Annals of Work Exposures and Health, 2022, Vol. 66, No. 6, 815–821
https://doi.org/10.1093/annweh/wxac009
Advance Access publication 25 February 2022
Short Communication

Short Communication

Development of a Crosswalk to Translate Italian Occupation Codes to ISCO-68 Codes

Andrea Spinazzè[1,*], Dario Consonni[2], Francesca Borghi[1], Libero Andrea Mazzucchelli[1], Sabrina Rovelli[1], Andrea Cattaneo[1], Carolina Zellino[2,3], Barbara Dallari[2], Angela Cecilia Pesatori[2,4], Hans Kromhout[5], Susan Peters[5], Luciano Riboldi[2], Carolina Mensi[2] and Domenico Maria Cavallo[1]

[1]Department of Science and High Technology, University of Insubria, Como, Italy; [2]Occupational Health Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; [3]Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [4]Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy; [5]Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands

*Author to whom correspondence should be addressed. Tel: +39 031 2386629; e-mail: andrea.spinazze@uninsubria.it

Submitted 23 August 2021; revised 30 October 2021; editorial decision 24 January 2022; revised version accepted 16 February 2022.

Abstract

In occupational epidemiology, job coding is an important—but time-consuming—step in assigning exposure. We implemented a tool (i.e. a crosswalk) to translate occupation codes from the Italian (ISTAT-CIP-91, $n = 6319$ five-digit job codes) to the International Standard Classification of Occupations (ISCO-68, $n = 1881$ five-digit job codes). The former is currently used in Italy for various purposes (e.g. in the National Mesothelioma Registry). The latter has been used in several studies on occupational cancers because it facilitates communication of results to the scientific community and, most importantly, because some job exposure matrices (JEMs) are based on international codes. Three authors created a table containing the crosswalk structure, providing an interpretation for each of the ISTAT-CIP-91 codes job descriptions and then manually recoding them according to ISCO-68. Two other authors independently revised it. The performance of the final version was assessed by comparison with results obtained by manual ISCO-68 coding performed in two previous case-control studies on asbestos and mesothelioma. More specifically, the automatically obtained ISCO-68 codes were merged with a JEM (DOM-JEM). The resulting individual asbestos exposure estimates (ever versus never exposed) were compared to those originally obtained (using the same DOM-JEM) from manual translation of ISTAT-CIP-91 to ISCO-68 (considered as the 'gold standard'). In the first study, among 159 peritoneal mesothelioma cases (400 job codes), Cohen's kappa was 0.91, sensitivity 0.95, and specificity 0.96. In the second study, among 716 pleural mesothelioma cases and controls (4400 job codes) kappa was 0.86, sensitivity 0.94, and specificity 0.91. Performance was better among in

What's Important About This Paper?

Job coding is a critical step in the development of job exposure matrices for epidemiologic studies and is made more complicated by the use of different job classification and coding schemes around the world. This study addresses this challenge by developing a crosswalk to translate job codes used in Italy, ISTAT-CIP-91, into the International Standard Classification of Occupations, ISCO-68. Using the example of mesothelioma, the crosswalk was found to be robust in the classification of jobs among men and women, and among cases and controls.

women. For men, performance was lower among cases than among controls (kappa 0.70, sensitivity 0.95, specificity 0.72 versus kappa 0.87, sensitivity 0.97, and specificity 0.92). In conclusion, the proposed tool allowed a rapid translation of thousands of job codes with good to excellent accuracy. The table containing ISTAT-CIP-91 codes and job descriptions and the corresponding ISCO-68 codes and job descriptions is made publicly available and can be freely used for epidemiological analyses in Italy and international collaborations.

