Job-exposure matrices addressing lifestyle to be applied in register-based occupational health studies

Petersen, Sesilje Bondo; Flachs, Esben Meulengracht; Prescott, Eva Irene Bossano; Tjønneland, Anne; Osler, Merete; Andersen, Ingelise; Juel, Knud; Budz-Jørgensen, Esben; Kolstad, Henrik A.; Schlunssen, Vivi; Bonde, Jens Peter

Methodology

ORIGINAL ARTICLE

Job-exposure matrices addressing lifestyle to be applied in register-based occupational health studies

Sesilje Bondo Petersen,[1] Esben Meulengracht Flachs,[1] Eva Irene Bossano Prescott,[2] Anne Tjønneland,[3] Merete Osler,[4] Ingelise Andersen,[5] Knud Juel,[6] Esben Budz-Jørgensen,[7] Henrik A Kolstad,[8] Vivi Schlünssen,[9,10] Jens Peter Bonde[1]

► http://dx.doi.org/10.1136/oemed-2018-105425

Check for updates

ABSTRACT

Objectives Information about lifestyle factors in register-based occupational health studies is often not available. The objective of this study was therefore to develop gender, age and calendar-time specific job-exposure matrices (JEMs) addressing five selected lifestyle characteristics across job groups as a tool for lifestyle adjustment in register-based studies.
Methods We combined and harmonised questionnaire and interview data on lifestyle from several Danish surveys in the time period 1981–2013 for 264 054 employees registered with a DISCO-88 code (the Danish version of International Standard Classification of Occupations (ISCO)-88) in a nationwide register-based Danish Occupational Cohort. We modelled the probability of specified lifestyles in mixed models for each level of the four-digit DISCO code with age and sex as fixed effects and assessed variation in terms of intraclass correlation coefficients (ICCs) and exposure-level percentile ratios across jobs for six different time periods from 1981 through 2013.
Results The ICCs were overall low (0.26%–7.05%) as the within-job group variation was large relative to the between job group variation, but across jobs the calendar period-specific ratios between highest and lowest predicted levels were ranging from 1.2 to 6.9, and for the 95%/1% and the 75%/5% percentile ratios ranges were 1.1–2.8 and 1.1–1.6, respectively, thus indicating substantial contrast for some lifestyle exposures and some occupations.
Conclusions The lifestyle JEMs may prove a useful tool for control of lifestyle-related confounding in register-based occupational health studies where lacking information on individual lifestyle factors may compromise internal validity.

INTRODUCTION

Job-exposure matrices (JEMs) have for decades been applied in studies addressing occupational risk of disease when individual exposure data are not available or too costly to collect.[1 2] A JEM is a cross-tabulation of occupations with exposure data for a certain well-defined occupational exposure in a given time window and geographical area. Information on exposure can be based on measurements, observations, expert assessments, self-reported information or combinations of those.[2 3]

There are limitations when using JEMs in epidemiological studies. First of all, JEMs do not capture variation in occupational exposure within a given

Key messages

What is already known about this subject?
► Information about lifestyle factors in register-based occupational health studies is often not available, which may raise concern about inappropriate control of confounding.

What are the new findings?
► This study describes six different job-exposure matrices (JEMs) with predicted estimates of exposure averages for lifestyle factors for specific job groups.

How might this impact on policy or clinical practice in the foreseeable future?
► The JEMs provide us with new possibilities to conduct large nationwide register-based studies controlling for lifestyle habits.

job group, and errors in job coding may also lead to misclassification.[1 4] Another limitation is the potential risk of confounding by individual characteristics and health behaviour, such as smoking, alcohol consumption and physical activity level that rarely are included in community-based studies using data retrieved from public registries and JEMs.[1 5] It is known that both socioeconomic status and educational levels are strong predictors for health and healthy lifestyle[6 7]; however, lifestyle factors also vary across and within socioeconomic status and education, and therefore it is a key issue to include information about lifestyle in studies dealing with occupational exposures and health outcomes.[8] Especially regarding smoking, the social context seems to be an underplayed factor.[9] Although level of education—readily available in many public registries—captures some of the variation in lifestyle factors, a lifestyle JEM is providing a higher level of detail corresponding to the level of detail in occupational JEMs.

In large register-based studies, it is often not economically feasible to acquire data on lifestyle information for the main part of the study subjects, and for some study subjects it is not possible because of time lag or deceased individuals. Gathering information on just a small part is also costly and time-consuming, which emphasises the advantage of alternative methods accounting for lifestyle in epidemiological studies based on register data.

Traditionally, indirect methods have been used to evaluate the magnitude and direction of the potentially confounding in observational studies, when lifestyle information is missing.[10 11] An alternative approach is development of survey-based lifestyle JEMs with predicted gender, age, calendar time and job-specific estimates of lifestyle characteristics, even when jobs are not direct causes of certain lifestyles (as, for instance, higher body mass index (BMI) in sedentary work and higher alcohol consumption in brewery workers). The need in JEM-based occupational studies addressing health effects of specific and explicit workplace exposures is to obtain systematic information on the distribution of potentially confounding lifestyle factors across job titles at the same level of detail as used for the construction of occupational exposures. The present study is the first, as we know, that aims to develop and document lifestyle JEMs on a very large study population by combining questionnaire and interview data from several Danish research studies.

