© International Epidemiological Association 2003 Printed in Great Britain

International Journal of Epidemiology 2003;**32**:419–428
DOI: 10.1093/ije/dyg080

The use of occupation and industry classifications in general population studies

A 't Mannetje[1,2] and H Kromhout[2]

Accepted 15 November 2002

Occupation and industry classifications are used in epidemiological studies to classify study subjects according to their job and subsequently to study risk by job, to infer social class indicators, or to infer exposure to specific agents through job-exposure matrices. However, documentation on methodological aspects concerning the use of occupation and industry classifications is sparse within epidemiology. This paper reviews the diverse applications of occupation and industry classifications in population-based epidemiological studies. The different classifications in use are discussed, and criteria are given for choosing a classification in an epidemiological study. Finally, the reliability of coding for occupation and industry is reviewed. A further standardization of the use of occupation and industry classifications in epidemiology is recommended, in order to facilitate future comparisons between studies and fully exploit their possibilities, especially when occupational exposures are to be inferred.

Keywords Standard classifications, job codes, industry codes, job-exposure matrices, social class, general population studies

Occupation and industry classifications categorize occupations and industries into clearly defined groups. As such they provide a common basis for collecting, presenting, and comparing of labour statistics. Occupational classifications group people based on job and tasks performed, and are commonly used in sociology and population studies. Industry classifications group people based on the sector of economic activity in which they are employed and are mainly used for economic analysis.

Although not primarily developed for use in epidemiological studies, occupation classifications, and to a lesser extent industry classifications, are often used in this field. Population-based epidemiological studies frequently include questions about job title and specific tasks, after which the information is coded using either national or international classifications.

Reviewing the literature in the *British Medical Journal, American Journal of Epidemiology*, and *International Journal of Epidemiology* (published between 1995 and 2000) indicated that information on occupation in epidemiological studies (n = 129), was mostly used as an indicator for social class (38%). In 27% of the studies, occupation was studied directly in relation to disease, and in 24% occupation was used to infer occupational exposure. In the remaining 11% of the studies occupation was treated as a confounding factor or used to describe the study population.

Although widely applied in epidemiological studies, only limited methodological information is available on the use of job and industry classifications. The lack of a theoretical basis may hamper full exploitation of occupation information within epidemiology and limits the potential to optimize its reliability.

In this paper we review the potential for occupation and industry classifications in epidemiology. The main classifications available are reviewed and different methods for coding occupation and industry are discussed. In addition, we will address issues of reliability of the coding process.

Options for analysis

Labour statistics

For population-based studies, a classification can be used to describe the occupational profile of the study sample, which subsequently can be compared with that of the national population from which the study sample is drawn. The International Labour Office (ILO) provides labour statistics for most countries in the ILO yearbook of labour statistics, using the International Standard Classification of Occupations (ISCO) from 1968 and 1988 and the International Standard Industrial Classification of all Economic Activities (ISIC), Revisions 2 and 3. Statistics for major groups of these classifications are available on the Internet.[1]

[1] International Agency for Research on Cancer, Unit of Environmental Cancer Epidemiology, Lyon, France.

[2] Utrecht University, Institute for Risk Assessment Sciences, Environmental and Occupational Health Division, Utrecht, The Netherlands.

Correspondence: Andrea 't Mannetje, IARC, Unit of Environmental Cancer Epidemiology, 150 Cours Albert Thomas, 79008 Lyon, France. E-mail: mannetje@iarc.fr

For the US, national statistics are also available in the National Industry-Occupation Employment Matrix based on SIC (Standard Industrial Classification) and SOC (Standard Occupational Classification), presenting employment statistics for over 240 industries and 500 occupations.[2]

Social class

In epidemiology, occupation is most often used for distinguishing between socioeconomic groups. Whereas in fields such as sociology social class is often treated as an outcome, in epidemiology it is more often considered as a risk factor or confounder, since many diseases such as cancer appear to have a social class gradient.[3] Measures of social class can be based on occupation, education, income, or a composite of these.[4] Different scales for occupational class have been developed, of which the British Registar General's Scale is the most widely used.[4] This scale has proven to be highly predictive of inequalities in morbidity and mortality, especially among employed men. Its five categories (I professional; II managerial and technical; IIIN skilled non-manual; IIIM skilled manual; IV partly skilled; V unskilled) are based on a graded hierarchy of occupations ranked according to skill.[5] Some more recently revised occupational classifications already include a rating for occupation-related characteristics such as skill, status, or education. The ISCO-88 (International Standard Classification of Occupations) includes for each occupation a reference to one of four broad levels of formal education via the International Standard Classification of Education (ISCED).[6] O*NET, the US Occupational Information Network,[7] provides links between the newest SOC and knowledge, skills, abilities, educational levels, and work values. Some links between social class indicators and occupation have been made *ad hoc* for ISCO 68 and ISCO 88.[8–10]

Prevalence of confounders

Death certificate or census-based studies typically lack information on variables such as smoking and alcohol consumption, which can act as confounders in studies of tobacco and alcohol-related diseases. In absence of individual-level data, the prevalence of these variables in each occupation can be used for adjustment.[11] Data on cigarette smoking prevalence by occupation are generally available,[12] although most data are based on smoking patterns in the US. Country, culture, age, and sex differences as well as differences in trends over time need to be taken into account when using these data.