Keywords: epidemiology; job coding; job exposure matrix (JEM); occupational exposure; retrospective exposure assessment

Introduction

Background

In the framework of occupational hygiene, occupational health and epidemiology, exposure assessment (qualitative or quantitative) is needed to investigate the relationship between the exposure to a risk agent and the likelihood of disease (and thus to obtain an appropriate risk characterization) (Sahmel *et al.* 2010). Different methodological approaches are available for exposure assessment, depending on the target. In general, if the interest is directed towards personal exposure, the level of exposure should be measured or estimated at the point of contact with the subject. Then, obtained results should be interpreted in the light of scenario-specific contextual information and related to the duration of the exposure, to refine the estimate. The exposure estimate can be carried out with a prospective (estimates of present or future exposures; typically used for regulatory purposes) or retrospective (reconstructions of past exposures) approach. The retrospective exposure assessment (REA) is necessary for chronic (i.e. long-latency) diseases like cancer. Direct measurement of the exposure of each subject for the entire duration of the period of interest would be the most accurate method for assessing exposure, but this approach is rarely feasible due to the lack of adequate past measurements. Therefore, epidemiological studies on occupational diseases often require an estimate of past exposure through REA in relation to industry, occupation, and specific tasks, ideally for the entire working life (Sahmel *et al.* 2010; Borghi *et al.* 2020). REA approaches vary according to the purpose of the study and data availability, from simple classification into 'exposed' and 'unexposed' workers to more sophisticated assessments, involving use of statistical models to provide quantitative (or semi-quantitative) estimates of exposure for each job performed by the worker. Finally, individual estimates of exposure are produced. These may be dichotomous (e.g. ever-never), ordinal or quantitative (e.g. duration of exposure, cumulative exposure) estimates. Job exposure matrices (JEMs) and individual expert exposure assessment (IEEA) can be regarded as classical REA methods (Borghi *et al.* 2020). IEEA is a commonly adopted practice for the estimation of past exposure in relatively small studies. IEEA relies on the knowledge and the experience of the assessor, which assigns a level of exposure based on the analysis of a person's occupational history and job-specific questionnaires. However, IEEA applicability in large population (cohort or case–control) studies may be limited, since IEEA is a time- and resource-demanding process (Johnson *et al.* 2017). JEMs represent an efficient and reproducible methodology, that allows obtaining exposure assessment free of differential misclassification with respect to the health outcome. A JEM typically assigns an estimated exposure (dichotomous, ordinal or quantitative) to each job code; some JEMs include a temporal axis in addition to job codes. This allows to easily translate job histories into exposure history. A JEM, once defined and validated, is relatively easy to apply and much less costly than IEEAs (Peters 2020). In fact, JEMs have been applied in a wide range of settings in large-scale studies (Ge *et al.* 2018).

Problem statement

The high level of standardization and ease of use represent an advantage and make JEMs useful for large-scale studies and in compensation and surveillance efforts in occupational health, while at the same time represent drawbacks (Fadel *et al.* 2020; Peters 2020). More in detail, other than the lack of complete validation, JEMs generally fail to characterize the inter-individual variability and heterogeneity of exposure of different workers classified in the same job title, or temporal variations in exposure levels (Peters 2020), but given a group-based approach and the associated Berkson-type error, it will result in no or little bias of risk estimates but with loss of precision(Armstrong 1998). However, coding of job titles is a major challenge. Manual coding of each individual's job history is time-consuming, so that many large-scale studies could not apply JEMs (Peters 2020; Savic *et al.* 2021). In this regard, efforts in developing automatic coding have shown quite low agreement [in the order of 50%, Cohen's Kappa (κ = 0.4–0.8) with manual coding] (Burstyn *et al.* 2014; Russ *et al.* 2016). To overcome this methodological issue and facilitate the use of JEMs, standardized and validated methods have been proposed to translate individual's job history in a format compatible with the use of JEMs (e.g. standardized occupational coding systems such the ILO—International Labour Organization 2021). Several computer-based tools have been developed for this purpose (Patel *et al.* 2012; Russ *et al.* 2016; De Matteis *et al.* 2017; Rémen *et al.* 2018; Savic *et al.* 2021), based on different premises and technologies and with different objectives. Generally, the available tools cover only one national job classification and use English as language. The only exceptions are 'CAPS-Canada' (Rémen *et al.* 2018) (Language: English and French; applied to seven International, Canadian and US classifications), 'CASCOT' (Computer Assisted Structured Coding Tool; Languages: Arabic; Chinese; Dutch; English; Finnish; French; German; Hindi; Indonesian; Italian; Portuguese; Romanian; Russian; Slovak; Spanish) (CASCOT 2018) and 'Procode' (Savic *et al.* 2021) (Language: English, German, French and Italian; applied to seven International classifications). It must be noted that automatic translations of job coding may be useful whenever there is the need to adapt a large database to a different classification system. However, since job information is only available as a code (not as a free-text entry), automated coding systems is not applicable to most registry data. For this reason, a crosswalk (a tool capable of translating one code to another) would be very helpful (Koeman *et al.* 2013). Translations can be a solution for pooling different coding systems, but this may also introduce further disagreement between codes and jobs, possibly leading to further misclassification of exposure (Peters 2020) although this bias has been shown to be small in previous studies (Kromhout and Vermeulen 2001; Koeman *et al.* 2013; Burstyn *et al.* 2014; Petersen *et al.* 2020).