The objective of the present study was to develop JEMs addressing smoking, alcohol consumption, leisure-time physical activity, intake of fruits and vegetables, and BMI across job groups as a tool for lifestyle adjustment in register-based JEM studies where individual lifestyle information is not available and where adjustment for one or more lifestyle factors is essential considering study populations and outcomes. Thus the aim is not to examine the occurrence of health behaviours in certain jobs per se, but rather to enable adjustment for lifestyle in studies addressing workplace exposures where individual lifestyle information is unavailable.

METHODS
Study population
We retrieved individual self-reported data on lifestyle from four large Danish population-based studies (figure 1). Combined questionnaire and interview data on lifestyle were available from the years 1981 to 2013 (table 1). In total, we had 16 different data files from the respective surveys. They were merged into
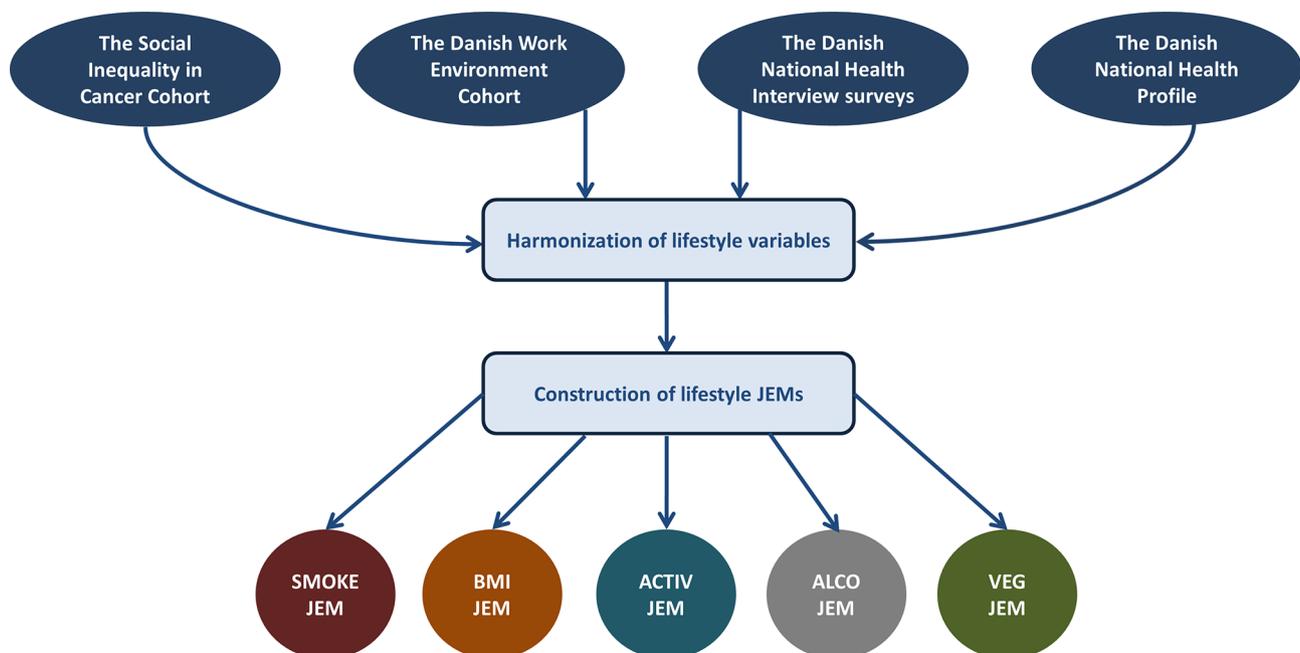
one data file by use of the Danish personal identification number (CPR).[12]

Social Inequality in Cancer Cohort (SIC) comprises pooled data from seven Danish cohorts including 83 006 individuals aged 20–93 years of age at entry, primarily living in the Danish

Table 1 Observations and individuals in the aggregated study population by time and data source

Year	SIC	DNHI	DWEC	DNHS	Cumulative number of observations	Cumulative number of individuals
1981	3089				3089	3089
1982	3985				7074	7074
1983	4098				11 172	11 172
1984	24				11 196	11 196
1986	899				12 095	12 095
1987	288	3063			15 446	15 350
1990			7019		22 465	22 186
1991	1288	3076			26 829	26 451
1992	128				26 957	26 579
1993	80				27 037	26 659
1994	3326	2943			33 306	32 844
1995	13 636		6774		53 716	48 011
1996	17 294				71 010	63 305
1997	7418				78 428	72 723
1999	3133				81 561	75 856
2000	2194	9997	6003		99 755	87 200
2001	139				99 894	87 339
2005		7701	8013		115 608	97 522
2010			10 312	92 126	218 046	191 319
2013				81 104	299 150	264 054

DNHI, Danish National Health Interviews; DNHS, Danish National Health Survey; DWEC, Danish Work Environment Cohort; SIC, Social Inequality in Cancer Cohort.

