Occupational exposure assessment in the general population:

improvements, innovations, and impact

Calvin B Ge

Occupational exposure assessment in the general population:
improvements, innovations, and impact

Calvin Benjamin Ge

Doctorate thesis, Utrecht University

Cover art and layout by: Nan He.

Printing, binding, and content layout by:
Print Service Ede, Ede, the Netherlands.

ISBN: 978-90-831295-5-6

Occupational exposure assessment in the general population:
improvements, innovations, and impact

Karakterisering van beroepsmatige blootstelling in de algemene populatie:
verbeteringen, innovaties en impact

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 4 maart 2021 des middags te 12.45 uur

door

Calvin Benjamin Ge

geboren op 22 december 1982
te Shenzhen, China

Promotor:
Prof. dr. R.C.H. Vermeulen

Copromotor:
Dr. S.M. Peters

Table of contents

CHAPTER 1

General Introduction

Since the agricultural revolution around 10,000 years ago, humans have been hard at work. Specialized jobs, trades, and careers have become such an important part of our identity that some of the most common surnames around the world describe occupations (e.g. Bakker in Dutch, Müller in German, Smith in English). Because those who work typically spend a significant portion of their lives at work, it is no surprise that occupational exposures account for a significant share of non-genetic risk factors for disease. The Global Burden of Disease (GBD) study estimated that more than 850,000 deaths globally in 2017 were due to occupational risk factors for non-communicable chronic diseases (NCD) such as various cancers and respiratory diseases (Stanaway et al., 2018). The real overall burden of chronic occupational diseases is likely even larger, as the GBD estimates did not include cardiovascular, mental, metabolic, and neurodegenerative diseases, yet studies have suggested that occupational exposures such as stress, light-at-night and particulate matter may contribute to these conditions (Fang et al., 2010; Hall et al., 2018; Shan et al., 2018; Sieurin et al., 2018).

Benefits of advances in occupational health research extend beyond the workplace: many harmful environmental exposures were first identified as hazardous to health in occupational settings. Benzene, for instance, was first identified as a potent bone marrow poison in workplaces in the late 19th century before becoming a well-established environmental pollutant in air, food, and drinking water (EEA, 2008). In fact, among human carcinogens classified by the International Agency for Research on Cancer (IARC), 70 of 120 exposures classified as Group 1 human carcinogens were likely encountered in the workplace, and 59 evaluations were supported by epidemiological evidence from occupational studies (Loomis et al., 2018). Better quantification of the risks and impacts of occupational exposures is an important element in promoting well-being both in the workplace and the overall community.

In the post-industrial and digital age, however, the field of occupational health research faces challenges both from the past and future. "Classical exposures" that began in the industrial era such as asbestos, benzene, and crystalline silica continue to occur and cause disease around the world (albeit with very different prevalence depending on region). At the same time, the "classical exposure scenarios" where many workers have lifelong trades with few job changes are quickly disappearing in developed nations. New chemicals are also being produced and put into use at an astonishing rate (Vermeulen et al., 2020). More complex exposure patterns involving novel exposures, more job changes, and simultaneous jobs are increasingly likely in the modern workforce. This trend will likely continue into the future and adds to the existing challenge of accurately characterizing occupational exposures and their relationships with human diseases.

It is intuitive to investigate the health effects of occupational exposures in specific workplaces where theses exposures occur. In fact, industry-based studies on occupational exposure and diseases have provided significant knowledge in the field. Within a specific workplace or industry, exposures may be directly measured or estimated from company records or by industry experts. There are, however, several important disadvantages that limit industry-based studies' utility in causal inference and risk estimation for occupational exposure and disease. For instance, industry-based studies may not have access to important subject information related to past occupational exposures, socioeconomic status, and lifestyle factors such as smoking. Sample sizes in industry-based studies are also limited by the size of the relevant industries. A cohort of offshore oil and gas workers may provide only a few hundred subjects to investigate exposures in petroleum production such as benzene. The low statistical power is particularly restrictive for studying rare conditions with long latency periods, such as cancers and neurodegenerative diseases.