Aim of the study

The aim of the study was to implement a tool (i.e. a crosswalk) to translate occupation codes from the Italian ISTAT-CIP-91 coding system (ISTAT—Istituto Nazionale di Statistica 1991) into the International Standard Classification of Occupations 'ISCO-68' coding system (ILO—International Labour Organization 2021). This tool would allow (i) to communicate results to the scientific community using international rather than country-specific classifications and (ii) facilitate the use of JEMs, many of which are based on international codes. The particular interest in developing a translation tool between these two coding systems is due to the fact that ISTAT-CIP-91 is one of the most widely used coding system for job histories description in Italy, especially in epidemiological surveillance networks of occupational diseases (ISPESL 2003). Furthermore, ISCO-68 was chosen for this study despite the availability of newer versions because (i) ISCO-68 served as a model developing or revising several national occupational classifications; (ii) latest ISCO versions are oriented more toward job status and less toward job tasks; (iii) ISCO-68 has been widely used in epidemiologic studies; and (iv) is more appropriate for historical exposures than more recent classifications (Ahrens and Merletti 1998). ISCO-68 has also been implemented in widely used JEMs such as SYN-JEM (Peters *et al.* 2013, 2016) and DOM-JEM (Peters *et al.* 2011).

Materials and methods

Job coding systems

The Italian ISTAT-CIP-91 coding system (ISTAT—Istituto Nazionale di Statistica 1991) includes 9 Major Groups (first digit), 35 Groups (second digit), 119 Classes (third digit) and 599 Categories (fourth digit) and more than 5000 subcategories (fifth digit): within the structure just outlined there are 6319 job codes (ISTAT—Istituto Nazionale di Statistica 2013b). ISCO-68 (ILO—International Labour Organization 2021) is made up of 8 major groups (first digit), 83 groups (second digit), 284 occupational units (third digit) and 1506 occupations (fourth and fifth digit). This version includes the definitions of each of the 1881 groupings and describes the general functions assumed by the professions, as well as the main tasks performed by the respective workers.

Coding translation and crosswalk implementation

A first step was performed by three of the authors (A.S., F.B., L.A.M.) to translate the ISTAT-CIP-91 classification to the ISCO-68 coding system. The three authors divided equally the number of ISTAT-CIP-91 codes to be translated and provided for an interpretation of the job descriptions and then to manually recode each code according to ISCO-68. The crosswalk basic structure consists of a table in which ISTAT-CIP-91 codes and job descriptions were paired with the corresponding ISCO-68 codes and job descriptions. The coding activity was carried out blindly by one author with respect to the other. Subsequently, a first verification step was carried out by the same three authors, each of whom evaluated the translation of the other two, thus ensuring a triple check of the translation work. In this phase, any discrepancies in interpretation/translation of job description were reported. Each identified discrepancy was discussed by the three authors, and the final translation to ISCO-68 coding was decided upon by a majority decision.

