69	0.62	0.65	0.62

Table 3 Proportions of coding differences by coder pairs (same coder or different coders)

	Questionnaires with coding differences				
	Total	At least one difference for ISIC or ISCO	Difference for ISIC	Difference for ISCO	Difference for ISIC and ISCO
<i>Questionnaires coded twice by the same coder</i>					
Episodes	755	272	150	186	64
(%)		(36)	(19.9)	(24.6)	(8.5)
Subjects	146	109	83	84	37
(%)		(74.6)	(56.8)	(57.5)	(25.3)
<i>Questionnaires coded once by two different coders</i>					
Episodes	1344	782	496	500	214
(%)		(58.2)	(36.9)	(37.2)	(15.9)
Subjects	279	245	193	193	100
(%)		(87.8)	(69.2)	(69.2)	(35.8)

analysed the concordance of the exposure probability levels by episode, the reproducibility coefficient was 0.66. The percentage of agreement varied with the exposure probability: 85% of the episodes classified as non-exposed in the first series were classified as non-exposed in the second and 80% of exposed episodes with probability superior or equal to 0.7 in the first series were therefore classified as exposed. The agreement decreased when probability was 0.3 (73.6%) and when it was 0.05 (68.5%).

In the most frequent occupational group (plant and machine operators and assemblers), it was 0.65 (details not shown), but we observed differences between corresponding occupational subgroups ($\kappa = 0.84$ for painters, 0.76 for builders and carpenters, 0.35 for metal manufacturing and fashioning workers, and 0.03 for food and drink workers). In the other occupational groups, we observed variations to the reproducibility coefficient for the episode exposure probability (details not shown): kappa was very high ($\kappa = 0.81$) for specific occupations ($50\,000 \leq \text{CITP} < 60\,000$, services workers) but low ($\kappa = 0.29$) for less specific occupations ($20\,000 \leq \text{CITP} < 30\,000$, directors and administrative managers).

Finally, 84.7% of the subjects had the same exposure status in both coding series (table 1). Status changed from exposed to unexposed (or vice versa) in 4.5% of the cases. Of the subjects classified as exposed during the first series, 95% were also classified as exposed in the second, while 2% were classified as not exposed and 3% were not classified. For the subjects classified as non-exposed in the first series, 64.7% retained the same exposure status in the second series, 19.1% were classified as exposed, and 16.2% could not be classified. The kappa reproducibility coefficient for the subjects' exposure status was thus 0.64, and that for their exposure probability 0.60 (table 2).

The distribution of coding differences according to coder pairs (table 3) showed that 36% of the episodes (74.6% of subjects) coded twice by the same coder had at least one difference, while 58.2% of those (87.8% of subjects) coded by two different coders differed. Simultaneous differences for both codes (ISIC and ISCO) were found for 8.5% of the episodes (25.3% of subjects) coded by the same person and 15.9% of the episodes (35.8% of subjects) coded by two different people.

Reproducibility coefficients were also calculated with the coders taken into account (table 2). The coefficient of reproducibility for all the questionnaires treated twice by the same coder was 0.78 for episode exposure status, 0.77 for episode exposure probability, 0.70 for subject exposure status, and 0.66 for subject exposure probability. We noted differences in reproducibility according to the coders, in particular for subject exposure status, with coefficients ranging from 0.86 for coder 1 to 0.53 for coder 2.

When looking at the questionnaires treated by different pairs of coders, reproducibility was 0.64 for episode exposure status, 0.60 for episode exposure probability, 0.62 for subject exposure status, and 0.57 for subject exposure probability. The coefficients of reproducibility calculated according to coder pairs varied from 0.57 to 0.69 for episode exposure and 0.49 to 0.65 for subject exposure.

We were able to compare the matrix classification with the physician's assessment (after interview) of those subjects who came for a medical examination (table 4). In the first coding, 29.5% of the subjects classified as exposed by the matrix were judged to be non-exposed by the physician ($n = 31$) and in the second coding, there were 31.8% ($n = 35$).