Occupation and industry as a risk factor

Occupations can be regarded as a proxy for exposure to a substance, a mixture of substances or other workplace characteristics. In occupational epidemiology the risk for a disease has often been analysed using occupation or industry information, leading to useful hypotheses with respect to more specific exposures. Besides calculating risks for each possible occupation/industry, epidemiologists may use clusters of high-risk occupations. Ahrens *et al.*[13] proposed lists, A and B, for high-risk occupations for lung cancer, based on ISCO 68 and ISIC Rev. 2. List A consists of occupations and industries known to be associated with lung cancer, and list B consists of occupations and industries that are suspected to be associated with the disease. A similar approach was used for a bladder cancer case-control study.[14]

Occupational exposure databases

If the occupational profile of a study population is available, the prevalence of exposure to occupational exposures can be inferred using databases, such as CAREX (International Information System on Occupational Exposure to Carcinogens).[15] CAREX contains estimates of the numbers of workers occupationally exposed to carcinogens by industry in 15 countries of the European Union 1990–1993. Statistics are available for 55 industrial categories of the ISIC Rev. 2 classification.[16]

A similar database is available for the US: NOES (National Occupational Exposure Survey),[17] containing, besides carcinogens, also other occupational exposures, and providing statistics for two-digit 1972 SIC codes.

The advantages and limitations of these databases have been discussed[15,17] and data collection methods need to be considered when using the databases.

Also available are several national databases in which quantitative workplace measurements have been assembled. For each measurement, information is included such as the sampling strategy applied, the time and place the sample was taken, the purpose and origin of the measurement, a description of the workplace and codes of the job title, and industry the sample was taken from. Although these data are not necessarily representative for a whole occupational group or industry, these databases can be used for exposure modelling and as a source for individual quantitative exposure estimates. Examples of such databases are MEGA from Germany,[18] NOEDB from UK,[19] ATABAS and BIOBAS from Denmark,[19] EXPO from Norway,[19] and the international database WAUNC.[20] The majority of these databases use the ISIC and ISCO codes for industries and occupations respectively.

Job-exposure matrices

Since the 1980s job-exposure matrices (JEM) have enhanced the value of coding by occupational classifications in epidemiological studies. Job-exposure matrices are cross-classifications of occupation and exposure. When linked with the occupation and industry codes of the study subjects, JEM place subjects from different industry–occupation combinations in the same exposure category.[21] Its automatic application avoids recall bias and differential misclassification of exposure. Job-exposure matrices can be categorized in general population JEM (GPJEM) and industry-specific JEM (ISJEM). The ISJEM[22] cross-classify exposures only for a limited group of occupations and tasks within one certain industry, and often include more-detailed definitions on the exposure axis. In a GPJEM,[23] all possible occupations that can occur in a population are represented and standard or national classifications are often adopted.

Kromhout and Vermeulen[24] presented an overview of all presently available GPJEM. In total 19 GPJEM were reviewed, 5 of which used ISCO and ISIC or a direct derivation of these classifications[25–29] and 5 used SIC/SOC combinations (any revision).[21,30–33] Nine GPJEM were based on national or *ad hoc* classifications.[34–42] Most GPJEM cross-classify occupation and industry with occupational (chemical) agents, while some are made for psychosocial risk factors such as occupational stress[28] or exposures related to asthma.[43]

The first consideration before using a GPJEM is the exposure of interest since each available GPJEM includes a limited list of occupational exposures. A second consideration would be the occupation and industry classification used in the GPJEM and the study. To make optimal use of the information in the JEM, the level of detail of the occupational classification in the study should be equal or higher compared with the level of detail in the JEM. Most of the time when applying an external GPJEM to a study population, a certain amount of recoding of occupational codes is needed. Recoding based on the original task description would be most ideal, but often the only feasible option is a direct conversion. Some conversion keys may be available in the literature or may be obtained through the Internet. The effect of recoding occupational codes to another classification for the application of a JEM was studied by Kromhout and Vermeulen.[24] Agreement was measured between codes obtained from a direct conversion key and those from new coding based on the full job descriptions. Recoding to a highly similar classification resulted in an agreement of 84% between direct codes and recodes of the same jobs. Recoding to a dissimilar coding scheme resulted in an agreement of only 49%. This low agreement needs to be considered when using direct conversion keys for multicentre studies. However, the agreement in exposures resulting from the application of a JEM was not effected by the low agreement of the job codes, indicating that although coded with different occupational codes, subjects were nonetheless assigned a similar exposure by the JEM.

In the absence of more sophisticated methods of occupational exposure assessment, JEM provide an easy and low-cost way to assess exposure based on occupation and industry title alone, but the occurrence of non-differential exposure misclassification should not be underestimated when using a JEM.[44,45]

Standard classification systems in use

National classifications for occupation and industry were first developed in the last decades of the 19th century. These first classifications started as listings of occupations without any hierarchical structure, and had a tendency to reflect social strata rather then tasks performed. With the increasing need for international and inter-discipline comparisons, the necessity for standardization emerged. International standards were discussed at the International Conferences of Labour Statisticians organized by the ILO in the 1920s, but concrete results only emerged during the 1940s.[46] To date, many countries base their national classifications (to some degree) on standardized classifications.

Hoffmann and Chamie[47] distinguish different types of classifications i.e. 'reference', 'derived', or 'related' classifications. *Reference classifications* are a product of international agreements approved by the United Nations Statistical Commission or another competent intergovernmental board. ISCO (International Standard Classification of Occupations, ILO)[48] and ISIC (International Standard Industrial Classification of All Economic Activities, UN) are reference classifications and recognized as such in the family of international economic classifications adopted at the Second Meeting of the Expert Group in International Classifications.[49] Reference classifications are used as guidelines for the preparation of national classifications and for international comparison. *Derived classifications* are based on reference classifications using the same structure, but in defining detailed categories they will go beyond the existing reference structure. For example, NACE Rev. 1 (Statistical Classification of Economic Activities in The European Community, Eurostat)[50] is derived from ISIC Rev. 3.[51] *Related classifications* might follow part of the reference classification's structure, but are associated with the reference classification at specific levels of the structure only.

Occupation classifications

Table 1 includes the main standardized classification systems used in epidemiological studies. The main standardized occupation classification used in Europe and other countries, besides the US, is the ISCO classification. It was first developed in 1958 by the International Labour Office (ILO). The most recent version is the 1988 revision[48,52] and in epidemiology the 1968 edition[53] has been used frequently. The aim of the new 1988 revision was to become the international standard for occupation classification. To make the classification applicable for other regions with specific requirements, ILO has provided advice for the development of three common regional classifications based on ISCO-88:[46] the European Union variant of ISCO-88 (ISCO-88(COM)),[47,54] a commonwealth variant (ISCO-88(CIS))[55] and an Asian variant (ISCO-88(OCWM)[56]).