Figure 1 Overview of the data flow linking the respective cohort studies in one data file, where in the lifestyle job-exposure matrices (JEMs) are generated. BMI, body mass index.

cities of Aarhus or Copenhagen. Lifestyle data were harmonised across the seven cohorts (n=76 294) as described elsewhere.[13]

The Danish Work Environment Cohort (DWEC) studies from the National Research Centre for Working Environment were initiated in 1990 and include nationwide random samples of employees aged 18–59.[14–16]

The Danish National Health Interviews (DNHI) surveys from the Danish National Institute of Public Health were carried out from 1987 through 2005. All samples were randomly drawn using the Danish Civil Registration System. Study design and characteristics of the DNHI surveys have been described elsewhere.[17 18]

The Danish National Health Survey (DNHS) was carried out in 2010 and 2013 at the Danish National Institute of Public Health, together with the five Danish regions. National representative surveys were conducted to provide an overview of the health, morbidity and well-being of the Danish adult population.[17 19]

Assessment of occupation

Individual information of occupation for all individuals in the study was available from the DOC*X database at Statistics Denmark,[20] and linked to the survey data by the use of the CPR.[21] In DOC*X, all employed Danish citizens are included from the age of 15, with annual information on occupational code, industry code and income. All occupational codes in DOC*X have been harmonised to the DISCO-88 classification system based on the four-digit International Standard Classification of Occupations (ISCO)-88 classification system,[22] including 372 job groups. For descriptive purposes, we grouped DISCO-88 codes according to procedures from Statistic Denmark.[23 24]

Permission to link variables from the included cohorts and surveys with the DOC*X database was obtained from the Danish Data Protection Agency.

Assessment of lifestyle factors

Information about lifestyle factors was extracted and harmonised from questionnaire and interview data from the four data sources. We assembled a categorical variable for current smoking (0=non-smoker; 1=smoker), where ex-smokers were defined as non-smokers. Among smokers we calculated the total amount of tobacco smoked per day (gram of tobacco for each unit: cigarette=1, cheroot=3, cigar=4; pipe=3). Alcohol consumption was calculated as units of alcohol per week (12 g of alcohol in a unit). We further made a categorical variable for alcohol (0=0 units/week; 2= >0–7 units/week; 3= >7–14 units/week; 4= >14 units/week). Leisure-time physical activity was combined in one categorical variable (1=sedentary activity (≈ no sport/training); 2=low/easy waking or biking (≈ 1–2 hours/week); 3=moderate training (≈ 2–4 hours/week); 4=hard training/competitive sport (≈ >4 hours/week)). Information on height and weight was available either from questionnaires/interviews or clinical examination data for the calculating of the BMI (kg/m^2). Vegetable consumption was calculated as the highest frequency of fruit or vegetables indicated in the question(s) of each data material. The frequencies were divided into three groups for frequency of total intake of fruits and vegetables per week (1=never/rarely; 2=1–6 per week; 3=daily).

Statistical methods

We restricted our study sample to men and women with an occupational code registered in the DOC*X cohort for the same year as for completion of either the questionnaire or interview.

The exposure level in a job group was calculated as best linear unbiased predictions (BLUPs) by fitting mixed models in SAS (V.9.4, SAS Institute). We used linear models (Proc Mixed procedure) except for probability of smoking, where we used mixed logistic regression (Proc Glimmix, link=logit, dist=binary). Age group (1= <30; 2=30–39; 3=40–49; 4= ≥50 years of age), gender (men/women) and source of data[1–16] were included as fixed effects and DISCO-88 codes as random effect. Data were divided into 5-year intervals, except for the first 10-year interval with less participants (1=1981–1990; 2=1991–1995; 3=1996–2000; 4=2001–2005; 5=2006–2010; 6= >2010). Only job -groups with at least 10 observations were included. If <10 observations were available for a DISCO-code, the predicted value for the less detailed DISCO-code was imputed to the final JEM (eg, the predicted value for the DISCO-code 931 was imputed for the missing value of 9311). The final JEMs included predicted values for each combination of DISCO-code, gender, age group and time period. Furthermore, the number of study subjects used for the predictions and the SD for the predicted measures was included. In this article, only the results for the most detailed level of DISCO are presented.

For evaluation of each JEM on the most detailed four-digit DISCO-level, we calculated the intraclass correlation coefficient (ICC) and the highest/lowest, 95/5 and 75/5 percentile ratios for each time period. The ICC is equal to the variance between the job groups divided by the sum of the variance within and between job groups plus the residual value, and it ranges from 0 to 1 (all variation is between job groups).

We calculated Spearman correlations between JEMs to investigate any agreement between the different lifestyles.

To illustrate that the lifestyle JEMs carries independent information in addition to education, we made analyses using sex, age and the cumulated smoking at age 50 (by the smoking proportion JEM, categorised according to quartiles) as predictors, both with and without including education in Poisson regression of all-cause mortality. Education was defined as highest attained education at age 50, categorised as short (primary or secondary school or vocational), medium or long.

The population included was persons from the Danish population (employed between 1976 and 2015) with at least 20 years of (JEM-)recorded smoking exposure before the age of 50, followed -up from first employment between age 50–60 and until 2015 or death whichever came first. In total, 976 264 persons were followed for 9 458 032 years (43 326 deaths).

RESULTS

In our study sample, 57.4% of the subjects had a job title registered in the DOC*X database for the current year, which overall resulted in a final sample of 264 054 study subjects for construction of the JEMs (table 1). Characteristics of the surveys are provided in table 2.

The calendar period-specific 95/5 and 75/25 percentile ratios were 1.5–2.8 and 1.6–1.9, respectively (table 3).

Smoking

We saw a linear decline in the proportion of smokers from 56% in the period 1981–1990 to 19% after 2010 (table 3), whereas the amount of tobacco consumed by smokers only changed slightly across the years (table 3).