A second round of verification was carried out by two of the authors (D.C., C.M.), to compare and revise the coding performed for this study with that of manual coding of job histories for two previous studies. The first was a multicentre population-based case–control study on pleural mesothelioma performed in five Italian regions including Lombardy (unpublished). The second was a population-based case-control study on peritoneal mesothelioma in Lombardy (Italy) (Consonni *et al.* 2019). In both studies, ISTAT-CIP-91 codes had been manually coded into ISCO-68. The latter codes were then merged with DOM-JEM or SYN-JEM to obtain estimates of asbestos exposure in each job to estimate an individual's exposure history.

Here, we applied the crosswalk to obtain automatically ISCO-68 codes. Then we merged these codes with the DOM-JEM. The resulting individual asbestos exposure estimates (ever versus never exposed) were compared to those originally obtained from manual coding. We considered the latter as the gold standard to calculate sensitivity and specificity of the proposed crosswalk. We also calculated Cohen's kappa coefficient of agreement. Confidence intervals (CI) were calculated with the Agresti–Coull formula. We additionally stratified analyses by case–control status and gender. For the peritoneal study ISTAT-CIP-91 codes were available only for cases. Data management and statistical analyses were performed with Stata 16 (StataCorp 2019).

Results

Results of the crosswalk application showed good sensitivity and specificity (0.91 or higher) against manual (expert-based) coding among peritoneal mesothelioma cases, overall and within genders (Table 1). Agreement was 0.91 overall, 0.86 in men, and 0.95 in women. Among cases and controls in the pleural mesothelioma study, we found high sensitivity (0.96), specificity (0.90), and agreement (0.84) of automated coding (Table 2, upper). Agreement was better for controls. In men (Table 2,

Table 1. Comparison of occupational asbestos exposure (ever versus never, according to the DOM-JEM) among peritoneal mesothelioma cases, obtained by automatic coding using ISTAT-CIP-91 to ISCO-68 crosswalk and that obtained using expert-based coding (taken as gold standard), Lombardy, Italy, period 2008–2015 (see Consonni *et al.* 2019)

Automated coding	All		Men		Women	
	Expert-based coding		Expert-based coding		Expert-based coding	
	Ever	Never	Ever	Never	Ever	Never
Ever	*59*	*4*	*48*	*3*	*11*	*1*
Never	*3*	*93*	*3*	*32*	*0*	*61*
Sensitivity	0.95		0.94		1.00	
95% CI[a]	0.86; 0.99		0.83; 0.99		0.70; 1.00	
Specificity		0.96		0.91		0.98
95% CI[a]		0.90; 0.99		0.77; 0.98		0.91; 1.00
Cohen's kappa	0.91		0.86		0.95	
95% CI[a]	0.84; 0.98		0.74; 0.97		0.85; 1.00	

CI, confidence interval.

[a]Calculated with the Agresti–Coull method.

middle), we found a similar pattern, again with a better agreement for controls. In women (Table 2, bottom), very good performance was found either among cases or among controls. The differences in agreement between sexes and between cases and controls could be most likely due to differences in number of (unexposed) jobs between men (more) and women (fewer) and cases (more) and controls (fewer). Once defined and verified, the crosswalk structure was transferred into a table (Excel spreadsheet) containing ISTAT-CIP-91 codes and job descriptions and the corresponding ISCO-68 codes and job descriptions (Supplementary Table S1, available at *Annals of Occupational Hygiene* online). This table can be freely used for epidemiological analyses. If results using this tool are published, the present paper should be acknowledged.

Discussion

The proposed coding tool provided a good performance (i.e. sensitivity, specificity, agreement with manual coding) and, overall, the crosswalk is considered reliable and applicable in epidemiological studies. The translation is based on the numerical code ISTAT-CIP-91 (five digits) and allows translation into the corresponding ISCO-68 code (five digits). This allows a unique

Table 2. Comparison of occupational asbestos exposure (ever-never, according to the DOM-JEM) among pleural mesothelioma cases and controls, obtained by automatic coding using ISTAT-CIP-91 to ISCO-68 crosswalk and that obtained using expert-based coding (taken as gold standard), Lombardy, Italy, period 2008–2015