Table 1 Overview of the main standardized classifications for occupations used in epidemiology

Running name	Full name	Organization	Year	Direct translations	Use
ISCO[48,52]	International Standard Classification of Occupations	International Labour Office (ILO) Geneva	1988 previous: 1958 1968	French (CITP-88) Spanish (CIUO-88) Catalan (CIOU-88) Russian (MCK3-88)	Used to collect ILO labour statistics. Many national occupation classifications are based on ISCO-88 or on the same basic structure
DOT	Dictionary of Occupational Titles	Washington (DC): US Department of Labor. US Government Printing Office	1991 previous: 1930 1965 1977		Mainly used in the US for vocational counselling. The Department of Labor has developed O*NET[7] as a replacement. DOT is becoming more and more outdated
SOC[57,58]	Standard Occupational Classification	US Department of Commerce, Office of Federal Statistical Policy and Standards. Washington DC: US Government Printing Office	1998 previous: 1966 1977 1980		Mainly used in the US, by US federal statistics agencies

In the US, occupational classifications have developed separately from those used in Europe. The main classifications that have been used in epidemiology are the Bureau of Census classification, the Dictionary of Occupational Titles (DOT), and the Standard Occupational Classification (SOC).[57] Currently most frequently used in epidemiological studies from the US is the SOC. It was first developed in 1966 and its latest revision dates from 1998.[58] A new revision of SOC is envisaged to start in 2005, in preparation for the 2010 Census of Population. The SOC was developed to reconcile the Census and the DOT systems to meet the need for one universal classification. The DOT[59] will be replaced by O*NET,[7,60] and the Bureau of the Census adopted the 1998 SOC[58] for the 2000 Census.

Industry classifications

The main international industry classifications used in epidemiology are presented in Table 2. The ISIC is the most widely used for statistics of economic activities and a new revision of ISIC is envisaged for 2007.[49] In the European Union, NACE Rev. 1[50] is widely used, which was designed as a more detailed version of ISIC Rev. 3[51] targeted towards the European circumstances.

In the US the most widely used is the SIC (Standard Industrial Classification),[61] that has undergone several revisions since the 1930s. In 1997 it was replaced by the NAICS (North American Industry Classification System).[62,63] The NAICS harmonized the separate schemes in use by the US, Canada, and Mexico, and is available in slightly different versions in English, French, and Spanish.

International comparability

The preceding paragraphs list the main standardized classifications for industry and occupation, but many countries have developed their own classification, often based on a standard classification.

Table 3 lists some of the national classifications currently in use in several European countries (not necessarily used in epidemiology), and summarizes how they compare to the standard classification ISCO-88 (COM). Ten out of 15 countries use a classification that compares well or reasonably well with ISCO-88. For other countries outside the European Union such as Japan, Mexico, and the US, the mapping of their national classifications tends to be complex resulting in poor comparability. The US SOC was not based on the ISCO-88 classification since the latter was not thought to be flexible enough for use in the US.[64] For Australia, Canada, and New Zealand the comparability with ISCO-88 is good.

Worldwide, 76% of the national occupation classifications have established a link with either ISCO-68 or ISCO-88.[6] However, only 28% of those national classifications that have a link with ISCO-88 have it on the most detailed level of ISCO-88.[6]

Table 2 Overview of the main standardized classifications for industries used in epidemiology

Running name	Full name	Organization	Year	Direct translations	Use
NACE[50,79]	Statistical Classification of Economic Activities in the European Community	Eurostat (EU) (subject of legislation at the European Union level. Imposes the use uniformly within all the Member States)	NACE Rev. 1: 1990 Previous: 1970	Full version (codes, headings, and explanatory texts): German, French, English Codes and headings: FR-EN-DE-NL-ES-PT-FI-SE-GR-AT-IT	NACE Rev. 1 has been designed as a more detailed version of ISIC Rev. 3 appropriate to European circumstances. The classification has been adopted by a large number of European States including most of the EFTA countries and a significant number of countries of Central and Eastern Europe
ISIC[51,65]	International Standard Industrial Classification of all Economic Activities	New York, United Nations	ISIC Rev. 3: 1989 ISIC: 1948, ISIC Rev. 1: 1958 ISIC Rev. 2: 1968/1971	Arabic, Chinese, English, French, Spanish, Russian	A reference classification recognized in the family of international classifications. Widely used for international statistics
SIC[61,80]	Standard Industrial Classification	Published by: US Office of Management and Budget. Washington (DC): US Governing Printing Office	1987 Previous: 1972 1977		Mainly used in the US. Has been replaced by NAICS
NAICS[62,63]	North American Industry Classification System	Prepared by the Office of Management and Budget's Economic Classification Policy Committee and printed by NTIS	1997	Spanish: Sistema de Clasificación Industrial de América del Norte (SCIAN) French: Système de classification des industries de l'Amérique du Nord (SCIAN)	New US industry classification replacing SIC. Designed as the index for statistical reporting of all economic activities of the US, Canada, and Mexico

Table 3 National classifications of occupations in Economic Union countries and their relationship to ISCO-88 (COM)

Country	Name of classification	Mapping to ISCO-88 (COM)	Comparability rating
Austria	ÖBS-72	Not available	Not applicable
Belgium	INS-91	Simple	Average
Denmark	DISCO-91	Simple	Average
Finland	TLN-95	Simple	Not known
France	PCS-82	Complex (+NACE, workplace size)	Good
Germany	Kld-92 Rev	Complex (+workplace size)	Good
Greece	STEP-92	Simple	Good
Ireland	SOC	Complex (+NACE, workplace size)	Average
Italy	CP-91	Simple	Poor
Luxembourg	ISCO-88	Simple	Average
The Netherlands	CBS 90/91	Direct	Average
Portugal	CNP-94	Simple	Good
Spain	CNO-94	Simple	Good
Sweden	SSYK-95	Simple	Not known
UK	SOC-2000	Simple	Average

