Original article

Scand J Work Environ Health. 2020;46(3):259–267. doi:10.5271/sjweh.3857

Influence of errors in job codes on job exposure matrix-based exposure assessment in the register-based occupational cohort DOC*X

by Sesilje B Petersen, PhD,[1] Esben M Flachs, PhD,[1] Susanne W Svendsen, PhD,[2] Jacob L Marott, MSc,[3] Esben Budtz-Jørgensen, PhD,[4] Johnni Hansen, PhD,[5] Zara Ann Stokholm, PhD,[6] Vivi Schlünssen, PhD,[6, 7] Johan Hviid Andersen, PhD,[2] Jens Peter Bonde, MD [1]

Petersen SB, Flachs EM, Svendsen SW, Marott JL, Budtz-Jørgensen E, Hansen J, Stokholm ZA, Schlünssen V, Andersen JH, Bonde JP. 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 Health.* 2020;46(3):259–267. doi:10.5271/sjweh.3857

Objective Job-exposure matrices (JEM) may be efficient for exposure assessment in occupational epidemiological studies, but they rely on valid job information. We evaluated the agreement between JEM-based exposure estimates according to self-reported job titles converted to DISCO-88 codes and according to register-based DISCO-88 codes in the Danish Occupational Cohort with eXposure data (DOC*X). Furthermore, we evaluated the agreement between these two sets of DISCO-88 codes.

Methods We used JEM regarding wood dust, lifting, standing/walking, arm elevation >90°, and noise from DOC*X. Participants from previous questionnaire studies were assigned JEM-based exposure estimates using (i) self-reported job titles converted to DISCO-88 codes and (ii) DISCO-88 codes registered in DOC*X, in four time periods (1976–78: N=7707; 1981–83: N=2193; 1991–94: N=2664; 2004: N=11 782). Agreement between the exposure estimates and between the DISCO-88 codes (four-digit levels, 1–4) was evaluated by kappa (κ) statistics. Sensitivities were calculated using the self-reported observation as the gold standard.

Results We found substantial agreement (κ>0.60) between exposure estimates for all types of job-exposures and all time periods except for one κ. Low sensitivity (30–65%) was found for the period 1981–83, but for the other time periods the sensitivities varied between 60–91%. For individual 4-digit DISCO-88 codes, the sensitivities varied substantially and overall the sensitivities increased by lower digit level of DISCO-88.

Conclusion The validity of the DISCO-88 codes in DOC*X was generally high. Substantial agreement was found for the JEM-based exposure estimates and the DISCO-88 codes per se, although the DISCO-88 code-specific agreement varied across digit levels and time periods.

Key terms arm elevation; epidemiology; ISCO-88; JEM; job code error; lifting; metal dust; noise; occupation; shoulder; standing; validity; walking; wood dust.

Since the late 1970s, job-exposure matrices (JEM) have been increasingly used to obtain exposure estimates in occupational epidemiological studies. A JEM is a cross-tabulation of job titles or occupational codes and occupational exposures, preferably for a specific time window (1–4). JEM can be used in large epidemiological studies where methods based on individual interview data, observation, or technical measurements would be very costly. Other important advantages are that JEM can be used to estimate both current and past exposures and minimize the risk of information bias compared to individual-based self-report methods (2, 5, 6).

1 Department of Occupational and Environmental Medicine, Bispebjerg and Frederiksberg Hospital, University of Copenhagen, Copenhagen, Denmark.
2 Danish Ramazzini Centre, Department of Occupational Medicine, Regional Hospital West Jutland – University Research Clinic, Herning, Denmark.
3 Copenhagen City Heart Study, Frederiksberg University Hospital, Copenhagen, Denmark.
4 Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
5 Danish Cancer Society Research Center, Copenhagen, Denmark.
6 Danish Ramazzini Centre, Department of Occupational Medicine, Aarhus University Hospital, Aarhus, Denmark.
7 National Research Centre for Working Environment, Copenhagen, Denmark.
8 Department of Public Health, Occupation, Environment and Health, Danish Ramazzini Centre, Aarhus University, Denmark.

Correspondence to: Sesilje Bondo Petersen, Department of Occupational and Environmental Medicine, Bispebjerg and Frederiksberg Hospital, University of Copenhagen, Copenhagen, Denmark. [E-mail: sesilje.elise.bondo.petersen@regionh.dk]

The validity of occupational exposure estimates assigned to individuals by means of JEM depends on the quality of information about exposures in specific jobs in different time periods, as well as on correct job titles or occupational codes (7). The latter aspect of JEM validity is particularly important when occupational codes are retrieved from national registers, without occupational research as the primary objective. While the validity of exposures assigned by JEM has been examined in a number of publications (8–13), the validity of the job titles and occupational codes per se has seldom been examined (7, 14). Incorrect occupational codes in registers may be the result of erroneous reporting from the primary sources (eg, tax agents, companies) and – if classification systems have changed over time – errors in translation from one classification system to another. Therefore, the validity of registered occupational codes may vary between industries and occupations and across time periods.

The Danish Occupational Cohort with eXposure data (DOC*X) is a nationwide cohort for occupational research containing occupational histories in terms of year-by-year codes according to the Danish version of the International Standard Classification of Occupations (DISCO) on an individual level from 1970 through 2015 with ongoing updates. DOC*X is an open research resource that provides opportunities to perform register-based epidemiological studies of occupational exposures by use of JEM (15). The validity of the DISCO codes in the nationwide registers, which form the foundation of DOC*X, has not been investigated.