Studies performed in the general population offer several advantages that complement the shortcomings of industry-based studies. In population-based studies, full subject occupational histories are typically collected, which allows for assessment of lifetime occupational exposures to the substance of interest, as well as other exposures as potential confounders. In addition, population-based studies typically have more complete information on non-occupational factors such as education and lifestyle habits. Because of the wide variety of jobs and industries included, the range of exposure in population-based studies is typically broader compared to industry-based studies, which usually focuses on workplaces with higher exposures. Better coverage of lower concentration ranges enables the detection and description of exposure-disease relationships at lower exposure levels and makes the results from population-based studies more generalizable to the general public.

Some of the key advantages of population-based studies are related to their potential for a large sample size. Unlike industry-based studies, study sizes in the general population are not limited by the number of workers in particular industries. In fact, individual population-based studies may be pooled to obtain samples sizes in the tens of thousands or even millions (EPHOR, 2020; Kogevinas et al., 2020; Peters et al., 2020). In addition to giving more statistical power for the exploration and characterization of exposure-disease relationships, a large sample size allows for better investigation of differential risks among exposure subgroups, such as subjects with different physical characteristics (e.g. female, specific genotype), temporal exposure characteristics (e.g. exposure age, exposure rate), and different socioeconomic levels or lifestyle habits. For case-control studies in the population, there is an added benefit of being able to include more cases in the study, which allows the study of rare diseases, disease subtypes or uncommon pre-clinical conditions as health endpoints. Due to large sample sizes, population-based studies are well suited to investigate more complex issues surrounding exposure and disease, such as the effect of confounding and interactions between multiple exposures, as well as the exploration of novel exposure-disease associations.

Naturally, population-based studies are not without disadvantages. One of the most significant limitations is the difficulty in assessing occupational exposure in population-based studies, as exposure may originate from a variety of workplaces and different time periods. If population-based studies are unable to accurately quantify disease risks associated with a unit of exposure, their utility and impact become limited in risk assessment and policymaking. The following chapters in this thesis aim to review, explore, and discuss potential means of improving exposure assessment and disease risk quantification in population-based studies.

Chapter 2 is a systematic literature review on occupational exposure assessment approaches used in population-based case-control studies on carcinogens in the previous four decades. The review explores the reliability performance of assessment approaches and highlights important areas of new development.

Chapter 3 demonstrates the performance of a real-time, algorithmic approach in assigning flexible, modular questionnaires during work history interviews in order to gather exposure information related to specific tasks. The manuscript covered the evaluation of a set of interview algorithms applied in 11,409 jobs episodes from subjects from the AsiaLymph project, which is a large, multi-centered hospital-based case-control study on lymphoma and leukemia in Asia.

Chapter 4 describes the procedure, training, and coordination employed to complete the occupational coding for 34,353 jobs from 12,590 subjects in the AsiaLymph project and assesses the reliability performance in job coding work performed by different coders. The reliability performance of exposure assessment resulting from the different sets of job codes is further explored for different exposures via a generic job-exposure matrix.

Chapter 5 estimates the association between occupational benzene exposure and risks of lymphohaematopoietic cancers in the Swiss National Cohort, which is a census-based cohort of 2.97 million Swiss citizens with 13,415 lymphohaematopoietic cancer cases. The study serves to demonstrate that registry-based cohorts in the general population can be useful for investigating the relationships between workplace exposures and human diseases, particularly for rare disease subtypes.

Chapter 6 describes the estimation of lung cancer risks associated with occupational exposure to crystalline silica in the SYNERGY study, which is a pooled analysis of 14 case-control studies from Europe and Canada with 16,901 lung cancer cases and 20,965 controls. In addition to overall lung cancer risks, risks by sex, silicosis status, smoking habit, and lung cancer subtype were investigated. The impacts of crystalline silica exposure under different exposure scenarios were expressed using lifetime excess lung cancer risks.

Chapter 7 details the development of a novel quantitative diesel engine exhaust job-exposure matrix and its application to assess associated risk of lung cancer in the SYNERGY study. Cancer risks by sex, smoking habit, and lung cancer subtypes were also explored. For risk assessment, lifetime excess lung cancer risks based on different workplace exposure scenarios were estimated.