Similarly, we compared the matrix classification and the classification reported by the subjects in the questionnaires

Table 4 Subjects' asbestos exposure status according to the matrix and according to the physician in both coding series

		Exposure according to physician			
		Not classified	Exposed	Not exposed	Total
<i>Exposure according to matrix</i>					
Not classified	1st series	3	5	6	14
	2nd series	1	5	2	8
Exposed	1st series	15	59	31	105
	2nd series	17	58	35	110
Not exposed	1st series	0	0	5	5
	2nd series	0	1	5	6
Total	1st series or 2nd series	18	64	42	124

Table 5 Subjects' asbestos exposure status according to self report and according to the matrix in both coding series

		Exposure according to subject				
		Not completed	Exposed	Not exposed	Doesn't know	Total
<i>Exposure according to matrix</i>						
Not classified	1st series	0	18	18	18	54
	2nd series	0	14	18	16	48
Exposed	1st series	3	99	117	84	303
	2nd series	3	102	126	86	317
Not exposed	1st series	2	3	45	18	68
	2nd series	2	4	36	18	60
Total	1st series or 2nd series	5	120	180	120	425

(table 5). In the first coding, 65% of the subjects who reported that they had not been exposed to asbestos were finally classified as exposed by the matrix ($n = 117$), while 2.5% of the subjects who stated they were exposed were classified as non-exposed by the matrix ($n = 3$). In the second coding, these figures were 70% ($n = 126$) and 3.3% ($n = 4$), respectively.

DISCUSSION AND CONCLUSION

In this study, we compared the results of two series of coding of the occupations and activity sectors of a sample of men. More than half the episodes, corresponding to more than 83% of the subjects, were coded differently between the two series. This high proportion of differences is attenuated by the analysis of the distribution of sectors (first ISIC digit) and categories of activity (first two ISIC digits) and the main groups (first ISCO digit) and subgroups (first two ISCO digits) of occupations; we found a similar distribution in both series since the Spearman rank correlation coefficients were very high (≥ 0.93). Moreover, only 13.2% of the episodes (32.2% of the subjects) differed for both activity and occupation. These coding differences are often minimal, because many involve just the last two digits of the codes used for a detailed definition of the occupation, but not systematically involving any change in the exposure evaluation: most of the episodes (81.7%) and subjects (84.7%) had the same exposure status (exposed, not exposed, or not classified) in both series and a change from exposed to non-exposed status or a change from non-exposed to exposed status affected only 4.5% of all the study subjects and 5.4% of those for whom a coding difference

was found. The unweighted kappa coefficient of reproducibility was 0.64 for subject exposure status, a figure indicating satisfactory agreement. The agreement for episode exposure is also acceptable ($\kappa = 0.69$). The analysis of the exposure probability agreement also indicates satisfactory results since $\kappa = 0.66$ for the episodes and 0.60 for the subject (maximum episode probability).

Inter-coder differences were more common than intra-coder differences: 58.2% of the episodes coded by two people had at least one difference, compared with 36% for those coded by the same person; the respective figures for at least one difference for both codes were 15.9% and 8.5%. This difference was also observed for the coefficients of reproducibility, which were higher for the questionnaires coded both times by the same person: κ ranged from 0.66 to 0.78, depending on the exposure variables and tended towards very good agreement. For two coders, κ ranged from 0.57 to 0.64, a less satisfactory result. Similarly, the analysis by pairs of coders showed good agreement for the episodes coded by the same coder (coder 1, 2, or 3) (κ ranged from 0.75 to 0.80 for exposure status and from 0.75 to 0.77 for exposure probability), while the agreement for episodes measured by pairs of coders (1–2, 1–3, and 2–3) was systematically lower (κ ranged from 0.57 to 0.69). The reproducibility of subject exposure status varied more according to the specific pair of coders: κ ranged from 0.53 (pair 2–2) to 0.86 (pair 1–1) for the same person, and from 0.58 (pair 1–3) to 0.65 (pair 2–3) for two different coders.