Source: Elias, 1997.[6]

For the national industry classifications used in Europe the comparability tends to be better, because NACE Rev. 1 is the official format in which to send statistical data to Eurostat, and therefore used in most European countries. Table 4 lists the national industry classifications used in European countries and how they refer to standard classifications. Many countries in Africa and Asia have based their national classification on ISIC Rev. 3,[65] which compares well with NACE Rev. 1. The classification used in North America (NAICS) is comparable with NACE Rev. 1 and ISIC only on the two-digit level.[47,66]

The choice of a standard classification for the use in an epidemiological study

When choosing an occupational classification for use in an epidemiological study, the following characteristics should be considered.

Hierarchical structure

A classification with a hierarchical structure provides codes for detailed as well as aggregated groups. A hierarchical structure gives the coder the possibility to assign a more aggregated code to a subject, in case the choice between the more detailed descriptions is difficult to make. Furthermore, in case the detailed groups do not contain enough subjects for separate analysis, the detailed groups can be easily collapsed to more general groups.

Availability of a full version and translations

Old classifications are often non-hierarchical listings of occupation or industry title, without a detailed description for each title. To date, occupational classifications include a description of the tasks performed and industry classifications include a description of the products manufactured or services provided, to improve the interpretation of each title.

The availability of exact translations to other languages besides English will facilitate the use of the classification in multicentre studies. When using a translation, it needs to be checked to see if the version is an exact translation of the original standard classification, or a translation adjusted to the local situation. For example, the Brazilian classification for occupations Classificação Brasileira de Ocupaçőes (CBO)[67] is highly similar to ISCO 68, but many codes do not correspond.

The possibility of linkage with other data

Standardized classifications for occupation and industry can often be linked with other data or information (see paragraph on 'options for analysis'). Which classification to apply depends on the classification that is used, for example in the GPJEM or social class scale that will be used. Application of a JEM will generally require an occupation as well as industry classification while occupational class only considers occupation.

Also the level of detail needed will depend on the intended use, with social class indicators needing less detail than for example GPJEM.

However, in order not to lose any potential information it is recommended that the coding be as specific as the crude data (the job description) will allow,[68] so that the possibility of future comparison or linkage with other data will be maximized.

The use of a standard reference international classification such as ISCO and ISIC will generally give most flexibility in terms of comparing and linking possibilities, and will facilitate multicentre studies, while in the US the use of SOC and SIC can have certain advantages.

Methods for coding of occupation and industry and reliability

Different methods are available for obtaining and coding occupational and industry information from subjects in an epidemiological study.

Table 4 National classifications of industry in European countries and their relationship to standard classifications

Country	Name of classification	Based on
Albania	NVE	NACE Rev. 1
Austria	ÖNACE 1995	NACE Rev. 1
Belgium	NACEBEL	NACE Rev. 1
Bulgaria	NCEA	NACE Rev. 1
Croatia	NKD	NACE Rev. 1
Czech Republic	OKEC	NACE Rev. 1
Denmark	DB 93	NACE Rev. 1
Estonia	EMTAK	NACE Rev. 1
Finland	SIC 1995	NACE Rev. 1
France	NAF	NACE Rev. 1
Germany	WZ93	NACE Rev. 1
Greece	GRIC-91	NACE Rev. 1
Hungary	SZJ and TEAOR	NACE Rev. 1
Ireland	NACE Rev. 1	Exactly NACE Rev. 1
Italy	ATECO '91	NACE Rev. 1
Latvia	NACE	NACE Rev. 1
Lithuania	EVRK	NACE Rev. 1
The Netherlands	SBI	NACE Rev. 1
Norway	SN 94	NACE Rev. 1
Poland	PKD	NACE Rev. 1
Portugal	CAE-Rev. 2	NACE Rev. 1
Romania	CAEN	NACE Rev. 1
Russian Federation	OKDP	ISIC
Slovakia	OKEC	Exactly NACE Rev. 1
Slovania	SKD	NACE Rev. 1
Spain	CNAE-93	NACE Rev. 1
Sweden	SNI 92	NACE Rev. 1
UK	SIC 92	NACE Rev. 1

Source: United Nations, Department of Economic and Social Affairs.[65]

Self-classification

The study subject (or proxy) indicates the code or category of occupations to which the subject belongs. Although costs of this coding method are low, the use of it is limited, because a respondent will have difficulty in choosing between categories that are not clearly distinguishable without training. Detailed classification can only be done through one of the following methods.

Clerical coding

A coding expert finds the most applicable job title or economic activity title and code, based on what the study subject or proxy reports, by choosing between the different job descriptions given in a classification book. For most epidemiological studies, the coding for occupation and industry has been done through clerical coding based on questionnaire information.

Computer-assisted coding

A computer-based classification book generates some alternatives for job titles, based on keywords in the job description through word matching and algorithms. The coding expert chooses the most applicable job title. To date, more interviews are computer-assisted (CAPI: computer-assisted personal interview or CATI: computer-assisted telephone interview), enabling computer-assisted coding directly at time of interview.[69,70]

An evaluation of computer-assisted coding compared with clerical coding,[71] showed that computer-assisted coding of occupation by interviewers did not necessarily improve the quality of coding; however, it did reduce the coding time by 13–23%. For large batch operations, such as for census data, fully computerized coding is often applied. One study indicated that approximately two-thirds of the job title information provided by the respondents may be classified in a valid and reliable fashion by fully automated coding methods.[6] The presence of coding errors through computer-assisted and automated coding needs to be considered when using occupational codes from census data in an epidemiological study.