The overall aim of this study was to evaluate the validity of DISCO codes in DOC*X. Specific aims were to evaluate (i) the agreement between JEM-based exposure estimates according to self-reported job titles converted to DISCO codes and according to register-based DISCO codes in DOC*X; and (ii) the agreement between these two sets of DISCO codes per se.

Methods

Danish Occupational Cohort with eXposure data (DOC*X)

DOC*X is a nationwide database including 6.4 million residents in Denmark from the age of 16, who have been gainfully employed at a private or public workplace in Denmark from 1970 through 2015 (15–17). The database has been compiled and is updated at a secured platform at Statistics Denmark. The backbone of the database is the information on occupation and industry, which includes calendar specific DISCO-88 codes for each individual based on the 1970 Census (16) and the Employment Classification Module (1976–2015) (17). The Employment Classification Module has used three classifications: (i) a scheme developed by Statistics Denmark based on ISCO-68 (1976–1990), (ii) DISCO-88 (1991–2009), and (ii) DISCO-08 (2010 onwards) (15). In DOC*X, the different coding versions have been harmonized to DISCO-88 codes in a code-by-code manner as described previously (15). The codes vary in detail from 1- to 4-digit levels, of which the last-mentioned is the most detailed. The annual DISCO-88 code for each individual is defined by the job with the highest income during each calendar year. We extracted annual DISCO-88 codes by use of the personal identifier (18).

Population used for validation

From 1976–1994, we used occupational data from the Copenhagen City Heart Study (CCHS). In total, 19 698 men and women from the center of Copenhagen were randomly drawn from the Copenhagen Population Register. The sample was age-stratified within 5-year age groups from 35–70 years of age. All participants completed a self-administrated questionnaire in 1976–1978, including a freeform question about current job title (N=14 223). Follow-up studies with information on job title were completed in 1981–83 (≥500 20–25-year-olds) and in 1991–94 (≥3000 20–49 year-olds) (19, 20). The proportions that responded were 73.6% at baseline and 70.2% and 61.2% at follow-up. In the beginning of 2016, the job title text strings from the stored questionnaires were digitalized and assigned DISCO-88 codes by three librarians, who worked independently. The codes were cross-checked and a supervising occupational health specialist resolved discrepancies.

For 2004, we used data from the ASUSI cohort of 14 266 men and women, who completed a questionnaire in a population-based study of working environment and sickness absence (ASUSI is a Danish acronym for working environment, sickness absence, premature exit from the labor market, social inheritance, and intervention) (21). Two trained sociologists digitalized the job title text strings from the questionnaires assigned DISCO-88 codes. Only persons who had been in employment for ≥80% of the time during the previous year or had been employed for 6 out of the 12 weeks preceding 1 July 2004 were included.

Assessment of occupational exposure intensities

We assessed five types of exposure using four JEM:

Wood dust estimates were assessed using a wood dust JEM based on expert ratings and 12 704 measurements collected in 1978–2007 in wood related industries in six European countries (22, 23). We dichotomized the exposures as non-exposed and exposed because wood dust exposure was rare in the study population.

Lifting and standing/walking estimates were assessed using the Lower Body JEM (24) . Five Danish occupational health physicians with a minimum of 10 years of experience rated the exposures. We categorized the lifting exposures as described previously (25–28) (0=non-exposed, 1=medium exposed (>0–<1000 kg/day), and 2=highly exposed (≥1000 kg/day)) and divided the exposure estimates for standing/walking into three groups [(0=non-exposed (0 hours/day), 1=medium exposed (>0–5.9 hours/day), and 2=highly exposed (≥6.0 hours/day)] according to previously used categories (27, 28).

Work with the arms elevated >90° estimates were assessed using the Shoulder JEM, which is based on expert ratings by five Danish occupational health physicians with a minimum of 10 years of experience (29–32). The expert rated estimates of time spent working with the arms elevated >90° (hours/day) have been validated against technical measurements (13). We divided the exposure estimates according to previously used cut-off value for high exposure (0=non-exposed, 1=medium exposed (>0-0.4 hours/day), and 2=highly exposed (≥0.5 hours/day) (32, 33).

Noise was assessed using the Noise JEM (35, 36), which is based on personal dosimeter measures of occupational noise exposure in the periods 2001–03 and 2009–10 among 1140 workers (1343 measurements) within the ten industries with the highest reporting of noise induced hearing loss according to the Danish Working Environment Authority. The measurements represented 100 occupational titles according to the DISCO-88 system. Four experts rated the noise intensity levels for the remaining jobs using 35 benchmark groups. Their ratings were used to construct an expert score dependent on sex, age, and calendar time (34, 35). We used the categorical variable for noise exposure (0=<80 dB, 1=80-84 dB, 2=≥85dB), based on ISO-1999 thresholds (35, 36).

We assigned exposure estimates to individuals in the CCHS/ASUSI cohorts with DISCO-88 codes for which a JEM exposure estimate was available. The estimates were assigned by connecting the JEM with their calendar-year specific DISCO-88 codes based on self-report and their DISCO-88 codes in DOC*X for the specific calendar year.

Statistical methods

From both cohorts (CCHS and ASUSI) and each time period, we excluded persons, who stated that they were unemployed or had retired. For each exposure and time period, the final population included only individuals with both sets of DISCO-88 codes and only DISCO-88 codes with ≥10 self-reported observations (37). Furthermore, we only included observations where JEM-based exposure estimates were available for both sets of codes.