This inter-coder variability in the consistency of exposure assessments has been observed in other studies for exposure

assessments. A population based case-control study in Montreal showed similar results. When comparing the exposure status to several occupational chemicals as assessed by two groups of experts, one belonging to the core research team, the other being external, the authors found a better reproducibility between internal experts ($\kappa = 0.63$), than between internal and external experts ($\kappa = 0.55$)¹⁷; in another comparison between exposure assessments given by the internal expert group, and a second performed four years later by the same group, they found that the reproducibility was better when exposure was classified in two levels ($\kappa = 0.73$) than when it was classified in four levels ($\kappa = 0.67$).¹⁸ Another case-control study, which examined the association between a neurological disease and occupational exposure to metals and used industrial hygienists to assess exposure to different metals, tested double coding and found that consistency was higher when the same hygienist made an assessment twice (κ ranged from 0.57 to 0.26, depending on the metal, for the same hygienist, and from 0.49 to 0.15 for two different people).¹⁹ In a different field of research, where exposure assessment is also an important concern, a nested case-control study that examined the association between the consumption of Baltic Sea fish and the risk of a low birthweight child calculated a weighted kappa to assess the quality of the fish consumption data collection. Two surveys conducted a year apart showed that the weighted kappa ranged from 0.50 to 0.53 when two different surveyors questioned the subjects and reached 0.87 when the same investigator questioned them twice.²⁰

Another approach to assessing coding quality is to compare self reported asbestos exposure with that determined by the matrix in two coding series. Our results were similar in both series. We noted a high percentage of people who reported they had not been exposed to asbestos but whom the matrix considered to be possibly exposed (65% and 70% in each series); this result shows that some workers are ignorant of their own exposure, an observation made in other studies.²¹ In the ESPACES study, we observed that the disagreement between the subjects and the matrix decreased while the probability of exposure increased. Among the subjects classified as exposed by the matrix with a low probability (0.05), 15.7% have declared that they have been exposed, whereas the percentage reached 67% for subjects exposed with the maximal probability given by the matrix.

All the subjects who have declared that they have been exposed during their job history and the subjects classified as exposed by the JEM with a probability higher than 0.05, were invited for an interview with a physician specially trained for this exercise. Only 124 among the 303 eligible persons came to the health centre for the interview.

During the interview, the exposure circumstances were discussed between the subject and the physician. A comparison of subject exposure according to the JEM in two coding series, with expert physicians as the reference, found that 56.2% of the subjects reported as possibly asbestos exposed were confirmed in the first series and 52.7% in the second. Agreement between the JEM and the physicians varied depending on the exposure probability attributed by the matrix: in the first coding series, the percentages of concordance were 42% when the JEM probability of exposure was 0.05, 64.5% when probability was 0.7, and 100% when probability was 1; similar results were observed for the second series of coding. These differences between evaluation by experts and by the matrix probably have something to do with our choice to favour matrix sensitivity—all subjects with a non-zero probability in at least one job episode were classified as exposed—and with the diversity of circumstances related to asbestos exposure and to the difficulty of apprehending it. A general population case-control study of occupational exposure to formaldehyde and wood dust showed that this agreement varied greatly according to the particular exposure stud-

ied: it was clearly better for wood dust, which is a less heterogeneous exposure than formaldehyde.²²

Three coders conducted the coding and double coding: for all three, it was their first experience doing such coding, although they had all undergone specific training. To alleviate this handicap, the selection of questionnaires for the double coding was made more than two months after the coding process began, so that the coders could gain experience. Each coding series of 150 questionnaires per coder was conducted over a three day period, for 50 questionnaires per day per coder, a relatively high productivity in view of the complexity of the classifications used. We found missing data to be twice as frequent for activity sectors (6%) as for occupations (3%). This may be explained by the fact that the subjects sometimes neglected to specify their employer's activity sector, although they usually completed the question about their occupation. The activity sector was easy to establish for large industrial companies, but in some cases, additional research, based on the employer's address in particular, was required. The missing occupation codes involved either self administered questionnaires that had not been filled out completely or very specific job titles that could not be coded with the international classification system we used. Accordingly, the code for unspecified labourers (ISCO = 99910) was used fairly frequently in both series (8.3% and 8%); in particular, it was often systematically used to code occupations for which insufficient information was available.

The use of national codes (PCS for occupation and NAF for activity sector),^{23, 24} better adapted to this type of study and allowing greater precision, would certainly have limited some of these problems, but would not have been compatible with the matrix we used, which was constructed on the basis of the international classifications.