Chapter 8 reviews this thesis' main findings and discusses potential avenues of improvement in occupational exposure assessment and epidemiological analysis in population-based studies.

REFERENCES

EEA. (2008). *Late lessons from early warnings: The precautionary principle 1896-2000*. EEA. https://www.eea.europa.eu/publications/environmental_issue_report_2001_22

EPHOR. (2020). *EPHOR Work packages*. EPHOR Project. https://www.ephor-project.eu/work-packages

Fang, S. C., Cassidy, A., & Christiani, D. C. (2010). A systematic review of occupational exposure to particulate matter and cardiovascular disease. *International Journal of Environmental Research and Public Health*, *7*(4), 1773–1806. https://doi.org/10.3390/ijerph7041773

Hall, A. L., Franche, R.-L., & Koehoorn, M. (2018). Examining Exposure Assessment in Shift Work Research: A Study on Depression Among Nurses. *Annals of Work Exposures and Health*, *62*(2), 182–194. https://doi.org/10.1093/annweh/wxx103

Kogevinas, M., Schlünssen, V., Mehlum, I. S., & Turner, M. C. (2020). The OMEGA-NET International Inventory of Occupational Cohorts. *Annals of Work Exposures and Health*, *64*(6), 565–568. https://doi.org/10.1093/annweh/wxaa039

Loomis, D., Guha, N., Hall, A. L., & Straif, K. (2018). Identifying occupational carcinogens: An update from the IARC Monographs. *Occupational and Environmental Medicine*, *75*(8), 593–603. https://doi.org/10.1136/oemed-2017-104944

Peters, S., Turner, M. C., Bugge, M. D., Vienneau, D., & Vermeulen, R. (2020). International Inventory of Occupational Exposure Information: OMEGA-NET. *Annals of Work Exposures and Health*, *64*(5), 465–467. https://doi.org/10.1093/annweh/wxaa021

Shan, Z., Li, Y., Zong, G., Guo, Y., Li, J., Manson, J. E., Hu, F. B., Willett, W. C., Schernhammer, E. S., & Bhupathiraju, S. N. (2018). Rotating night shift work and adherence to unhealthy lifestyle in predicting risk of type 2 diabetes: Results from two large US cohorts of female nurses. *BMJ (Clinical Research Ed.)*, *363*, k4641. https://doi.org/10.1136/bmj.k4641

Sieurin, J., Andel, R., Tillander, A., Valdes, E. G., Pedersen, N. L., & Wirdefeldt, K. (2018). Occupational stress and risk for Parkinson's disease: A nationwide cohort study. *Movement Disorders: Official Journal of the Movement Disorder Society*, *33*(9), 1456–1464. https://doi.org/10.1002/mds.27439

Stanaway, J. D., Afshin, A., Gakidou, E., Lim, S. S., Abate, D., Abate, K. H., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, M., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., … Murray, C. J. L. (2018). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1923–1994. https://doi.org/10.1016/S0140-6736(18)32225-6

Vermeulen, R., Schymanski, E. L., Barabási, A.-L., & Miller, G. W. (2020). The exposome and health: Where chemistry meets biology. *Science*, *367*(6476), 392–396. https://doi.org/10.1126/science.aay3164

CHAPTER 2

Use and Reliability of Exposure Assessment Methods in Occupational Case–Control Studies in the General Population: Past, Present, and Future

C Ge[1], M Friesen[2], H Kromhout[1], S Peters[1,3], N Rothman[2], Q Lan[2] and R Vermeulen[1,4]

1. Institute for Risk Assessment Sciences, Utrecht University, PO Box 80178, 3508 TD Utrecht, The Netherlands;
2. Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, MSC 9776, Bethesda, MD 20892, USA;
3. Departmentof Neurology, University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands;
4. Julius Center for Health Sciences and Primary

Published: Annals of Work Exposures and Health, 2018, Vol. 62, No. 9, 1047–1063.
doi: 10.1093/annweh/wxy080

ABSTRACT

Introduction: Retrospective occupational exposure assessment has been challenging in case–control studies in the general population. We aimed to review (i) trends of different assessment methods used in the last 40 years and (ii) evidence of reliability for various assessment methods.