Automated coding	All		Cases		Controls	
	Expert-based coding		Expert-based coding		Expert-based coding	
	Ever	Never	Ever	Never	Ever	Never
All						
Ever	*512*	*84*	*296*	*54*	*216*	*30*
Never	*20*	*725*	*15*	*260*	*5*	*465*
Sensitivity	0.96		0.95		0.98	
95% CI[a]	0.94; 0.98		0.92; 0.97		0.95; 0.99	
Specificity		0.90		0.83		0.94
95% CI[a]		0.87; 0.92		0.78; 0.87		0.91; 0.96
Cohen's kappa	0.84		0.78		0.89	
95% CI[a]	0.81; 0.87		0.73; 0.83		0.85; 0.92	
Men						
Ever	*447*	*77*	*262*	*52*	*185*	*25*
Never	*18*	*409*	*13*	*136*	*5*	*273*
Sensitivity	0.96		0.95		0.97	
95% CI[a]	0.94; 0.98		0.92; 0.97		0.94; 0.99	
Specificity		0.84		0.72		0.92
95% CI[a]		0.81; 0.87		0.66; 0.78		0.88; 0.94
Cohen's kappa	0.80		0.70		0.87	
95% CI[a]	0.76; 0.84		0.63; 0.77		0.83; 0.92	
Women						
Ever	*2*	*7*	*34*	*2*	*31*	*5*
Never	*65*	*316*	*2*	*124*	*0*	*192*
Sensitivity	0.97		0.94		1.00	
95% CI[a]	0.89; 1.00		0.81; 0.99		0.87; 1.00	
Specificity		0.98		0.98		0.97
95% CI[a]		0.96; 0.99		0.94; 1.00		0.94; 0.99
Cohen's kappa	0.92		0.93		0.91	
95% CI[a]	0.87; 0.97		0.86; 1.00		0.84; 0.99	

CI, confidence interval.

[a]Calculated with the Agresti–Coull method.

translation of the original description into the destination coding system, unlike systems based on the identification of the presence of keywords in free-text entries of a given classification, which instead allow identifying different solutions for the translation of a single job description and must eventually be evaluated for the final decision. At the same time, this means that the crosswalk developers have manually prepared the translation for code pairs between two classifications, attributing the most appropriate code in classification. The structure of the crosswalk also assumes that this prefixed schema will not be dynamically updated, as defined in some similar examples based on machine learning methods (Savic *et al.* 2021), but future editions could be eventually updated manually, if revisions become necessary.

The proposed crosswalk was specifically developed to cope with issues related to the recoding of subjects' work history from a national (ISTAT-CIP-91) to an international (ISCO-68) coding system. It must be noted that the translation of ISTAT-CIP-91 codes to ISCO-68 coding, implies the need to consider a reduction in the number of available codes to classify occupations (from 6319 to 1881 codes). This could lead to a reduction of the contrast in exposure, but on the other hand, the possibility of conforming to an international classification system represents an advantage for the simplification of studies within the REA, and the possibility to apply more advanced approaches (e.g. quantitative JEMs) in the Italian context.

It should be noted that, both the coding systems have been updated, to document the economic transformations and the consequent evolution of professions. Subsequent editions of the Italian official classification of professions (2001 and 2011) are based on the same classification logic of ISTAT-CIP-91 and therefore continue to fully adhere to the chosen approach at an international level (ISTAT—Istituto Nazionale di Statistica 2013b). Concerning ISCO-68, the subsequent edition (ISCO-88) is organized in 10 major groups (first digit), 28 sub-major (second digit) groups, 116 minor groups (third digit), and 390 unit groups (fourth digit) (ILO—International Labour Organization 2021). For the first time, the concepts of skill level and skill specialization are developed, which represent the backbone of the taxonomic system. The most recent ISCO classification of professions is ISCO-08, published in 2008 (ILO—International Labour Organization 2021). The conceptual model of ISCO-08 uses is fully in line with that of the previous version, based on the concept of competence, and the definitions of the ten major groups have therefore not been changed. ISCO-08 is more detailed in the lower levels of the classification hierarchy (43 sub-major groups, 131 minor groups and 425 unit groups). Interestingly, both ISTAT-CIP-91 and ISCO-68 classifications may be converted into subsequent editions (ISTAT-CIP-2001 and ISTAT-CIP-2011; ISCO-88 and ISCO-08), by means of already available tools (ISTAT—Istituto Nazionale di Statistica 2013a; ILO—International Labour Organization 2021).