The predicted proportion of smokers ranged from 6% to 40% between job groups for the latest period (>2010) with a ratio between highest and lowest predicted level of 6.8, while

Table 2   Distribution of gender, age and lifestyle by time period and data source

Time period	Age (years)	Gender (% men)	Smoking (% smokers)	Smoking (g/day among smokers)	Alcohol (units/week)	BMI (kg/m^2)	Physical activity†	Fruits and vegetables†
1981– 1990								
SIC81-87	47.7*	53.1***	58.6*	9.8*	9.8*	25.0*	2.1	–
DNHI87	38.2	53.8	49.2	7.5	7.9	23.5	–	–
DWEC90	37.6	52.1	46.9	7.4	–	23.5	–	–
1991–1995								
SIC91-95	54.4*	52.1***	39.9*	7.2*	12.5*	25.9*	2.4*	2.6
DNHI91	38.6	51.5	58.4	5.3	–	–	2.2	–
DNHI94	39.0	52.2	43.6	6.7	8.7	24.0	2.2	–
DWEC95	38.9	51.8	40.8	6.3	–	24.0	–	–
1996–2000								
SIC96-00	53.4*	52.5*	32.3*	5.5***	11.8*	26.0*	2.6*	2.6*
DNHI00	39.4	50.7	57.0 ‡	8.8‡	8.7	24.6	2.2	2.5
DWEC00	39.2	49.8	36.7	5.8	5.9‡	24.4	2.6	2.7
2001–2005								
SIC01-05	36.7*	38.8*	36.7***	5.5***	6.4**	25.6***	1.9*	–
DNHI05	41.2	49.5	28.1	4.4	9.5	24.9	2.3	2.6*
DWEC05	41.5	46.1	27.6	4.2	9.7	24.9	2.1	2.7
2006–2010								
DWEC10	43.5*	45.1***	22.7***	2.8***	10.2*	25.4**	2.5*	2.8*
DNHS10	42.6	45.3	23.0	2.7	7.9	25.3	2.3	2.7
>2010								
>2010 DNHS13	44.3	45.5	19.1	2.1	7.0	25.4	2.3	2.7

*p < 0.001 in an ANOVA (continuous variable) or χ^2 test (class variable) for each time period.

**p < 0.01 in an ANOVA (continuous variable) or χ^2 test (class variable) for each time period.

***p > 0.05 in an ANOVA (continuous variable) or χ^2 test (class variable) for each time period.

† Data censored from the final study population due to unrealistic estimates (not included in ANOVA or χ^2 tests).

‡ Frequency on a scale of 1–4 physical activity (sedentary activity (≈ no sport/training)/ low/easy waking or biking (≈ 1–2 hours/week)/moderate training (≈ 2–4 hours/week)/ hard training/competitive sport (≈ >4 hours/week)) and 1–3 for the frequency of eating fruits and vegetable (never/rarely/1–6 per times per week/daily).

ANOVA, analysis of variance; DNHI, Danish National Health Interviews; DNHS, Danish National Health Survey ; DWEC, Danish Work Environment Cohort; SIC, Social Inequality in Cancer Cohort.

the amount of smoking ranged from 8.0 to 19.5 g/day, corresponding to a ratio on 2.4 (figure 2).

The calculated ICCs for the SMOKE-JEM for proportion increased linearly from 2.7% in the first time period to 5.7% in the last, which indicates increased variability between the job groups with calendar year. Similarly the ICCs increased by time period in the SMOKE-JEM for amount of smoking (table 3).

Alcohol
Information on alcohol was available from 1981 until 2013, and no major changes in the overall intake of alcohol were observed (table 3).

Average consumption of alcoholic beverages varied between the job groups with predicted values ranging from 2.4 to 10.0 units/week (ratio=3.7) (figure 2). The calculated ICCs were small with decreased ICCs by time period from 3.80% to 1.07% (table 3).

Body mass index
In the first time period, only 23% (n=84) of the DISCO codes had BMI information at the most detailed four-digit DISCO level, whereas in the latest time period BMI information was available for 71% (n=264). The mean BMI increased almost linearly from 24.2 kg/m^2 in the first time period to 25.2 kg/m^2 in the latest time period (table 3). In general, men had a higher BMI than women and the BMI increased by age in all time periods. The

ICCs were small with no time trend (table 3), and the predicted values ranged from 21.3 to 28.0 kg/m^2 with a ratio of 1.3. Analysis of the relationship between matrix estimates for BMI and socioeconomic status indicated a positive linear relationship (R^2=0.46 in >2010).

Leisure-time physical activity
Information on leisure-time physical activity was available from 1981 until 2013 but only small differences in the mean levels were found across the time periods (table 3). The level of physical activity was highest among the oldest and youngest age groups before the year 2000, but thereafter the level was lowest among the oldest age group. In general, men had a higher level of physical activity than women. The ICCs varied from 0.26% to 2.21% with no clear time trend (table 3). The predicted levels in the job groups ranged from 1.7 to 3.2 (>2010) with a ratio on 1.8.

Fruits and vegetables
Information on intake of fruits and vegetables was available from 1991 until 2013. We saw a small increase in the predicted intake during the time periods from 2.5 to 2.7 on a scale ranging from 1 to 3 (table 3). In general, women had a higher intake than men, and the intake of fruits and vegetables increased by age. The ICCs were small without any time trend (table 3). The average intake of fruits and vegetables ranged from 2.4 to 2.9 with a ratio of 1.2 in the latest time period.