Methods: Two separate literature reviews were conducted. We first reviewed all general population cancer case–control studies published from 1975 to 2016 to summarize the exposure assessment approach used. For the second review, we systematically reviewed evidence of reliability for all methods observed in the first review.

Results: Among the 299 studies included in the first review, the most frequently used assessment methods were self-report/assessment (*n* = 143 studies), case-by-case expert assessment (*n* = 139), and job-exposure matrices (JEMs; *n* = 82). Usage trends for these methods remained relatively stable throughout the last four decades. Other approaches, such as the application of algorithms linking questionnaire responses to expert-assigned exposure estimates and modelling of exposure with historical measurement data, appeared in 21 studies that were published after 2000. The second review retrieved 34 comparison studies examining methodological reliability. Overall, we observed slightly higher median kappa agreement between exposure estimates from different expert assessors (~0.6) than between expert estimates and exposure estimates from self-reports (~0.5) or JEMs (~0.4). However, reported reliability measures were highly variable for different methods and agents. Limited evidence also indicates newer methods, such as assessment using algorithms and measurement-calibrated quantitative JEMs, may be as reliable as traditional methods.

Conclusion: The majority of current research assesses exposures in the population with similar methods as studies did decades ago. Though there is evidence for the development of newer approaches, more concerted effort is needed to better adopt exposure assessment methods with more transparency, reliability, and efficiency.

INTRODUCTION

Retrospective exposure assessment in occupational case–control studies in the general population has been a major challenge (Kromhout et al., 1987; Teschke et al., 2002; Fritschi et al., 2003; Friesen et al., 2015a). For chronic diseases with a lengthy induction period, exposure has to be reconstructed for a subject's entire working lifetime. Accurate lifetime exposure assessment for any substance in the population is a difficult endeavour, as study subjects may have been employed in a large variety of occupations in different industries spanning different periods. Almost all retrospective occupational disease studies tackle this problem by first collecting detailed occupational histories from participants as a foundation for assessing work-related exposure.

The challenge then becomes estimating past exposures from full work histories. With the exception of a few widely studied and data-rich exposures such as crystalline silica, benzene, and asbestos, relevant historical exposure measurements in the population are often scarce, making fully quantitative assessment infeasible in most study settings (Stewart et al., 1996). As a result, qualitative and semi-quantitative assessment methods have been commonly used in population studies. The 'classical' qualitative/semi-quantitative assessment methods include the use of expert assessors to estimate exposure on a case-by-case basis, application of job-exposure matrices (JEMs), and reliance on self-reported exposure provided by study subjects or their next of kin. These methods may be used alone or in combination with each other to approximate lifetime exposures. For instance, studies may ask subjects to report previous exposures, then have expert assessors estimate exposures based on subject-reported exposures and job histories.

Teschke et al. (2002) published a comprehensive review on occupational exposure assessment in case–control studies. The review examined various exposure assessment techniques used at the time, and concluded based on reliability tests that case-by-case expert assessment generally have slightly better performance compared to other methods and 'is usually the best approach' for retrospective occupational exposure assessment in case–control studies (Teschke et al., 2002). The authors also proposed numerous suggestions to improve assessment reliability and efficiency, including the use of available exposure measurements to assist experts in assessing exposure, asking subjects about determinants of exposures rather than about exposures directly, and building measurement-based statistical models to predict exposure.

Since the report's publication 16 years ago, reliable and efficient exposure assessment in case–control studies in the general population has become even more important in the field of occupational epidemiology (Friesen et al., 2015a). As many hazardous exposures with large disease risks have been well characterized, recent efforts in occupational disease research aim to uncover new exposure–disease relationships with relatively small risks. In terms of statistical power, there is an advantage for large-scale population studies to detect small risk increases compared to industry-based studies. However, case-by-case expert exposure assessment often becomes cost- and time-prohibitive in these studies (Friesen et al., 2015a). There is a clear, growing need for more efficient and scalable assessment approaches, especially for large studies with multiple exposures of interest. There is also an increasing interest in uncovering specific shapes of exposure-response curves especially at the lower end of the exposure distributions, as well as characterizing gene–environment interactions. Discovery and quantification of these more nuanced relationships between exposure and effect require higher quality assessment to limit misclassifications. For instance, when working in dry-cleaning occupation was used as a proxy for perchloroethylene exposure, no significant association was found for liver cancer in a Nordic population (Lynge et al., 2006); however, a positive exposure–disease association was reported in the same population when exposure was assessed more quantitatively using a JEM (Vlaanderen et al., 2013).