The reliability of the coding of occupation and industry, defined as the degree to which the results can be replicated, will depend on the following two parts of data collection. (1) The collection of the occupational information on which the coding will be based, for which the reliability depends on the design of the questionnaire, the interviewer, the recall of the subject etc. (2) The translation of this occupational information to a single code, for which the reliability depends on the coding expert's familiarity with the coding book, logical structure of the coding book, availability of clear coding rules, training of the coding experts, etc.

(1) Within epidemiology there is no standard question formulation used to collect information on occupation and industry, and the detail of questioning will depend on what the information will be used for. The minimum information required to choose the correct occupation code is the job title and a description of the main tasks or duties performed in the job description being coded. For industry, the name of the industry or business and the primary goods produced or services provided by the employer are required.

The importance of the availability of a full description of the job and not just a title is illustrated by an anecdote reported by Bushnell;[69] during a social survey in the UK a sudden increase in the number of able seamen was observed, which later appeared to be due to a trend for staff working in McDonalds fastfood restaurants to be titled 'crew members'.

Different methods for obtaining work histories are discussed in Stewart *et al.*,[72] and a short overview of work history reliability is given in McGuire *et al.*[73] When compared with occupational records, the self-reported occupational history has been shown to be reliable.[74,75] Only a few studies have looked at the repeatability of occupational history reporting by respondents,[24,76,77] generally showing good repeatability.

(2) The validity of the translation of the occupational information to a single occupational (and/or industry) code, has not often been studied in epidemiology. However, most national statistical institutes and other organizations which have responsibility for the coding of occupational data have undertaken coding/recoding studies at various times. An overview of some re-coding trials in the UK is given by Elias[6] and Bushnell[71] and represented in Table 5, with results of some coding reliability studies that were undertaken as part of an epidemiological study.

Table 5 Agreement rates (%) for occupation and industry classification coding reliability studies

Reference	4–5 digits	3 digits	2 digits	1 digit
Occupation coding in survey data				
Elias, 1997[6]	56–78	70–87	75–87	
Bushnell, 1997[71]	75–80		87–90	
Occupation coding in epidemiology				
Rona & Mosbech, 1989[77]		61–70		
Kennedy *et al*., 2000[43]	60			
Kromhout & Vermeulen, 2001[24]	44–89			
unpublished results from Porru *et al.*	44	59	79	95
unpublished results from the Eastern Europe lung cancer study	47	72	92	97
Industry coding in epidemiology				
Rona & Mosbech, 1989[77]		75–79		
Kromhout & Vermeulen, 2001[24]	71–98			
unpublished results from Porru *et al.*	74	74	84	91
unpublished results from the Eastern Europe lung cancer study	59	66	89	90
All occupation	44–89	56–80	61–92	75–97
All industry	59–98	66–74	75–89	90–91

As can be expected, agreement rates are better for higher levels of aggregation of occupational groups. Agreement rates in excess of 75% at the three-digit level appeared to be hard to obtain and in only few of the studies was agreement above 90% obtained for the highest level of aggregation (one-digit). Agreement depended on the type of jobs selected with the highest agreement rates obtained for non-problematic jobs.[24]

The study of Kromhout and Vermeulen[24] indicated that clear instructions on decision-making for the coders can improve the reliability of the coding of both occupation and industry codes (agreement for occupation improved from 69% to 89% and for industry from 92% to 98%). A study by Ahrens[78] indicated that coders show a learning curve and perform better after good familiarity with the classification is achieved and regular feedback on coding errors is given (agreement for occupation improved from 82% to 92%).

Discussion

Classifications of occupation and industry are frequently used tools in population-based epidemiological studies. Their use is, however, far from being as standardized as that of other classifications applied in this field, such as the International Statistical Classification of Diseases (ICD). The reason for this is probably the availability of many different national and international classifications, none of them specifically made for epidemiological studies. The lack of standardization is also due to the little importance that is granted to the methodological aspects of using these classifications in the field of epidemiology. Often the full reference of the classification that was used is not provided and only seldom is a study done to assess its reliability.

The reliability studies reported here indicated that job information can be translated reliably to an occupation and industry code given that coders are trained, have access to the full version of a classification book and guidelines for its use. Therefore, it is recommended that regular evaluations of coding work are implemented as part of the study, until the learning curve is completed. Any study using occupation and industry classifications should aim at an agreement rate between coders of at least 75% at the three-digit level.

Here we discussed the criteria that need to be considered when choosing an occupation or industry classification for use in population-based studies. The most important criteria are probably that the classification has a hierarchical structure and can be linked easily to other information systems such as labour statistics, socioeconomic indicators, JEM, and occupational exposure databases. The reference classifications of ISCO for occupation and ISIC/NACE for industry meet these needs, as do the American classifications SIC/NAICS and SOC. Exact translations into different languages are available for these classifications (except SOC), making them suitable for international use and comparisons.

Many national classifications are directly based on these reference classifications, and for many of these classifications links or crosswalks are available that can re-code in case one wants to compare or pool results or data from different studies. These links are, however, not always available for the most detailed level of the reference classifications, and re-coding of dissimilar classifications has shown to lead to considerable misclassification,[24] an evident shortcoming in epidemiology, which is a field that relies on valid comparisons of results. Before using a national classification it is therefore recommended that the level at which links have been established with reference classifications is verified. If national classifications do not give any additional advantage within the epidemiological study, the use of a reference classification is recommended.

Occupation and industry classifications have been, and will continue to be, an important tool in population-based epidemiological studies that study work-related risk factors, since this easily obtainable information can be put to different uses in population-based epidemiological studies. Their full exploitation in this field will, however, depend on a valid choice and a valid application of the classification.

KEY MESSAGES

- Documentation on methodological aspects of occupation and industry classifications is sparse within epidemiology, and hampers full exploitation of these tools.
- The applications of occupation and industry classifications in general population studies include the possibility of comparison with labour statistics, indirect adjustment for confounders, and inference of social class indicators and occupational exposures.
- Reliability of coding can be improved through training and the availability of clear instructions.
- Epidemiological studies would benefit from a better standardization in the use of these classifications.