We computed kappa coefficients (κ) with 95% confidence intervals (CI) for exposures with two exposure categories (wood dust) and weighted κ with 95% CI for exposures with three exposure categories (all other exposures). Additionally, we in 3×3 tables computed sensitivity (the percentage of true exposure categorizations for the highest exposed individuals) and specificity (the percentage of true exposure categorizations for the non-exposed individuals) based on self-report as the gold standard. This means that the medium exposed groups not were included in the interpretation of sensitivity and specificity. We also assessed the sensitivity and agreement (weighted κ) between the DISCO-88 codes per se (specificity was not assessed because it would always be very high due to the low frequency of persons in any DISCO-88 group compared to the total number of persons in the study). Sensitivity was calculated as the percentage of true registrations within each DISCO-88 code digit level (1–4) taking the DISCO-88 codes based on self-report as the gold standard. In addition to the agreement at 1-, 2-, 3-, and 4-digit levels, we computed weighted κ coefficients by time period (1976–78; 1981–1983; 1991–1994; 2004) at DISCO-88 1-digit level (DISCO-88 major groups). We interpreted the κ coefficients as: <0=poor, 0.00–0.20=slight, 0.21–0.40=fair, 0.41–0.60=moderate, 0.61–0.80=substantial, and 0.81–1.00=almost perfect agreement (38). SAS software, version 9.4, (SAS Institute Inc, Cary, NC, USA) was used.

Results

Table 1 presents the number of DISCO-88 codes according to time period, including all digit levels of DISCO-88 (based on self-reported job titles), that met the inclusion criteria of minimum ten observations in our final study dataset. These codes represented 29–56% of the total number of codes, including all digit levels of the DISCO-88 system, with the lowest percentage in 1991–94 and the highest in 2004. The number of individuals in each time period is also shown; their distribution across DISCO-88 groups is presented in supplementary table S1, www.sjweh.fi/show_abstract.php?abstract_id=3857.

As seen in table 2, our data showed substantial agreement between JEM-based exposure estimates according to the two sets of DISCO-88 codes based on self-reported job titles and registrations in DOC*X, except for noise in 1981–83. Across time, both the sensitivities and κ estimates were lowest for the time period 1981–83. Overall, the specificities were high showing substantial agreement for the non-exposed individuals.

Table 1. Number of individuals with two sets of DISCO-88 (Danish version of the International Standard Classification of Occupations from 1988) codes including all (1–4) digit levels of DISCO that met the inclusion criteria of ≥10 self-reported observations.

Time period	N (≥10) [a] DISCO-88 codes	% (≥10) [b] DISCO-88 codes	Final population number of individuals
1976–78	215	44	7707
1981–83	180	37	7193
1991–94	142	29	2664
2004	271	56	11 782

[a] Number of DISCO-88 groups available for validation in the final population out of 486 groups in the DISCO-88 classification system including all (1–4) digit levels.

[b] Percent of DISCO-88 groups, including all (1–4) digit levels of DISCO, available for validation in the final study sample.

Table 3 shows that the agreements between the two sets of DISCO-88 codes were substantial across 1-, 2-, 3-, and 4-digit levels. The highest κ estimates were seen for the 4-digit DISCO-88 group level with estimates between 0.73–0.81. The sensitivities varied between 51.5–73.2% and were highest for the 1-digit DISCO-88 level. As seen in table 4, the DISCO-88 code specific agreement at 1-digit level varied from fair to almost perfect across time periods (κ=0.34–0.91). Group 0 (armed forces) had almost perfect agreement, whereas group 1 with legislators, senior officials, and managers showed the lowest agreement; no time trends were evident. The sensitivities generally showed the same pattern as the κ-values.

Sensitivities for individual DISCO-88 codes, according to time period, are presented in supplementary table S1. The highest sensitivities across all time periods were found for dentists (2222; 96.2%); nursing associate professionals (3231; 95.0%); police officers (5162; 92.2%); medical doctors (2221; 91.5%); jewelry and precious-metal workers (7313; 91.3%); bakers, pastry-cooks and confectionery-makers (7412; 89.3%); and primary education teaching professionals (2331; 89.8%). Prison guards (5163) and travel attendants (5111) and travel stewards had 100% sensitivity in 2004, but not enough observations for the other time periods. In general, low sensitivities were found across all time periods for business services agents and trade brokers not elsewhere classified (3429; 1.7%); production clerks (4132; 6.2%); other teaching associate professionals (3340; 6.5%); advertising and public relations managers (1234; 7.4%); finance and sales associate professionals not elsewhere classified (3419; 10.1%); safety, health and quality inspectors (11.7%; 3152); receptionists and information clerks (4222; 12.3%); and buyers (3416; 13.5%).

Discussion

Job titles and occupational codes constitute a crucial basis for the use of JEM, but errors in job titles and assignment of occupational codes have received minimal scientific attention. The present study benefitted from exposure data from JEM concerning five airborne, mechanical, and physical exposures. Self-reported job titles for the CCHS/ASUSI cohorts were translated into DISCO-88 codes, which were connected with the JEM to provide exposure estimates, which were then compared to JEM-based exposure estimates according to DISCO-88 codes registered in DOC*X. High sensitivities and substantial agreement was found for the JEM-based exposure estimates and for the DISCO-88 codes per se, although the DISCO-88 code-specific agreement varied across digit levels and across time periods.