Finally, it appears that, despite intra- and inter-coder differences, the coding of job episodes was satisfactory, since the agreement between the two series was adequate; we thus conclude that the subjects' asbestos exposures were assessed without major distortions. Overall, the methodological difficulties encountered in the coding stage do not appear to have overly influenced the results for asbestos exposure. Moreover, the interview with a trained reference physician allowed verification of the job history coding and validation of the exposure assessment provided by the JEM; we were thus able to limit the extent of misclassification generally associated with this method.^{7-9, 25, 26} One particularly interesting finding from this work comes from the observation of substantial differences between the coders in individual reproducibility coefficients. This indicates that adequate training of specialist coders is likely to improve the quality of coding for activity sectors and occupations. We can therefore recommend the use of specialist coders in epidemiological studies in general populations that use a job-exposure matrix method, to improve the quality of the exposure data. Our results suggest also that the quality of exposure assessment in population based studies could be ameliorated by the use of interview to obtain subjects' job histories rather than self administered questionnaires.

The ESPACES pilot study was intended to identify retirees who may have been exposed to asbestos during their working lives. For this objective, we decided to tune the JEM to its maximal sensitivity, regarding as potentially exposed each person with a non-zero probability of exposure, and for the other steps of the pilot study, we used only the ISIC-ISCO codes yielded by the first coding procedure. The aim of this work was to validate the method used before its extension to a larger population.

ACKNOWLEDGEMENTS

The authors thank Drs Dominique Coste, Patrick Lépinau, Xavier Pagnon, Brigitte Varsat, Bertrand Wadoux, and M Jean-François Meyer for

Main messages

- More than half the episodes (more than 83% of the subjects) were coded differently between the two series, but most of the episodes (81.7%) and subjects (84.7%) had the same exposure status (exposed, not exposed, or not classified).
- The unweighted kappa coefficients of reproducibility were 0.64 for subject exposure status, 0.69 for episode exposure, 0.66 for episode exposure probability, and 0.60 for subject episode probability.

Policy implication

- To improve the quality of the exposure data in general population based epidemiological studies that use a job-exposure matrix, the coding of jobs and activity sectors should be performed by highly experienced coders.

their collaboration in the health examination centres, Mrs Joëlle Févotte for her participation in coder training, and also the coders.

.....

Authors' affiliations

C Pilorget, M Goldberg, S Bonenfant, A Schmaus, INSERM Unité 88-IFR 69, 14 rue du Val d'Osne, 94415 Saint-Maurice, France
E Imbernon, Institut de Veille Sanitaire (InVS), Département Santé Travail, 94415 Saint-Maurice, France
Y Spyckerelle, B Fournier, J Steinmetz, Centre technique d'appui et de formation des centres d'Examens de Santé (CETAF), 54500 Vandoeuvre lès Nancy, France