References

1 LABORSTA. An International Labour Office database on labour statistics operated by the ILO Bureau of Statistics. http://laborsta.ilo.org. 2002.

2 US Department of Labor. Bureau of Labor Statistics. *Industry-Occupation Employment Matrix*. http://stats.bls.gov/emp/nioem/empioan.htm. 2002.

3 Kogevinas M, Pearce N, Susser M, Boffetta P. *Social Inequalities and Cancer*. IARC scientific publication No. 138, 1997.

4 Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. *Epidemiol Rev* 1988;**10:**87–121.

5 Krieger N, Williams DR, Moss NE. Measuring social class in US public health research. *Annu Rev Public Health* 1997;**18:**341–78.

6 Elias P. *Occupation Classification: Concepts, Methods, Reliability, Validity and Cross-National Comparability*. http://www.lisproject.org/publications/leswps/leswp5.pdf. Institute for Employment Research, University of Warwick, 1997.

7 US Department of Labor. *Employment and Training Administration. O*NET, the Occupational Information Network*. http://www.doleta.gov/programs/onet/. 2002.

8 Salavany AD, Alonso JM. Proposal of an indicator of social class based on the occupation (in Spanish). *Gac Sanit* 1989;**3:**320–26.

9 Ganzeboom HBG, De Graaf PM, Treiman DJ. A standard international socio-economic index of occupational status. *Soc Sci Res* 1992;**21:**1–56.

10 Ganzeboom HBG, Treiman DJ. Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Soc Sci Res* 1996;**25:**201–39.

11 Beaumont JJ, Singleton JA, Doebbert G, Riedmiller KR, Brackbill RM, Kizer KW. Adjustment for smoking, alcohol consumption, and socioeconomic status in the California Occupational Mortality Study. *Am J Ind Med* 1992;**21:**491–506.

12 Nelson DE, Emont SL, Brackbill RM, Cameron LL, Peddicord J, Fiore MC. Cigarette smoking prevalence by occupation in the United States. A comparison between 1978 to 1980 and 1987 to 1990. *J Occup Med* 1994;**36:**516–25.

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

14 't Mannetje A, Kogevinas M, Chang-Claude J *et al.* Occupation and bladder cancer in European women. *Cancer Causes Control* 1999;**10:**209–17.

15 Kauppinen T, Toikkanen J, Pedersen D *et al.* Occupational exposure to carcinogens in the European Union. *Occup Environ Med* 2000;**57:**10–18.

16 Finnish Institute of Occupational Health. *CAREX. International Information System on Occupational Exposure to Carcinogens*. http://www.occuphealth.fi/list/data/CAREX/. 2002.

17 Greife A, Young R, Carroll M *et al.* National Institute for Occupational Safety and Health general industry occupational exposure databases: Their structure, capabilities, and limitations. *Appl Occup Environ Hyg* 1995;**10:**264–69.

18 Stamm R. MEGA-database: one million data since 1972. *App Occup Environ Hyg* 2001;**16:**159–63.

19 European Foundation for the Improvement of Living and Working Conditions. *HASTE The European Health and Safety Database*. http://www.occuphealth.fi/e/eu/haste/. 2002.

20 Kromhout H, Symanski E, Rappaport SM. A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. *Ann Occup Hyg* 1993;**37:**253–70.

21 Hoar SK, Morrison AS, Cole P, Silverman DT. An occupation and exposure linkage system for the study of occupational carcinogenesis. *J Occup Med* 1980;**22**:722–26.

22 Goldberg M, Kromhout H, Guenel P *et al.* Job exposure matrices in industry. *Int J Epidemiol* 1993;**22(Suppl.2):**S10–S15.

23 Bouyee J, Héman D. Retrospective evaluation of occupational exposures in population-based case-control studies: general overview with special attention to job exposure matrices. *Int J Epidemiol* **22(Suppl.2):**557–64.

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

25 Macaluso M, Delzell E, Rose V, Perkins J, Oestenstad K. Inter-rater agreement in the assessment of solvent exposure at a car assembly plant. *Am Ind Hyg Assoc J* 1993;**54:**351–59.

26 Macaluso M, Vineis P, Continenza D, Ferrario F, Pisani P, Andisio R. Job exposure matrices: experience in Italy. Acheson ED. Medical Research Council conference report. Job Exposure Matrices, pp. 22–30. Southampton, MRC, 1983.

27 Ferrarion F, Continenza D, Pisani P, Magnani C, Merletti F, Berrino F. *Description of a Job Exposure Matrix for Sixteen Agents which are or may be Related to Respiratory Cancer*. Amsterdam: Elsevier, 1988.

28 Johnson JV, Stewart WF. Measuring work organization exposure over the life course with a job-exposure matrix. *Scand J Work Environ Health* 1993;**19:**21–28.

29 Demers PA, Kogevinas M, Boffetta P *et al.* Wood dust and sino-nasal cancer: pooled reanalysis of twelve case-control studies. *Am J Ind Med* 1995;**28:**151–66.

30 Olshan AF, Teschke K, Baird PA. Paternal occupation and congenital anomalies in offspring. *Am J Ind Med* 1991;**20:**447–75.

31 Sieber WK, Sundin DS, Frazier TM, Robinson CF. Development, use, and availability of a job exposure matrix based on national occupational hazard survey data. *Am J Ind Med* 1991;**20:**163.

32 Wortley P, Vaughan TL, Davis S, Morgan MS, Thomas DB. A case-control study of occupational risk factors for laryngeal cancer. *Br J Ind Med* 1992;**49:**837–44.

33 Gomez MR, Cocco P, Dosemeci M, Stewart PA. Occupational exposure to chlorinated aliphatic hydrocarbons: job exposure matrix. *Am J Ind Med* 1994;**26:**171–83.