Table 3   Crude and best linear unbiased prediction (BLUP) statistics by lifestyle factor across calendar time

JEM	Crude data (surveys)		BLUP statistics applied to the entire population			Ratio of percentiles (high/low, %)		
	Mean	SD	Mean	SD	ICC (%)	100/0	95/5	75/25
SMOKE-JEM, proportion								
1981–1990	0.53	0.50	0.56	0.06	2.66	191.7	146.2	117.7
1991–1995	0.41	0.49	0.39	0.07	3.23	306.0	177.9	125,5
1996–2000	0.33	0.47	0.34	0.07	3.52	372.7	191.5	132.4
2001–2005	0.28	0.45	0.27	0.06	6.53	447.6	238.0	128.0
2006–2010	0.23	0.42	0.22	0.06	6.00	504.5	273.9	155.0
>2010	0.19	0.39	0.19	0.06	5.73	689.1	280.6	148.7
SMOKE-JEM, amount (g/day)								
1981–1990	16.1	8.8	15.9	2.3	2.28	225.1	159.8	124.7
1991–1995	16.9	9.7	16.5	2.9	2.85	265.5	175.6	132.5
1996–2000	16.2	9.6	15.6	2.8	2.80	300.4	191.9	134.8
2001–2005	15.5	8.1	15.2	2.2	4.83	253.2	161.0	123.7
2006–2010	14.1	7.4	14.0	2.1	7.05	257.3	164.1	122.0
>2010	13.6	8.1	13.2	1.8	5.04	244.0	158.4	122.8
ALKO-JEM, drinks/day								
1981–1990	1.4	1.0	1.5	0.4	3.80	434.2	210.2	158.8
1991–1995	1.7	0.9	1.6	0.3	2.79	274.7	188.1	141.5
1996–2000	1.7	0.9	1.6	0.3	2.12	276.5	181.4	135.2
2001–2005	1.5	1.0	1.6	0.3	1.95	291.9	184.1	137.1
2006–2010	1.4	0.9	1.4	0.3	1.34	304.5	199.4	134.3
>2010	1.3	0.8	1.3	0.2	1.07	266.6	188.1	131.7
BMI-JEM, kg/m^2								
1981–1990	24.2	3.8	24.2	1.3	3.14	130.3	120.2	108.5
1991–1995	25.3	3.9	24.7	1.2	3.53	132.3	116.8	107.3
1996–2000	25.5	4.0	24.3	1.1	2.81	127.3	115.1	106.1
2001–2005	24.9	4.0	24.8	1.0	2.18	123.5	113.8	105.5
2006–2010	25.3	4.4	25.1	1.2	2.96	131.8	117.9	107.1
>2010	25.4	4.5	25.2	1.2	2.71	131.7	118.2	107.3
ACTIVE-JEM, scale 1–4								
1981–1990	2.1	0.7	2.2	0.2	2.01	166.8	137.2	116.0
1991–1995	2.4	0.8	2.4	0.2	0.95	137.0	123.6	110.3
1996–2000	2.4	0.8	2.3	0.1	0.26	125.0	118.5	108.1
2001–2005	2.2	0.7	2.1	0.2	2.02	152.8	128.9	112.2
2006–2010	2.3	0.7	2.3	0.2	1.87	174.5	130.1	112.0
>2010	2.3	0.7	2.3	0.2	2.21	186.5	133.9	113.4
VEG-JEM, scale 1–3								
1991–1995	2.6	0.5	2.5	0.1	2.23	131.1	119.3	110.8
1996–2000	2.6	0.5	2.5	0.1	1.82	134.5	122.1	110.5
2001–2005	2.7	0.5	2.7	0.1	2.73	126.8	115.0	108.7
2006–2010	2.7	0.5	2.7	0.1	2.88	125.2	114.6	107.0
>2010	2.7	0.5	2.7	0.1	2.28	122.5	113.0	106.6

JEM, job-exposure matrix.

Performance of the lifestyle JEMs (an example)

Most lifestyle factors are moderately correlated, though smoking and activity and BMI, and alcohol and BMI are only very weekly correlated (online supplemental table S1). The JEM-based cumulative smoking exposure is a strong predictor of death from all causes, even when simultaneously controlling for education (online supplemental table S2).

DISCUSSION

We created health behaviour JEMs based on 264054 study subjects from different studies with interview and questionnaire data on smoking, alcohol consumption, leisure-time physical activity, BMI, and intake of fruits and vegetables. The between-job group variation was <5%–10% of the total variation, but across job groups a substantial contrast was evident with ratios between highest and lowest predicted levels ranging from 1.2 to 6.8 and with 95/5 and 75/25 percentiles ranges of 1.5–2.8 and 1.6–1.9, respectively. The analysis of all-cause mortality predicted by the smoking JEM with and without adjustment for length of education illustrates that the JEM carries substantial independent information in addition to usual lifestyle proxies as education.