The number of individuals in the study population from 1991–94 was low since only about one third of the individuals with a self-reported job title had a DISCO-88 code in DOC*X. An explanation may be the higher mean age in the population by calendar time as the main part of the population was included in 1976 with an age of up to 70 years at that time. For example, if they retired from the workforce before 1991, they have no DISCO code registered in DOC*X database for the time-period 1991–94. The classification system used by Statistics Denmark changed in 1981 and 1993, which may be an explanation for lower agreement observed in the period 1981–83, and again in 1991–94. In 1981–83, the classification system was less detailed than the DISCO-88 system. This means that it was very difficult to translate specific job groups from that time-period to DISCO-88 codes. Therefore, discrepancies between DISCO-88 codes may be because of translation difficulties rather than exact differences between jobs. Because of the less detailed job groups in 1981–83, the solution was to translate job titles to less detailed DISCO-88 group levels. The system for code assignment also changed in 1991, when the DISCO-88 classification system was introduced by Statistics Denmark. The DISCO-88 was based on the ISCO-88. Before 1991, the occupational codes were assigned by trained coders at Statistics Denmark based on self-reported information and union membership, but from 1991 the system was automatized and based on tax records and other personal register information. This shift in code assignment led to a temporary reduction of data reporting, which probably also contributed to the low number of individuals in the final study population for 1991–94.

The variation across DISCO-88 codes probably reflected variations in the accuracy by which DISCO codes are reported to the central authorities. Reporting to Statistics Denmark from large public and private com-

Table 2. Sensitivity, specificity, and agreement between occupational exposures assigned by job-exposure matrices (JEM) according to self-reported job titles converted to DISCO-88 codes and according to DISCO-88 codes registered in the Danish Occupational Cohort with eXposure data (DOC*X). For each exposure and time period, the final population included only individuals with both sets of codes, only DISCO-88 codes with ≥10 self-reported observations were included, and only DISCO-88 codes for which there is a JEM-exposure estimate. [CI=confidence interval; DISCO-88=Danish version of the International Standard Classification of Occupations from 1988; κ=kappa coefficient]

Exposure time period	Self-reported N non/medium/high	Registered N non/medium/high	Sensitivity [a]	Specificity [b]	Agreement weighted κ, (95% CI)
Wood dust [c]					
1976–1978 [d]	6448/–/119	6446/–/121	90.9	99.9	0.91 (0.88–0.95) [f]
1981–1983 [d]	4295/–/37	4294/–/38	63.2	99.7	0.64 (0.51–0.76) [f]
1991–1994 [d]	1712/–/14	1710/–/16	85.7	99.8	0.80 (0.64–0.96) [f]
2004 [e]	9465/–/230	9479/–/216	76.1	99.7	0.78 (0.74–0.82)[f]
Lifting					
1976–1978 [d]	2854/2755/944	2655/2695/1203	60.3	89.6	0.71 (0.70–0.72)
1981–1983 [d]	2196/1638/465	2108/1785/406	47.3	88.9	0.64 (0.63–0.66)
1991–1994 [d]	904/585/198	817/665/205	76.6	94.6	0.78 (0.75–0.81)
2004 [e]	4358/2777/1783	4383/2826/1371	80.2	84.8	0.72 (0.70–0.73)
Standing/walking					
1976–1978 [d]	2776/2032/1745	2619/2698/1236	78.8	89.5	0.68 (0.67–0.70)
1981–1983 [d]	2160/1242/897	2070/1140/1089	61.5	89.4	0.68 (0.66–0.70)
1991–1994 [d]	882/507/298	811/575/301	76.4	94.6	0.78 (0.75–0.80)
2004 [e]	4347/3224/1347	4379/3122/1417	68.2	84.8	0.69 (0.67–0.70)
Arm elevation >90°					
1976–1978 [d]	2790/2646/1083	2645/2988/886	86.0	91.4	0.78 (0.77–0.80)
1981–1983 [d]	2235/1503/594	2090/1344/898	57.9	89.8	0.70 (0.68–0.72)
1991–1994 [d]	941/624/179	891/632/179	71.9	92.1	0.78 (0.75–0.80)
2004 [e]	4875/3379/1384	5146/3016/1476	73.8	82.2	0.69 (0.68–0.70)
Noise					
1976–1978 [d]	4713/1516/387	4568/1663/385	75.1	94.0	0.75 (0.73–0.76)
1981–1983 [d]	3482/777/73	3251/954/127	29.9	93.8	0.56 (0.53–0.58)
1991–1994 [d]	1345/386/15	1319/400/27	73.3	94.3	0.78 (0.75–0.81)
2004 [e]	6504/2634/587	6703/2473/549	60.8	93.2	0.72 (0.70–0.73)

[a] The percentage of true registrations for the highest exposed individuals.
[b] The percentage of true registrations for the non-exposed individuals.
[c] Dichotomized (non-exposed/exposed)
[d] Observations from the Copenhagen City Heart Study.
[e] Observations from the ASUSI study. (ASUSI is a Danish acronym for working environment, sickness absence, premature exit from the labor market, social inheritance, and intervention)
[f] For wood dust the κ and 95% CI are not weighted.

panies is undertaken by trained staff according to written guidelines, while small private companies with fewer resources may provide less accurate DISCO codes. It is only mandatory for Danish companies with ≥10 employees to report information on occupation, and therefore significant differences in accuracy may be expected.