REFERENCES

- 1 **Gérin M**, Siemiatycki J, Kemper H, *et al*. Obtaining occupational exposure histories in epidemiologic case-control studies. *J Occup Med* 1985;**27**:420-6.
- 2 **InVS**, CETAF, INSERM, CNAM. *Enquête pilote ESPACES. Identification et suivi médical post-professionnel des salariés retraités ayant été exposés à l'amiante. Place et rôle des Centres d'Examens de Santé*. Institut de Veille Sanitaire, France: Rapport, 2001.
- 3 **Spyckerelle Y**, Imbernon E, Fournier B, *et al*. Identification et suivi post-professionnel des salariés retraités ayant été exposés à l'amiante (ESPACES). *Revue Médicale de l'Assurance Maladie* 2000;**3**:29-36.
- 4 **Orlowski E**, Pohlabein H, Berrino F, *et al*. Retrospective assessment of asbestos exposure—II. At the job level: complementarity of job-specific questionnaire and job-exposure matrices. *Int J Epidemiol* 1993;**22**(suppl 2):S96-S105.
- 5 **International Labour Office**. *International Standard Classification of Occupations (ISCO)*. Geneva: International Labour Office, 1968.
- 6 **United Nations Statistical Office**. *International Standard Industrial Classification of all economic activities (ISIC)*. New York: United Nations Statistical Office, 1975.
- 7 **Goldberg M**, Kromhout H, Guénel P, *et al*. Job exposure matrices in industry. *Int J Epidemiol* 1993;**22**(suppl 2):S10-S15.
- 8 **Bouyer J**, Hémon D. Studying the performance of a job exposure matrix. *Int J Epidemiol* 1993;**22**(suppl 2):S65-S71.
- 9 **Bouyer J**, Hémon D. Les matrices Emplois-Expositions. *Revue d'Epidémiologie et de Santé Publique* 1994;**42**:235-45.
- 10 **Zar JH**. Simple linear correlation. In: *Biostatistical analysis*, 2nd edn. Englewood Cliffs, NJ: Prentice-Hall, 1984:306-27.
- 11 **Iwatsubo Y**, Paireon JC, Boutin C, *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**:193-8.
- 12 **Goldberg M**, Bancaei A, Goldberg S, *et al*. Past occupational exposure to asbestos among men in France. *Scand J Work Environ Health* 2000;**26**:52-61.
- 13 **Luce D**, Leclerc A, Bégin D, *et al*. Sinonasal cancer and occupational exposures: a pooled analysis of 12 case-control studies. *Cancer Causes Control* 2002;**13**:147-57.
- 14 **Marchand JL**, Luce D, Leclerc A, *et al*. Laryngeal and hypopharyngeal cancer and occupational exposure to asbestos and man-made vitreous fibers: results of a case-control study. *Am J Ind Med* 2000;**37**:581-9.
- 15 **Fleiss JL**. *The measurement of interrater agreement*. In: *Statistical methods for rates and proportions*. New York: Wiley-Intersciences, 1981:212-36.
- 16 **Fermanian J**. Mesure de l'accord entre deux juges. Cas qualitatif. *Revue d'Epidémiologie et de Santé Publique* 1984;**32**:140-7.
- 17 **Goldberg MS**, Siemiatycki J, Gérin M. Inter-rater agreement in assessing occupational exposure in a case-control study. *Br J Ind Med* 1986;**43**:667-76.
- 18 **Siemiatycki J**, Fritschi L, Nadon L, *et al*. Reliability of an expert rating procedure for retrospective assessment of occupational exposures in community-based case-control studies. *Am J Ind Med* 1997;**31**:280-6.
- 19 **Rybicki BA**, Peterson EL, Johnson CC, *et al*. Intra- and inter-rater agreement in the assessment of occupational exposure to metals. *Int J Epidemiol* 1998;**27**:269-73.
- 20 **Rylander L**, Strömberg U, Hagmar L. Agreement between reported fish consumption obtained by two interviews and its impact on the results in a reproduction study. *Eur J Epidemiol* 1998;**14**:93-7.
- 21 **Fletcher AC**, Engholm G, Englund A. The risk of lung cancer from asbestos among Swedish construction workers: self-reported exposure and a job exposure matrix compared. *Int J Epidemiol* 1993;**22**(suppl 2):S29-S35.
- 22 **Luce D**, Gérin M, Berrino F, *et al*. Sources of discrepancies between a job-exposure matrix and a case by case expert assessment for occupational exposure to formaldehyde and wood-dust. *Int J Epidemiol* 1993;**22**(suppl 2):S113-S120.
- 23 **Institut National de la Statistique et des Etudes Economiques (INSEE)**. *Nomenclature des professions et catégories socioprofessionnelles (PCS)*. Paris: INSEE, 1994.
- 24 **Institut National de la Statistique et des Etudes Economiques (INSEE)**. *Nomenclature d'Activités Françaises (NAF)*. Paris: INSEE, 1994.
- 25 **Moulin JJ**, Romazini S, Lasfargues G, *et al*. Elaboration d'une matrice Emplois-Expositions dans l'industrie productrice de métaux durs en France. *Revue d'Epidémiologie et de Santé Publique* 1997;**45**:41-51.
- 26 **Imbernon E**, Goldberg M, Guénel P, *et al*. Validation of asbestos exposure assessment in a job-exposure matrix in the electricity and gas industry in France: the MATEX project. *Occup Hyg* 1996;**3**:193-8.