Conclusions

Manual translation of codes from a classification to another is a time-consuming procedure and often is a limiting factor in the use of JEMs to estimate individual lifetime exposure in large studies. A high-sensitivity, high-specificity, and reliable tool (crosswalk) was elaborated to efficiently translate Italian job codes (ISTAT-CIP-91) to international job codes (ISCO-68), thus allowing use of JEMs based on ISCO-68 in Italian studies and in large-scale international collaborative epidemiological studies.

Supplementary data

Supplementary data are available at *Annals of Work Exposures and Health* online.

Funding

This research was funded by 'Fondazione Regionale per la Ricerca Biomedica'—'Bando Progetto Speciale 2017 su Patologie Amianto-Correlate', in the framework of the project 'An integrated precision medicine approach to malignant mesothelioma: from mutation load to epidemiology and therapy' (acronym: PRIMATE-code ARL_2/2018; Principal Investigator: Tommaso Dragani).

Author contributions

Conceptualization: A.S., D.C., and C.M.; Data curation: A.S., D.C., F.B., L.M., and C.M.; Formal analysis: A.S., D.C., and F.B.; Funding acquisition: D.C.; Investigation: A.S., D.C., F.B., S.R., C.Z., and C.M.; Methodology: A.S., D.C., F.B., and C.M.; Project administration: D.C.; Resources: D.C., A.P., H.K., S.P., L.R., C.M., and D.C.; Software: D.C., F.B., and L.M.; Supervision: D.C., A.C., C.M., and D.C.; Validation: A.S., D.C., F.B., L.M., and C.M.; Writing—original draft: A.S., F.B., and C.M.; Writing—review and editing: D.C., L.M., S.R., A.C., C.Z., B.D., A.P., H.K., S.P., L.R., and D.C.

Conflict of interest

The authors declare no conflict of interest.

Data availability

The data were derived from sources in the public domain: ILO—International Labour Organization (2021) ISCO—International Standard Classification of Occupations. https://www.ilo.org/public/english/bureau/stat/isco/.
ISTAT—Istituto Nazionale di Statistica (2013) LA CLASSIFICAZIONE DELLE PROFESSIONI (Italian). ISBN 9788845817533.

References

Ahrens W, Merletti F. (1998) A standard tool for the analysis of occupational lung cancer in epidemiologic studies. *Int J Occup Environ Health*; **4**: 236–40.

Armstrong B. (1998) Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med*; **55**: 651–6.

Borghi F, Mazzucchelli LA, Campagnolo D *et al*. (2020) Retrospective exposure assessment methods used in occupational human health risk assessment: a systematic review. *Int J Environ Res Public Health*; **17**: 6190.

Burstyn I, Slutsky A, Lee DG *et al*. (2014) Beyond crosswalks: reliability of exposure assessment following automated coding of free-text job descriptions for occupational epidemiology. *Ann Occup Hyg*; **58**: 482–92.

CASCOT. (2018) Cascot: Computer Assisted Structured Coding Tool. Available at https://warwick.ac.uk/fac/soc/ier/software/cascot/. Accessed 16 July 2021.

Consonni D, Calvi C, De Matteis S *et al*. (2019) Peritoneal mesothelioma and asbestos exposure: a population-based case-control study in Lombardy, Italy. *Occup Environ Med*; **76**: 545–53.

De Matteis S, Jarvis D, Young H, *et al*. (2017) Occupational self-coding automatic recording (OSCAR): a novel web-based tool to collect and code lifetime job histories in large population-based studies. *Scand J Work Environ Heal*; **43**: 181–6. doi:10.5271/sjweh.3613

Fadel M, Evanoff BA, Andersen JH *et al*. (2020) Not just a research method: if used with caution, can job-exposure matrices be a useful tool in the practice of occupational medicine and public health? *Scand J Work Environ Heal*; **46**: 552–3. doi:10.5271/sjweh.3900

Ge CB, Friesen MC, Kromhout H *et al*. (2018) Use and reliability of exposure assessment methods in occupational case–control studies in the general population: past, present, and future. *Ann Work Expo Heal*; **12**: 1047–063. doi:10.1093/annweh/wxy080

ILO—International Labour Organization. (2021) ISCO—International Standard Classification of Occupations. Available at https://www.ilo.org/public/english/bureau/stat/isco/.