The misclassification of JEM-based individual exposures assigned by using DISCO-88 codes in DOC*X seems less than might be expected based on comparison of the sensitivities for the DISCO-88 codes per se; overall, the sensitivities were higher when comparing JEM-based exposure estimates than when comparing the two sets of DISCO-88 codes (especially at the 3- and 4-digit levels). This is because DISCO-88 codes belonging to similar job groups in the JEM are assigned similar job-exposures (7, 14). For example, the noise JEM will assign the same low level of noise exposure to all types of office workers regardless of the specific DISCO-88 code. Lack of agreement between two sets of DISCO-88 codes will therefore not necessarily affect the agreement between JEM-based exposure estimates.

The variation in agreement between the two sets of individual DISCO-88 codes seems to depend on characteristics of the jobs covered by the code. In general, the codes with lowest sensitivities are broadly defined and not specified, eg, business services agents and trade brokers not elsewhere classified, other teaching associate professionals, and finance and sales associate professionals not elsewhere classified. The two last-mentioned groups will probably be classified as other kinds of office workers, which will reduce the effect of the misclassification on the assigned JEM-based exposure estimates (see above). Another possibility is to exclude DISCO codes with low sensitivities in epidemiological studies (at least in sensitivity analyses) as they may increase the risk of misclassification of exposures. Thus, the actual validity of the DISCO-codes per se may be significantly higher in cleaned data prepared for analysis.

Strengths and limitations

One strength of our study is that we have data from four different time periods during a 24-year long period where Statistics Denmark used different classification systems of occupations in their registers. Furthermore, we have access to self-reported job titles. It may

Table 3. Sensitivity and agreement between self-reported job titles converted to DISCO-88 codes and DISCO-88 codes registered in the Danish Occupational Cohort with eXposure data (DOC*X) at 1-4-digit levels. [CI=confidence interval; DISCO-88=Danish version of the International Standard Classification of Occupations from 1988; κ= kappa coefficient]

	Self-reported [a]	Registered [b]	Final population [c]	Sensitivity [d]	Agreement
	N	% (N)	% (N)		κ, (95% CI)
1976–1978 [e]					
4-digit	10 443	73.8 (7708)	55.8 (5824)	66.3	0.77 (0.76–0.79)
3-digit	10 933	74.3 (8124)	65.7 (7182)	61.4	0.71 (0.70–0.72)
2-digit	11 335	71.7 (8128)	66.1 (7491)	67.1	0.71 (0.70–0.73)
1-digit	11 688	72.1 (8430)	65.9 (7707)	70.6	0.73 (0.72–0.74)
1981–1983 [e]					
4-digit	9319	55.8 (5204)	42.6 (3973)	59.4	0.74 (0.72–0.75)
3-digit	9744	69.8 (6804)	65.2 (6352)	54.4	0.69 (0.68–0.71)
2-digit	10 041	69.8 (7012)	68.3 (6856)	60.8	0.69 (0.67–0.70)
1-digit	10 311	69.8 (7199)	69.8 (7193)	65.5	0.69 (0.67–0.70)
1991–1994 [e]					
4-digit	8186	28.4 (2322)	17.6 (1443)	71.4	0.81 (0.79–0.83)
3-digit	8552	28.6 (2447)	23.9 (2042)	65.5	0.72 (0.70–0.75)
2-digit	8753	28.8 (2522)	28.0 (2451)	69.4	0.72 (0.70–0.74)
1-digit	8992	29.7 (2668)	29.6 (2664)	73.2	0.73 (0.71–0.75)
2004 [f]					
4-digit	13 858	77.9 (10 794)	65.4 (9064)	51.5	0.73 (0.72–0.74)
3-digit	13 892	78.4 (10 891)	72.9 (10 134)	56.6	0.72 (0.71–0.73)
2-digit	13 892	82.6 (11 469)	75.9 (10 540)	64.1	0.72 (0.71–0.73)
1-digit	14 266	84.5 (12 048)	82.6 (11 782)	65.8	0.71 (0.70–0.72)

[a] Number of individuals with a DISCO-88 code based on self-reported job-titles within each DISCO-88 code digit level.

[b] Number of individuals also registered in DOC*X within each DISCO-88 code digit level.

[c] For each exposure and time period, the final study sample includes observations with both sets of codes, and only DISCO-codes with at least 10 self-reported observations overall in the sample.

[d] The percentage of true registrations within each DISCO-88 code digit level based on self-reported job-title as the gold standard.

[e] Agreement between registered DISCO-88 codes in DOC*X and self-reported job titles converted to DISCO-88 codes based on the Copenhagen City Heart Study.

[f] Agreement between registered DISCO-88 codes in DOC*X and self-reported job titles converted to DISCO-88 codes based on the ASUSI cohort.