In recent years, several methodological developments have allowed for improved assessment and quantification of historical work-related exposures in the general population. Collectively, these new developments may be described as 'enhancements' to classical methods. One example of such enhancement is the application of expert-derived algorithms linking questionnaire responses to expert and measurement-based exposure estimates (hereafter, algorithmic assessment). Another example is the use of historical exposure data to calibrate existing population JEMs to create quantitative exposure estimates (Friesen et al., 2015a).

Our work aimed to provide an updated overview of methods for retrospective occupational exposure assessment for case–control studies in the general population. The specific goals of our review are 3-fold. First, through a review of published cancer case–control studies, we show trends of use for various retrospective exposure assessment methods. Second, for these identified retrospective assessment methods, we systematically review evidence of reliability. Third, we discuss recent progress in retrospective assessment methods and consider future possibilities for further improving occupational exposure assessment in population case–control studies.

METHODS

To gather publications for exposure assessment method trends in occupational cancer case–control studies of chemical agents in the general population over the last four decades, we searched the Medline database with combinations of the following Medical Subject Heading (MeSH) terms: 'occupational exposure', 'case-control studies', and 'neoplasms'. We limited ourselves to the systematic review of cancer case–control studies as this covers a well-defined research area, and evaluating all population case–control studies for all diseases would be too unwieldy. A total of 1783 matches published between 1 January 1975 and 1 January 2017 were kept for further selection. After removal of studies that were duplicates, were not in English, did not focus on occupational exposures, used job title exclusively as an exposure proxy, were not case–control studies in the general population, or focused on non-chemical (e.g. radiation, noise) exposures, 299 publications remained (Prisma diagram available in Supplementary Figure 1). Use of various exposure assessment methods in occupational cancer studies were summarized by decade for trends of different assessment methods used in the last four decades.

To gather publications on reliability performance of different assessment methods published since the review performed by Teschke et al. (2002), combinations of MeSH terms ('occupational exposure', 'case-control studies', and 'reproducibility of results') were used in conjunction with title keywords ('validity', 'comparison', 'estimation', 'performance', 'agreement', 'reliability', 'validation', 'sensitivity', 'specificity', and 'assessment') to search for relevant articles published from 1 April 2001 to 1 January 2017. Parallel searches with truncation (e.g. valid*) were also performed to capture articles that used alternate forms of the keywords (e.g. validate). Seven hundred and twenty-six articles matched the search criteria. After removal of studies that were duplicates, were not in English, were not case–control studies in the general population, did not focus on chemical occupational exposures, or did not contain comparison tests of assessment methods, 34 articles remained (Prisma diagram available in Supplementary Figure 2).

RESULTS

Assessment method trends in occupational cancer studies in the general population

All but two (Bhatti et al., 2011; Lee et al., 2015) of the 299 identified general population case–control occupational cancer publications assessed exposure using at least one of the three classical assessment methods, namely case-by-case expert assessment, JEM, and self-reported exposure (full list of reviewed publications available in Supplementary Table 1). Most included studies (221 of 299) reported relying on a single method for retrospective exposure assessment. From these single-method studies, 89 relied on self-reported exposure, 82 used job-by-job expert assessment, 48 applied JEMs, and 2 modelled exposure using task-based information in conjunction with measurements (Bhatti et al., 2011; Lee et al., 2015). Seventy-eight studies used more than one method to assess past work-related exposures.

Figure 1 shows both the type of occupational information collected and the exposure assessment method used in these 299 studies by decade from 1980. Approximately 80–90% of all included studies collected full occupational histories throughout different decades. Use of job- or task-specific questionnaires (hereafter, specific questionnaires), which study subjects respond to optional specific job- or task-based questions on determinants of exposure, was observed in approximately 10% of reviewed studies published in the 1990s. The frequency of using specific questionnaires rose subsequently to around 25% of studies in the 2000s and 35% of studies from 2010 onward.