34 Coggon D. Application of a job exposure matrix to occupational information obtained from death certificates. Acheson ED. Medical Research Council conference report. Job Exposure Matrices, pp. 83–86. Southampton, MRC, 1983.

35 Pannett B, Coggon D, Acheson ED. A job-exposure matrix for use in population based studies in England and Wales. *Br J Ind Med* 1985; **42:**777–83.

36 Magnani C, Pannett B, Winter PD, Coggon D. Application of a job-exposure matrix to national mortality statistics for lung cancer. *Br J Ind Med* 1988;**45:**70–72.

37 Steineck G, Plato N, Alfredsson L, Norell SE. Industry-related urothelial carcinogens: application of a job-exposure matrix to census data. *Am J Ind Med* 1989;**16:**209–24.

38 Imbernon E, Goldberg M, Bonenfant S *et al.* Occupational respiratory cancer and exposure to asbestos: a case-control study in a cohort of workers in the electricity and gas industry. *Am J Ind Med* 1995;**28:** 339–52.

39 Labreche FP, Cherry NM, McDonald JC. Psychiatric disorders and occupational exposure to solvents. *Br J Ind Med* 1992;**49:**820–25.

40 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.

41 Kauppinen T, Toikkanen J, Pukkala E. From cross-tabulations to multipurpose exposure information systems: a new job-exposure matrix. *Am J Ind Med* 1998;**33:**409–17.

42 Sunyer J, Kogevinas M, Kromhout H *et al.* Pulmonary ventilatory defects and occupational exposures in a population-based study in Spain. Spanish Group of the European Community Respiratory Health Survey. *Am J Respir Crit Care Med* 1998;**157:**512–17.

43 Kennedy SM, Le Moual N, Choudat D, Kauffmann F. Development of an asthma specific job exposure matrix and its application in the epidemiological study of genetics and environment in asthma (EGEA). *Occup Environ Med* 2000;**57:**635–41.

44 Kauppinen TP, Mutanen PO, Seitsamo JT. Magnitude of misclassification bias when using a job-exposure matrix. *Scan J Work Environ Health* 1992;**18:**105–12.

45 Bouyer J, Dardenne J, Hemon D. Performance of odds ratios obtained with a job-exposure matrix and individual exposure assessment with special reference to misclassification errors. *Scan J Work Environ Health* 1995;**21:**265–71.

46 Hoffmann E. *International Statistical Comparison of Occupational and Social Structures: Problems, Possibilities and the Role of ISCO-88*. http://www.ilo.org/public/english/bureau/stat/papers/index.htm. 1999.

47 Hoffmann E, Chamie M. *Standard Statistical Classifications: Basic Principles*. http://www.un.org/depts/unsd/statcom/principl.pdf, 1999.

48 International Labour Organization IS. *Resolution Concerning the Revision of the International Standard Classification Of Occupations*. http://www.ilo.org/public/english/bureau/stat/res/isco.htm. 2002.

49 United Nations, Department of Economic and Social Affairs, Statistics Division. Meeting of the expert group on International and Social Classifications. New York, 15–17 November 1999. ESA/STAT/AC.7514. http://www.un.org/Depts/unsd/class/ac75-4f.htm. 1999.

50 Eurostat (ES). *NACE Rev. 1: Statistical Classification of Economic Activities in the European Community*. 1996.

51 United Nations. *ISIC Rev. 3: International Standard Industrial Classification of all Economic Activities*. New York: UN (ST/ESA/STAT/SER.M/4/Rev.3), 1990.

52 International Labour Office (ILO). *International Standard Classification of Occupations: ISCO-88.* Geneva: ILO, 1990.

53 International Labour Office (ILO). *International Standard Classification of Occupations. Revised Edition 1968*. Geneva: ILO, 1981.

54 The University of Warwick. *ISCO 88 (COM)—The European Union Variant of ISCO 88.* http://www.warwick.ac.uk/ier/isco/frm-is88.html. 2002.

55 Statistical Committee of the Commonwealth of Independent States. *ISCO-88(CIS).* CIS-STAT, 1988.

56 ILO/UNDP Asian Regional Programme on International Labour. *Occupational Classification of Workers in Migration (ISCO-88(OCWM)).* ILO/UNDP, 1992.

57 US Department of Labor BoLS. *Standard Occupational Classification (SOC) System*. http://www.bls.gov/soc/home.htm. 2002.

58 US Office of Management and Budget. *Standard Occupational Classification 98 (SOC) Manual.* Washington DC: US Government Printing Office, 1999.

59 US Department of Labor. *Dictionary of Occupational Titles, Fourth Edn*. Washington DC: US Government Printing Office, 1977.

60 US Government Printing Office (GPO). *O*NET: The Occupational Information Network*. Washington DC: US Government Printing Office, 1998.

61 US Office of Management and Budget. *Standard Industrial Classification (SIC) Manual*. Washington DC: US Government Printing Office, 1977.

62 US Office of Management and Budget. *NAICS: North American Industry Classification System*. NTIS, 1997.

63 The NAICS Association. *NAICS, North American Industry Classification System*. http://NAICS.com/. 2002.

64 US Department of Labor. *Revising the Standard Occupational Classification System. Report 929*. Bureau of Labor Statistics, 1999.

65 United Nations, Department of Economic and Social Affairs, Statistics Division. *Methods and Classifications: National Classifications*. http://unstats.un.org/unsd/cr/ctryreg/default.asp. 2002.

66 United Nations, Department of Economic and Social Affairs, Statistics Division. *International Family of Economic and Social Classifications.* http://unstats.un.org/unsd/class/family/default.htm. 2002.

67 *Classificaçâo Brasileira de Ocupações (CBO)*. SINE Sistema Nacional de Emprego Brasilia, 1982.

68 Hoar SK. Epidemiology and occupational classification systems. In: Petro R, Schneiderman M (eds). *Banbury Report 9: Quantification of Occupational Cancer.* New York: Cold Spring Harbor Laboratory, 1981.