Strengths and limitations

Our study is by far the biggest Danish study on lifestyle exposures measured by pooling interview and questionnaire data
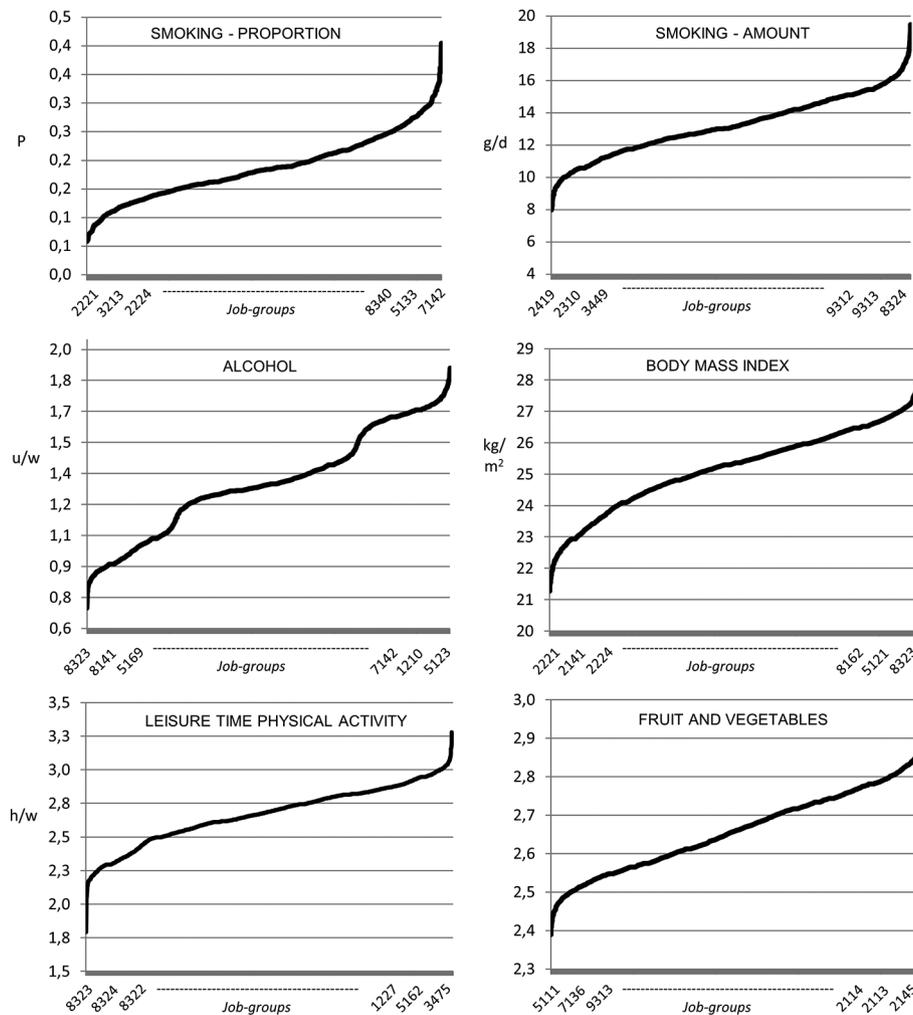
Figure 2 The population distribution of the predicted values of the lifestyle job-exposure matrices across all job groups with specification of the three jobs with lowest and highest predicted values for each lifestyle JEM. 1210=directors and chief executives; 1227=production and operations department managers; 2113=chemists; 2114=geologists and geophysicists; 2141=architects, town and traffic planners; 2145=mechanical engineers; 2221=medial doctors; 224=pharmacists; 2310=college, university and higher education teaching professionals; 2419=business professionals; 3213=farming and forestry advisers; 3449=customs, tax and related government associate professionals; 3475=athletes and sportspersons; 5111=travel and attendants; 5121=housekeepers and related workers; 5123=waiters and bartenders; 5133=home based personal care workers; 5162=police officers; 5169=protective service workers; 7136=plumbers and pipe fitters; 7142=varnishes and related painters; 8141=wood processing-plant operators; 8162=steam engine and boiler operators; 8322=car, taxi and van drivers; 8323=bus and tram drivers; 83244=heavy truck and lorry drivers; 8340=ships', deck crews and related workers; 9312=construction and maintenance labourers; 9313=building construction labourers.

from several Danish surveys. The large study sample allows for estimation of lifestyle factors for about 70% of the entire workforce on a detailed DISCO level. The large time span provides time-specific estimates, which is a key issue, as lifestyle habits have changed significant in the Danish population since 1981.[25][26] The lifestyle JEMs were refined by including gender, age and calendar period in the BLUP models which contributed significantly and thus served to decrease misclassification inherent in the JEM approach.

The construction of lifestyle JEMS relies in linkage of different cohort studies. Repeated cross-sectional surveys with fixed questions over time would be a useful alternative source but unfortunately no cross-sectional surveys with repeated information on both lifestyle and occupational titles are readily available.

The survey data included have used different recruitment strategies. The SIC consists of data from several cohort studies where the participants have been invited according to age group

and living in defined areas of Copenhagen and Aarhus[12] —the two largest cities in Denmark. The SIC data are therefore not representative for the whole country. The study populations of the DNHI, DWEC and DNHS have used other recruitment strategies, but overall with the purpose to create nationally representative samples. Overall, the number of study subjects and the representativeness increased by time period in our data, which may influence the estimated exposure values in the JEMs. The low number of jobs traditionally placed at the country side may leave us with less precise estimates for those jobs. For a given job group, however, we do not believe that the various sampling strategies introduce noteworthy bias. Participation bias may also be introduced in our estimates as a consequence of the healthy participant effect. We know from other studies that people participating in research studies in general are healthier than people who choose not to participate.[27] If the healthy participant effect is introduced in our data, the predicted values for

lifestyle factors may not reflect the general workforce, but only the healthiest part of it.