The proportion of studies that used self-reported exposures was approximately 45% in the 1980s, 55% in the 1990s and 2000s, and 35% in the current decade. These include studies that reported asking questions directly about specific exposures, providing a checklist of exposure substances, or having open-ended questions on exposure. Expert assessment on a case-by-case basis was used in approximately 40% of included studies from the first three decades and 55% of studies published in the current decade. The use of JEMs was reported in ~40% of studies published in the 1980s, 25% of studies from the next two decades, and 30% of studies published in or after 2010. Other methods, such as algorithmic assessment and measurement-calibrated JEMs, have appeared in 2% of studies in the 2000s and 20% in the 2010s.

Figure 1. Use of various retrospective occupational exposure assessment methods in general population case–control occupational cancer studies (*Others: includes methods that are distinct from other major assessment methods, such as exposure assessment using expert-derived algorithms, measurement calibrated job-exposure matrices, modelling of exposures based on historical measurements, and other learning or clustering statistical models).

Assessment method reliability and comparison studies

Of the 34 reliability studies identified, most (n = 30) compared exposure assessment results obtained from two methods; four (Daniels et al., 2001; Parks et al., 2004; Bourgkard et al., 2013; Friesen et al., 2013) compared assessment outcomes from three or more methods. All gathered studies compared candidate assessment methods against one or more assessment methods nominated as the comparison standard. For evaluating agreement between categorical measures of exposure, reliability studies often use the kappa statistic (κ), which may be interpreted as representing agreement that is almost perfect (κ = 0.81–1), substantial (κ = 0.61–0.8), moderate (κ = 0.41–0.6), fair (κ = 0.21–0.4), slight (κ = 0–0.2), and poor (κ <0) (Landis and Koch, 1977).

Table 1. Reliability of case-by-case expert assessment in estimating past occupational exposures in case–control studies in the population.

Authors, year	Exposure	Assessment method	Comparison method	Reliability test	Results
Daniels et al., 2001	Pesticides	Case-by-case expert review based on partial questionnaire data (e.g. job title, job tasks, industry, products and services)	Referent case-by-case expert assessment with full questionnaire data	Sensitivity and specificity against referent assessment; κ for presence of exposure	Sensitivity = 42.9–66.7%; specificity = 98.1–99.7%; κ = 0.5–0.6
Tinnerberg et al., 2001	13 agents	Case-by-case assessment performed individually by three occupational hygienists	Reassessment by two original hygienists after 1–3 years	κ for exposure status	κ between original and reassessment = 0.66. Inter-rater κ between different experts during reassessment = 0.72
Fritschi et al., 2003	19 agents	Consensus case-by-case assessment of exposure probability by three experts	Personal air measurements on select substances	Sensitivity for detecting substances present in air samples	Sensitivity = 73% for correct assessment with some certainty (probable/definite exposure)
Mannetje et al., 2003	70 agents	Case-by-case assessment by eight teams of experts in different study centres	Case-by-case assessment by reference chemist expert	Sensitivity, specificity versus reference rater; κ agreement for exposure presence, frequency, and intensity between all raters	Specificity >0.9; sensitivity = 0.48–0.75; overall κ across all agents = 0.41–0.45 between eight study centres and 0.53–0.64 between centres and reference rater
Tinnerberg et al., 2003	15 agents	Case-by-case assessment by occupational hygienist in five study centres	Comparison between different study centres	κ agreement for presence of exposure	Pair-wise comparison κ = 0.14–1.0 for the 15 agents, median = 0.74
Correa et al., 2006	Lead	Case-by-case assessment by three industrial hygienists independently	Comparison of between different experts	κ agreement	Inter-rater κ = 0.32–0.54 for presence/absence of exposure and for exposure probability, type, frequency, duration, and intensity
Richiardi et al., 2006	Diesel engine exhaust	Case-by-case assessment performed by three industrial hygienists independently	Comparison between different hygienists	Weighted κ for probability, intensity, and frequency of exposure	Weighted κ = 0.4–0.6