69 Bushnell D. *Computer Assisted Occupation Coding*. IBUC 95. Annual International Blaise Users Conference. 18–20 September 1995, Helsinki, Finland, pp. 23–34.

70 Frey R. *Developing a Blaise Instrument for the Spanish Bladder Cancer Study*. IBUC 2000. 6th International Blaise Users Conference. May 2000, Kinsale, Ireland.

71 Bushnell D. *An Evaluation of Computer-Assisted Occupation Coding: Results of a Field Trial*. IBUC 1997. Annual International Blaise Users Conference. May 1997, Paris, France, pp. 90–100.

72 Stewart WF, Stewart PA. Occupational case-control studies: I. Collecting information on work histories and work-related exposures. *Am J Ind Med* 1994;**26:**297–312.

73 McGuire V, Nelson LM, Koepsell TD, Checkoway H, Longstreth WT Jr. Assessment of occupational exposures in community-based case-control studies. *Annu Rev Public Health* 1998;**19:**35–53.

74 Baumgarten M, Siemiatycki J, Gibbs GW. Validity of work histories obtained by interview for epidemiologic purposes. *Am J Epidemiol* 1983;**118:**583–91.

75 Bourbonnais R, Meyer F, Theriault G. Validity of self reported work history. *Br J Ind Med* 1988;**45:**29–32.

76 Correa-Villasenor A, Batista C, Rothman N, Hakim R, Stewart W, Gray AS. Job histories obtained by interview from semi-conductor manufacturing workers: a reliability study. *Am J Epidemiol* 1991;**134:**737–38.

[77] Rona RJ, Mosbech J. Validity and repeatability of self-reported occupational and industrial history from patients in EEC countries. *Int J Epidemiol* 1989;**18:**674–79.

[78] Ahrens W. *Retrospective Assessment of Occupational Exposure in Case-Control Studies. Development, Evaluation and Comparison of Different Methods*. (Dissertation Bremen University, 1996). Fortschritte in der Epidemiology. Published in Germany by HE Wichmann HE, Jöckel K-H, Robra BP, 1999.

[79] The European Union on-line. *RAMON, Eurostat's Classification Server.* http://europa.eu.int/comm/eurostat/ramon/. 2002.

[80] US Department of Labor OSaHA. *Standard Industrial Classification Search.* http://www.osha.gov/oshstats/sicser.html. 2002.

International Journal of Epidemiology 2003;**32**:428–429
DOI: 10.1093/ije/dyg087

Commentary: Standardized coding of occupational data in epidemiological studies

Manolis Kogevinas

The evaluation of occupational exposures in epidemiological studies is complex because of the multiple potential exposures in the workplace, the varying determinants of exposure between people, the many jobs people hold in a lifetime, and the different reasons for taking or leaving a job. Mannetje and Kromhout[1] show that beyond these well-recognized difficulties there are several more basic issues that are not adequately dealt with in epidemiological studies, particularly the occupational classifications used, and the coding of this information. Their recommendations on the use of standard classifications (e.g. International Labour Organization [ILO]/International Standard Classification of Occupations [ISCO]), the improvement of coding, and the use of additional databanks, will help provide more reliable and comparable results in epidemiological studies regarding occupation. Pooling of studies evaluating occupational exposures typically requires extensive and very time-consuming re-coding of information on occupation and industry that are partially or even totally incompatible.[2] Furthermore, those of us involved in multicentre studies know that this is a particularly serious problem in those studies, since recording of information and coding varies between centres. Similar concerns in a wider context led to initiatives for the development of core questionnaires to be used in epidemiological studies.[3]

In earlier years, a basic classification by industry or major occupations sufficed to identify occupational risks.[4,5] Despite the serious limitations of exposure assessment methods that use only information on occupation and industry, these methods have helped identify specific risks and should continue being used. Surprisingly, in several situations they may even be the best proxy we have for the evaluation of combined and complex exposures. More powerful methods have been developed both for industry-based and population-based studies, including methods for the collection of more detailed information initially, and also elaborate ways for the evaluation of this information. Collection of occupational information can be done, for example, through computer-assisted interviews, repeated interviews with selected subjects, and use of modular questionnaires.[6,7] Evaluation of the data includes assessment by experts,[7,8] and the use of job exposure matrices based on extensive population-based measurements.[9] The identification of exposures still remains a complex issue, but in studies focusing on occupation the available methods have dramatically improved exposure assessment.

One of the main problems many of us encounter refers to the methods applied in studies that are not principally focusing on occupational exposures and in which the interview time dedicated to the evaluation of these exposures is limited. In these studies a balance has to be found between the need to restrict the questionnaire time or other resources for the evaluation of occupational exposures, and the need to get detailed and valid answers. This balance can be achieved if adequate preparatory work is done and priorities are set regarding the evaluation of specific exposures.

Whatever the aims of each study, one main message to be kept from the paper by Mannetje and Kromhout[1] is that information on occupation and industry is, in most studies, one of the basic variables to be collected, similar to smoking, education, or race. In addition, analyses by occupation and industry remain a main method for the identification of occupational risks. An effort should therefore be made to classify them correctly and in a generalizable way. The issues raised by Mannetje and Kromhout are important and have a fairly easy remedy since they depend mostly on researchers being better informed. Such improvements in study methodology do not cost much and, more importantly, they do not complicate the study design and the time requested from study participants. The guidelines mentioned by Mannetje

Respiratory and Environmental Health Research Unit, Municipal Institute of Medical Research (IMIM), Barcelona, Spain, and Division of Cancer Epidemiology and Genetics, National Cancer Institute, MD, USA.

Correspondence: Prof. Manolis Kogevinas, Respiratory and Environmental Health Research Unit, Municipal Institute of Medical Research (IMIM), 80 Dr. Aiguader Rd, Barcelona 08003, Spain. E-mail: kogevinas@imim.es