Another limitation in our study is that the questions used for estimation of lifestyle exposures differ between surveys. BMI is calculated in the same way for all surveys, but in the SIC cohort a large part of the measurements for weight and height is from clinical examinations. In the questionnaire surveys, weight and height are self-reported. We saw a tendency to lower BMI in self-reported data compared with data from clinical examinations, indicating that people may underestimate their weight and overestimate their height in self-administered questionnaires. The systematic bias in BMI is supported by findings in other studies.[28 29]

Between-group and within-group variation

The between-group variation in health behaviours was small compared with the within-group variation as reflected in ICC values <5%–10%. This is expected because the studied unhealthy behaviours are rather prevalent regardless of type of occupation which is in contrast to rare occupational exposures.[30 31] Small differences in the occurrence of unhealthy behaviours in the most prevalent occupations also contribute to small ICCs when calculations are based on the entire population. ICCs for subsets of the population defined to increase the occupational exposure contrast of interest may have substantially higher ICCs as evidenced by the ratio of unhealthy behaviour between jobs with highest and the lowest occurrence, which for some lifestyle factors proved substantial.

For smoking the ICC increased by time period, which indicated larger between-group variation by calendar year. During the 2000s, we saw major changes in the Danish Society with respect to smoking habits as educational institutions, the public transport systems, private companies, etc, began to introduce smoking policies. Furthermore, the Danish government introduced a nationwide smoking policy in 2007 with the purpose of limiting smoking at public places.[32] However, this is not directly reflected in our smoking data since we saw an almost constant decline in both smoking proportion and amount of smoking from the time period 1991–1995.

For alcohol consumption, we saw the opposite pattern with decreased ICC by time period, indicating that alcohol consumption in the last decade is happening equally in all socioeconomic status groups and in all job groups, as we did not see a social gradient for alcohol consumption. For BMI, leisure-time physical activity and intake of fruits and vegetables, the ICCs were in general small and did not show a clear time-dependent pattern. This indicates that there have been no major changes in the variability between job groups for the lifestyle parameters during the time periods. However, the low ICCs could also indicate that the harmonisation of the interview and questionnaire data for leisure-time physical activity and intake of fruits and vegetables has been inappropriate to measure the variability between job groups.

How to use the JEMs

The lifestyle JEM can be used simply as a systematic and transparent external information on average health behaviours in specific jobs or—depending on the research questions—it can be applied to the entire national population or subsets of the population defined by job title to obtain the optimal trade-off between contrast of exposure and statistical power. Thus the gender, age and calendar time-specific lifestyle JEMs are intended as tools to address potential confounding in occupational register-based studies where workplace exposure assessment is assigned by JEMs. The objective was not to provide a tool for registry-based studies of lifestyle per se, but rather to enable adjustment for lifestyle in studies addressing workplace exposures where individual lifestyle information is unavailable. In addition to community-based studies, this also includes occupational cohorts with industry-specific JEMs, but without information on lifestyle. The application of lifestyle JEMs will introduce non-differential misclassification and less efficient control of confounding than use of individual data, but since the behaviour JEMs are developed for use in register-based occupational JEM studies, the misclassification of lifestyle factors is balanced to the misclassification of the exposure of interest. Hereby the lifestyle JEMs represent a systematic alternative to other methods for control for lifestyle-related confounding in occupational studies such as ad hoc adjustment of excess risk based on hypothetical assumptions about health behaviours in exposed and controls[10] or use of antecedent ad hoc data on health behaviours that often are based on small samples. However, since health behaviours and lifestyle patterns are highly time-specific and country-specific the JEMs are not a priori useful in other countries.[33]

The estimated values for each job group can be applied to individual study subjects according to job title, gender, age and calendar time. The estimated values in the JEMs are not intended as exact exposure measures but should rather be seen as relative measures between job groups since they are not validated against external data sources. DOC*X is an open research resource, which means that researchers from all over the world can get access to the data after approval from the DOC*X steering committee, the Statistics Denmark and the Danish Data Protection Agency. More information can be found at www.doc-x.dk.

Author affiliations
[1]Department of Occupational and Environmental Medicine, Bispebjerg University Hospital, Copenhagen, Denmark
[2]Department of Cardiology, Bispebjerg University Hospital, Copenhagen, Denmark
[3]Danish Cancer Society Research Center, Copenhagen, Denmark
[4]Research Centre for Prevention and Health, Glostrup, Denmark
[5]Section of Social Medicine, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
[6]National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark
[7]Section of Biostatistics, Univetrsity of Copenhagen, Copenhagen, Denmark
[8]Department of Occupational Medicine, Danish Ramazzini Centre, Aarhus University Hospital, Aarhus, Denmark
[9]National Research Centre for Working Environment, Copenhagen, Denmark
[10]Department of Public Health, Danish Ramazzini Centre, Aarhus University, Aarhus, Denmark