Table 4. Sensitivity and agreement between DISCO-88 codes (major group level) registered in the Danish Occupational Cohort with eXposure data (DOC*X) and DISCO-88 codes assigned from self-reported job titles according to time period. [CI=confidence interval; DISCO-88=Danish version of the International Standard Classification of Occupations from 1988; κ= kappa coefficient]

DISCO Group [a]	1976–1978			1981–1983			1991–1994			2004		
	N [b]	Sensitivity [c] (%)	Agreement [d] κ, (95% CI)	N [b]	Sensitivity [c] (%)	Agreement [d] κ, (95% CI)	N [b]	Sensitivity [c] (%)	Agreement [d] κ, (95% CI)	N [b]	Sensitivity [c] (%)	Agreement [d] κ, (95% CI)
0	55	96.4	0.91 (0.86–0.97)	48	97.9	0.87 (0.80–0.94)	10	90.0	0.86 (0.70–1.00)	83	90.4	0.71 (0.64–0.78)
1	303	57.4	0.43 (0.39–0.48)	340	40.0	0.38 (0.33–0.43)	155	56.8	0.46 (0.40–0.53)	994	33.3	0.41 (0.37–0.44)
2	933	82.1	0.82 (0.80–0.84)	995	78.7	0.73 (0.71–0.75)	665	78.0	0.78 (0.75–0.80)	2258	71.4	0.69 (0.68–0.71)
3	1038	67.8	0.56 (0.53–0.59)	1054	54.8	0.53 (0.50–0.56)	557	68.4	0.62 (0.58–0.66)	2487	67.0	0.52 (0.50–0.53)
4	1549	81.7	0.70 (0.68–0.72)	1481	75.3	0.72 (0.70–0.74)	411	77.1	0.72 (0.69–0.76)	1130	68.0	0.52 (0.50–0.55)
5	1125	50.7	0.53 (0.50–0.56)	985	56.4	0.56 (0.53–0.59)	274	72.3	0.71 (0.67–0.76)	1630	80.7	0.76 (0.74–0.77)
6	20	75.0	0.77 (0.62–0.92)	20	55.0	0.40 (0.23–0.56)	11	54.5	0.54 (0.29–0.79)	61	57.4	0.52 (0.42–0.63)
7	1243	82.8	0.83 (0.81–0.84)	903	75.6	0.79 (0.77–0.81)	255	79.6	0.77 (0.73–0.81)	1150	77.0	0.70 (0.68–0.72)
8	489	68.7	0.60 (0.57–0.64)	504	39.3	0.34 (0.30–0.38)	98	64.3	0.57 (0.49–0.65)	974	54.1	0.56 (0.54–0.59)
9	952	55.8	0.57 (0.54–0.60)	863	69.9	0.48 (0.45–0.50)	228	72.4	0.63 (0.57–0.68)	1015	52.3	0.52 (0.49–0.55)

[a] 0=Armed forces; 1=Legislators, senior officials and managers; 2=Professionals; 3=Technicians and associate professionals; 4=Clerks; 5=Service workers and shop and market sales workers; 6=Skilled agricultural and fishery workers; 7=Craft and related trades workers; 8=Plant and machine operators and assemblers; 9=Elementary occupations.

[b] Number of observations with two sets of DISCO-88 codes at major (1-digit) group level.

[c] The proportion of true registrations within each major DISCO-88 group based on self-reported job-title as the gold standard.

[d] Agreement between registered DISCO-88 codes in DOC*X and self-reported job titles converted to DISCO-88 codes based on the Copenhagen City Heart Study.

[e] Agreement between registered DISCO-88 codes in DOC*X and self-reported job titles converted to DISCO-88 codes based on the ASUSI Cohort.

be questioned if self-reported job titles converted to DISCO-88 codes can be taken as a gold standard, but self-reported information on the current job is generally considered to have high validity (14, 39).

One limitation of our study is that we have no self-reported job titles from the years after 2004, and therefore no validation has been performed on DOC*X registrations from 2005 onwards. This limitation particularly pertains to DISCO-88 codes after the time point when Statistics Denmark introduced the DISCO-08 system in 2010 (15). Another limitation is that the DISCO-88 codes, which were available for validation, only represented around half of the codes in the DISCO-88 system so that only frequent occupational titles were validated at the 4-digit level. If the agreements are lower for rare DISCO-88 codes, we may have overestimated the general validity of the DISCO-88 codes in DOC*X. On the other hand, the sensitivities did not seem to depend on the number of observations (all ≥10) per DISCO-code.

In our analyses of agreement between exposure levels, we used categorical variables with two or three categories. The JEM exposures for wood dust and noise only exist as categorical variables while the other JEM contain continuous measures, which we categorized to ensure comparability. It may be a limitation that we only validated the DISCO-88 codes based on categorical variables instead of using continuous scales. We chose to focus on the lowest and highest exposure categories to examine whether they were correctly categorized. To the extent that DISCO-88 codes in DOC*X are misclassified so that highly exposed are categorized as medium or non-exposed, the data would not be of a quality that allows future exposure–response analyses.

Validity of DISCO-88 codes in future DOC*X studies

This study concerned selected airborne, mechanical, and physical exposures, and it remains open whether the validity of DISCO-88 codes in DOC*X is similar for other exposures, eg, chemicals. The validity varied across 4-digit DISCO-88 codes and time periods, which should be considered when planning studies in DOC*X. DOC*X also covers industry codes from 1976 and onwards (15) and it can be relevant to use those industry codes together with the DISCO-88 codes to reduce the risk of misclassification of occupations.

Concluding remarks

The validity of the DISCO-88 codes in DOC*X was generally high. Substantial agreement was found for the JEM-based exposure estimates and group-based DISCO-88 codes per se, although the DISCO-88 code-specific agreement varied across digit levels and time periods.

Funding

The Danish Working Environment Research Fund funded this study (grant no.: 43-2014-03 / 20140016763). The funding source played no role in the (i) study design, (ii) the collection, analysis and interpretation of the data, (iii) the writing of the report, or (iv) the decision to submit the paper for publication.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Johnson JV, Stewart WF. Measuring work organization exposure over the life course with a job-exposure matrix. Scand J Work Environ Health 1993 Feb;19(1):21–8. https://doi.org/10.5271/sjweh.1508.