ISPESL. (2003) *Linee guida per la rilevazione e la definizione dei casi di mesotelioma maligno e la trasmissione delle informazioni all'ISPESL da parte dei Centri Operativi Regionali*. Roma, Italy: Seconda edizione (Italian).

ISTAT—Istituto Nazionale di Statistica. (1991) *Classificazione delle professioni. Metodi e Norme Serie C*. n. 12. Rome, Italy: ISTAT—Istituto Nazionale di Statistica.

ISTAT—Istituto Nazionale di Statistica. (2013a) *Classificazione delle professioni—raccordo*. Rome, Italy: ISTAT—Istituto Nazionale di Statistica. Available at https://www.istat.it/it/archivio/18132. Accessed 9 July 2021.

ISTAT—Istituto Nazionale di Statistica. (2013b) *La Classificazione Delle Professioni*.

Johnson CY, Rocheleau CM, Hein MJ *et al*. (2017) Agreement between two methods for retrospective assessment of occupational exposure intensity to six chlorinated solvents: data from the national birth defects prevention study. *J Occup Environ Hyg*; **14**: 389–96.

Koeman T, Offermans NSM, Christopher-De Vries Y *et al*. (2013) JEMs and incompatible occupational coding systems: effect of manual and automatic recoding of job codes on exposure assignment. *Ann Occup Hyg*; **57**: 107–14.

Kromhout H, Vermeulen R. (2001) Application of job-exposure matrices in studies of the general population: some clues to their performance. *Eur Respir Rev*; **11**: 80–90.

Patel MD, Rose KM, Owens CR *et al*. (2012) Performance of automated and manual coding systems for occupational data: a case study of historical records. *Am J Ind Med*; **55**: 228–31.

Peters S. (2020) Although a valuable method in occupational epidemiology, job-exposure matrices are no magic fix. *Scand J Work Environ Heal*; **46**: 231–4.

Peters S, Kromhout H, Portengen L *et al*. (2013) Sensitivity analyses of exposure estimates from a quantitative job-exposure matrix (SYN-JEM) for use in community-based studies. *Ann Occup Hyg*; **57**: 98–106.

Peters S, Vermeulen R, Cassidy A *et al*. (2011) Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case–control study. *Occup Environ Med*; **68**: 148–53.

Peters S, Vermeulen R, Portengen L *et al*. (2016) SYN-JEM: a quantitative job-exposure matrix for five lung carcinogens. *Ann Occup Hyg*; **60**: 795–811.

Petersen SB, Flachs EM, Svendsen SW *et al*. (2020) Influence of errors in job codes on job exposure matrix-based exposure assessment in the register-based occupational cohort doc*X. *Scand J Work Environ Heal*; **46**: 259–67.

Rémen T, Richardson L, Pilorget C *et al*. (2018) Development of a coding and crosswalk tool for occupations and industries. *Ann Work Expo Heal*; **62**: 796–807.

Russ DE, Ho KY, Colt JS *et al*. (2016) Computer-based coding of free-text job descriptions to efficiently identify occupations in epidemiological studies. *Occup Environ Med*; **73**: 417–24.

Sahmel J, Devlin K, Paustenbach D *et al*. (2010) The role of exposure reconstruction in occupational human health risk assessment: current methods and a recommended framework. *Crit Rev Toxicol*; **40**: 799–843.

Savic N, Bovio N, Gilbert F *et al*. (2021) Procode: a machine-learning tool to support (re-)coding of free-texts of occupations and industries. *Ann Work Expo Heal*; **66**(1): 113–118. doi:10.1093/annweh/wxab037

StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.