REFERENCES

1   Greenland S, Fischer HJ, Kheifets L. Methods to explore uncertainty and bias introduced by job exposure matrices. *Risk Anal* 2016;36:10.
2   Kauppinen T, Heikkilä P, Plato N, *et al*. Construction of job-exposure matrices for the Nordic Occupational Cancer Study (NOCCA). *Acta Oncol* 2009;48:791–800.
3   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.
4   Kauppinen TP, Mutanen PO, Seitsamo JT. Magnitude of misclassification bias when using a job-exposure matrix. *Scand J Work Environ Health* 1992;18:105–12.
5   Blair A, Stewart P, Lubin JH, *et al*. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. *Am J Ind Med* 2007;50:199–207.
6   Balia S, Jones AM. Mortality, lifestyle and socio-economic status. *J Health Econ* 2008;27:1–26.
7   Pollitt RA, Rose KM, Kaufman JS. Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: a systematic review. *BMC Public Health* 2005;5:7.
8   Contoyannis P, Jones AM. *Socio-economic status, health and lifestyle*. 23 edn, 2004:965–95.
9   Poland B, Frohlich K, Haines RJ, *et al*. The social context of smoking: the next frontier in tobacco control? *Tob Control* 2006;15:59–63.
10 Steenland K, Beaumont J, Halperin W. Methods of control for smoking in occupational cohort mortality studies. *Scand J Work Environ Health* 1984;10:143–9.
11 Axelson O, Steenland K. Indirect methods of assessing the effects of tobacco use in occupational studies. *Am J Ind Med* 1988;13:105–18.
12 Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541–9.
13 Nordahl H, Hvidtfeldt UA, Diderichsen F, *et al*. Cohort profile: the Social Inequality in Cancer (SIC) cohort study. *Int J Epidemiol* 2014;43:1750–8.
14 Burr H, Bjorner JB, Kristensen TS, *et al*. Trends in the Danish work environment in 1990-2000 and their associations with labor-force changes. *Scand J Work Environ Health* 2003;29:270–9.
15 Det Nationale Forskningscenter for Arbejdsmiljø [National Research Centre for the Working Environment]. Den Nationale Arbejdsmiljøkohorte [The Danish Work Environment Cohort] - NAK. 2005 http://www.arbejdsmiljoforskning.dk/da/arbejdsmiljoedata/nak2005.
16 van der Ploeg HP, Møller SV, Hannerz H, *et al*. Temporal changes in occupational sitting time in the Danish workforce and associations with all-cause mortality: results from the Danish work environment cohort study. *Int J Behav Nutr Phys Act* 2015;12:71.
17 Christensen AI, Ekholm O, Glümer C, *et al*. The Danish National Health Survey 2010. Study design and respondent characteristics. *Scand J Public Health* 2012;40:391–7.
18 Ekholm O, Hesse U, Davidsen M, *et al*. The study design and characteristics of the Danish national health interview surveys. *Scand J Public Health* 2009;37:758–65.
19 Sundhedsstyrelsen [Danish Health Authority]. Den Nationale Sundhedsprofil 2013 [The Danish National Health Profile 2013]. 2014 https://www.sst.dk/da/udgivelser/2014/~/media/D7E7C675DAB04A37A8302CEAEF090181 (accessed 01 Apr 2017).
20 The Danish Occupational Cohort - DOC*X Få adgang til forskningsressourcer om arbejdsmiljø i Danmark fra 1964 - 2013. 2017 http://doc-x.dk/ (accessed 01 Apr 2017).
21 Pedersen CB. The Danish civil registration system. *Scand J Public Health* 2011;39:22–5.
22 Elias P. *Occupational Classification (ISCO-88) - Concepts, methods, reliability, validity and cross-national comparability. OECD Labour Market and Social PolicyOccasional Papers*. Paris: OECD Publishing, 1997.
23 Statistics Denmark.. *DISCO-88: statistics Denmark's standard classification of occupations*: Danmarks Statistik, 1996.
24 Danmarks Statistik [Statistics Denmark]. *SOCIO Danmarks Statistiks Socioøkonomiske Klassifikation [Statistics Denmark Socioeconomic Classification]*. Denmark, 1997.
25 Osler M, Prescott E, Gottschau A, *et al*. Trends in smoking prevalence in Danish adults, 1964-1994. The influence of gender, age, and education. *Scand J Soc Med* 1998;26:293–8.
26 Gerdes LU, Brønnum-Hansen H, Osler M, *et al*. Trends in lifestyle coronary risk factors in the Danish MONICA population 1982-1992. *Public Health* 2002;116:81–8.
27 Larsen SB, Dalton SO, Schüz J, *et al*. Mortality among participants and non-participants in a prospective cohort study. *Eur J Epidemiol* 2012;27:837–45.
28 Ng SP, Korda R, Clements M, *et al*. Validity of self-reported height and weight and derived body mass index in middle-aged and elderly individuals in Australia. *Aust N Z J Public Health* 2011;35:557–63.
29 Niedhammer I, Bugel I, Bonenfant S, *et al*. Validity of self-reported weight and height in the French GAZEL cohort. *Int J Obes Relat Metab Disord* 2000;24:1111–8.
30 Peters S, Vermeulen R, Portengen L, *et al*. SYN-JEM: a quantitative job-exposure matrix for five lung carcinogens. *Ann Occup Hyg* 2016;60:795–811.
31 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.
32 Aarestrup AK, Due TD, Juel K. *Evaluering af loven om røgfri miljøer: lov nr. 512 af 6 juni 2007 [Evaluation of the law on smoke free environments: law nr. 512 of 6 June 2007]*. Copenhagen: Sundhedsstyrelsen [Danish Health Authority] and Statens Institut for Folkesundhed [National Institute of Public Health], 2009.
33 Zatoński W, Przewoźniak K, Sulkowska U, *et al*. Tobacco smoking in countries of the European Union. *Ann Agric Environ Med* 2012;19:181–92.