2. Kauppinen T, Heikkilä P, Plato N, Woldbaek T, Lenvik K, Hansen J et al. Construction of job-exposure matrices for the Nordic Occupational Cancer Study (NOCCA). Acta Oncol 2009;48(5):791–800. https://doi.org/10.1080/02841860902718747.

3. Kauppinen T, Uuksulainen S, Saalo A, Mäkinen I, Pukkala E. Use of the Finnish Information System on Occupational Exposure (FINJEM) in epidemiologic, surveillance, and other applications. Ann Occup Hyg 2014 Apr;58(3):380–96.

4. 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 Sep;57(9):635–41. https://doi.org/10.1136/oem.57.9.635.

5. Greenland S, Fischer HJ, Kheifets L. Methods to explore uncertainty and bias introduced by job exposure matrices. Risk Anal 2016 Jan;36(1):74–82. https://doi.org/10.1111/risa.12438.

6. Kauppinen TP. Assessment of exposure in occupational epidemiology. Scand J Work Environ Health 1994;20(Spec No):19–29.

7. Mannetje A, Kromhout H. The use of occupation and industry classifications in general population studies. Int J Epidemiol 2003 Jun;32(3):419–28. https://doi.org/10.1093/ije/dyg080.

8. Solovieva S, Pehkonen I, Kausto J, Miranda H, Shiri R, Kauppinen T et al. Development and validation of a job exposure matrix for physical risk factors in low back pain. PLoS One 2012;7(11):e48680. https://doi.org/10.1371/journal.pone.0048680.

9. Sjöström M, Lewné M, Alderling M, Willix P, Berg P, Gustavsson P et al. A job-exposure matrix for occupational noise: development and validation. Ann Occup Hyg 2013 Jul;57(6):774–83.

10. Niedhammer I, Milner A, LaMontagne AD, Chastang JF. Study of the validity of a job-exposure matrix for the job strain model factors: an update and a study of changes over time. Int Arch Occup Environ Health 2018 Jul;91(5):523–36. https://doi.org/10.1007/s00420-018-1299-2.

11. Hanvold TN, Sterud T, Kristensen P, Mehlum IS. Mechanical and psychosocial work exposures: the construction and evaluation of a gender-specific job exposure matrix (JEM). Scand J Work Environ Health 2019 May;45(3):239–47. https://doi.org/10.5271/sjweh.3774.

12. Dale AM, Ekenga CC, Buckner-Petty S, Merlino L, Thiese MS, Bao S et al. Incident CTS in a large pooled cohort study: associations obtained by a Job Exposure Matrix versus associations obtained from observed exposures. Occup Environ Med 2018 Jul;75(7):501–6. https://doi.org/10.1136/oemed-2017-104744.

13. Dalbøge A, Hansson GA, Frost P, Andersen JH, Heilskov-Hansen T, Svendsen SW. Upper arm elevation and repetitive shoulder movements: a general population job exposure matrix based on expert ratings and technical measurements. Occup Environ Med 2016 Aug;73(8):553–60. https://doi.org/10.1136/oemed-2015-103415.

14. 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.

15. Flachs EM, Petersen SB, Kolstad HA, Schlünssen V, Svendsen SW, Hansen J et al. Cohort Profile: DOC*X: a nationwide Danish Occupational Cohort with eXposure data – an open research resource. Int J Epidemiol 2019;dyz110. https://doi.org/10.1093/ije/dyz110.

16. Denmark S. Folke og Boligtællingen 1970 [Census 1970]. 2018.

17. Petersson F, Baadsgaard M, Thygesen LC. Danish registers on personal labour market affiliation. Scand J Public Health 2011 Jul;39(7 Suppl):95–8. https://doi.org/10.1177/1403494811408483.

18. Pedersen CB. The Danish Civil Registration System. Scand J Public Health 2011 Jul;39(7 Suppl):22–5. https://doi.org/10.1177/1403494810387965.

19. Appleyard M, Hansen A, Schnohr P, Jensen G, Nyboe J. The Copenhagen City Heart Study. Østerbroundersøgelsen. A book of tables with data from the first examination (1976-1978) and a 5-year follow-up (1981-83). Scand J Soc Med Suppl 1989;170:1–160.

20. Schnohr P, Jensen G, Lange P, Scharling H, Appleyard M. The Copenhagen City Heart Study. Eur Heart J Suppl 2001;3 Suppl H:H1–83. https://doi.org/10.1016/S1520-765X(01)90110-5.

21. Hansen CD, Andersen JH. Going ill to work--what personal circumstances, attitudes and work-related factors are associated with sickness presenteeism? Soc Sci Med 2008 Sep;67(6):956–64. https://doi.org/10.1016/j.socscimed.2008.05.022.

22. Basinas I, Liukkonen T, Sigsgaard T, Andersen NT, Vestergaard JM, Galea K, et al. P096 Statistical modelling and development of a quantitative job exposure matrix for wood dust in the wood manufacturing industry. 2016;73(Suppl 1):A152–A3.

23. Basinas I, Wouters IM, Sigsgaard T, Heederik D, Spaan S, Smit LA, et al. O46-4 Development of a quantitative job exposure matrix for endotoxin exposure in agriculture. 2016;73(Suppl 1):A88-A.

24. Rubak TS, Svendsen SW, Andersen JH, Haahr JP, Kryger A, Jensen LD et al. An expert-based job exposure matrix for large scale epidemiologic studies of primary hip and knee osteoarthritis: the Lower Body JEM. BMC Musculoskelet Disord 2014 Jun;15:204. https://doi.org/10.1186/1471-2474-15-204.

25. Runge SB, Pedersen JK, Svendsen SW, Juhl M, Bonde JP, Nybo Andersen AM. Occupational lifting of heavy loads and preterm birth: a study within the Danish National Birth Cohort. Occup Environ Med 2013 Nov;70(11):782–8. https://doi.org/10.1136/oemed-2012-101173.

26. Juhl M, Larsen PS, Andersen PK, Svendsen SW, Bonde JP, Nybo Andersen AM et al. Occupational lifting during pregnancy and child's birth size in a large cohort study. Scand J Work Environ Health 2014 Jul;40(4):411–9. https://doi.org/10.5271/sjweh.3422.

27. Tabatabaeifar S, Frost P, Andersen JH, Jensen LD, Thomsen JF, Svendsen SW. Varicose veins in the lower extremities in relation to occupational mechanical exposures: a longitudinal study. Occup Environ Med 2015 May;72(5):330–7. https://doi.org/10.1136/oemed-2014-102495.

28. Vad MV, Frost P, Rosenberg J, Andersen JH, Svendsen SW. Inguinal hernia repair among men in relation to occupational mechanical exposures and lifestyle factors: a longitudinal study. Occup Environ Med 2017 Nov;74(11):769–75. https://doi.org/10.1136/oemed-2016-104160.

29. Dalbøge A, Frost P, Andersen JH, Svendsen SW. Cumulative occupational shoulder exposures and surgery for subacromial impingement syndrome: a nationwide Danish cohort study. Occup Environ Med 2014 Nov;71(11):750–6. https://doi.org/10.1136/oemed-2014-102161.

30. Dalbøge A, Frost P, Andersen JH, Svendsen SW. Surgery for subacromial impingement syndrome in relation to occupational exposures, lifestyle factors and diabetes mellitus: a nationwide nested case-control study. Occup Environ Med 2017 Oct;74(10):728–36. https://doi.org/10.1136/oemed-2016-104272.

31. Dalbøge A, Frost P, Andersen JH, Svendsen SW. Surgery for subacromial impingement syndrome in relation to intensities of occupational mechanical exposures across 10-year exposure time windows. Occup Environ Med 2018 Mar;75(3):176–82. https://doi.org/10.1136/oemed-2017-104511.

32. Svendsen SW, Dalbøge A, Andersen JH, Thomsen JF, Frost P. Risk of surgery for subacromial impingement syndrome in relation to neck-shoulder complaints and occupational biomechanical exposures: a longitudinal study. Scand J Work Environ Health 2013 Nov;39(6):568–77. https://doi.org/10.5271/sjweh.3374.

33. Dalbøge A, Frost P, Andersen JH, Svendsen SW. Cumulative occupational shoulder exposures and surgery for subacromial impingement syndrome: a nationwide Danish cohort study. Occup Environ Med 2014 Nov;71(11):750–6. https://doi.org/10.1136/oemed-2014-102161.

34. Kock S, Andersen T, Kolstad HA, Kofoed-Nielsen B, Wiesler F, Bonde JP. Surveillance of noise exposure in the Danish workplace: a baseline survey. Occup Environ Med 2004 Oct;61(10):838–43. https://doi.org/10.1136/oem.2004.012757.

35. Stokholm ZA, Bonde JP, Christensen KL, Hansen AM, Kolstad HA. Occupational noise exposure and the risk of hypertension. Epidemiology 2013 Jan;24(1):135–42. https://doi.org/10.1097/EDE.0b013e31826b7f76.

36. Lutman ME. What is the risk of noise-induced hearing loss at 80, 85, 90 dB(A) and above? Occup Med (Lond) 2000 May;50(4):274–5. https://doi.org/10.1093/occmed/50.4.274.

37. Hoozemans MJ, Burdorf A, van der Beek AJ, Frings-Dresen MH, Mathiassen SE. Group-based measurement strategies in exposure assessment explored by bootstrapping. Scand J Work Environ Health 2001 Apr;27(2):125–32. https://doi.org/10.5271/sjweh.599.

38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159–74. https://doi.org/10.2307/2529310.

39. Baumgarten M, Siemiatycki J, Gibbs GW. Validity of work histories obtained by interview for epidemiologic purposes. Am J Epidemiol 1983 Oct;118(4):583–91. https://doi.org/10.1093/oxfordjournals.aje.a113663.

40. Madsen IE, Gupta N, Budtz-Jørgensen E, Bonde JP, Framke E, Flachs EM et al. Physical work demands and psychosocial working conditions as predictors of musculoskeletal pain: a cohort study comparing self-reported and job exposure matrix measurements. Occup Environ Med 2018 Oct;75(10):752–8. https://doi.org/10.1136/oemed-2018-105151.

41. Bondo Petersen S, Flachs EM, Prescott EI, Tjønneland A, Osler M, Andersen I et al. Job-exposure matrices addressing lifestyle to be applied in register-based occupational health studies. Occup Environ Med 2018 Dec;75(12):890–7. https://doi.org/10.1136/oemed-2018-104991.

Received for publication: 14 May 2019