Table 1. *Continued*

Authors, year	Exposure	Assessment method	Comparison method	Reliability test	Results
Rocheleau et al., 2011	Seven agents	Case-by-case assessment by two industrial hygienists independently	Comparison between experts	κ agreement for presence of exposure	κ = 0.24–0.65 (median = 0.54) for different substances for the first 7229 jobs; after additional training κ = 0.51–0.91 (median = 0.51) for the remaining 4962 jobs
Gramond et al., 2012	Asbestos	Case-by-case assessment by six external experts individually and by consensus	Reference case-by-case assessment by two internal experts by consensus; inter-rater comparison between the six external experts	κ for exposure probability and cumulative exposure	Inter-rater weighted κ between external experts = 0.69–0.81; weighted κ against referent assessment = 0.79–0.84
Bourgkard et al., 2013	Asbestos and PAHs	Case-by-case assessment by two experts by consensus based on TBQ data	Reference case-by-case assessment by two different experts by consensus based on full interview data; population-based asbestos Matgéné JEM (Févotte et al., 2011)	Weighted κ for ordinal exposure levels	Weighted κ between TQB expert assessment and reference assessment was 0.68 for asbestos and 0.43 for PAHs; weighted κ between TBQ expert assessment and asbestos JEM was 0.31
Friesen et al., 2013	Diesel engine exhaust	Case-by-case assessment by three hygienists individually	Aggregate case-by-case assessment by three different experts.	Weighted κ for exposure probability, intensity, and frequency	Weighted κ = 0.50–0.76 (median = 0.59)
DellaValle et al., 2015	PCB	Case-by-case assessment of exposure probability by an industrial hygienist	Concentrations of 14 PCB congeners in serum	Variance in serum PCB explained by hygienist rating; regression model to compare serum PCB levels for subjects with different exposure ratings	38% of the variability in total serum PCB explained by hygienist rating; total serum PCB is 87% higher in workers rated probably exposed versus unexposed (no difference between those rated non-exposed and possibly exposed)

PCB = polychlorinated biphenyl; TBD = task-based questionnaire.

Expert case-by-case assessment was the most frequently included method in gathered studies, appearing in 12 studies as the candidate method. Three studies compared expert-assessed exposures to measured exposure (Fritschi et al., 2003; DellaValle et al., 2015) and to JEM-assessed exposure (Peters et al., 2011a); 10 other studies compared assessments made by the same experts at different times (Tinnerberg et al., 2001) or assessments made by different experts (Daniels et al., 2001; Tinnerberg et al., 2001; Fritschi et al., 2003; Mannetje et al., 2003; Tinnerberg et al., 2003; Correa et al., 2006; Rocheleau et al., 2011; Gramond et al., 2012; Friesen et al., 2013; Table 1). Fritschi et al. (2003) reported an average sensitivity of 73% for three experts who assessed exposure to 19 different agents for 47 fictional jobs constructed from personal air monitoring records. Another study on polychlorinated biphenyl (PCB) exposure reported that total serum PCB levels were 87% higher in subjects rated as exposed versus unexposed by an expert, with 38% of variability in serum PCB levels explained by the expert rating (DellaValle et al., 2015). Reported $\varkappa$ agreement for presence/ absence of exposure between different expert assessors ranged between –0.04 and 1, with a median of approximately 0.58 (Daniels et al., 2001; Tinnerberg et al., 2001, 2003; Mannetje et al., 2003; Correa et al., 2006; Rocheleau et al., 2011; Gramond et al., 2012; Friesen et al., 2013). Median intra-rater $\varkappa$ was 0.66 for assessments made at least 1 year apart for 13 different exposures by the same experts (Tinnerberg et al., 2001).

Nine comparison studies examined the reliability of self-reported exposures by comparison with expert assessment (Daniels et al., 2001; Parks et al., 2004; Nam et al., 2005; Westberg et al., 2005; Hepworth et al., 2006; Neilson et al., 2007), JEM-assessed exposure (Adegoke et al., 2004; Hardt et al., 2014), and repeated surveys of the same subjects (Duell et al., 2001; Table 2). The range of reported $\varkappa$ agreement between self-reported and expert-assessed presence of exposures was 0.19–0.70, with a median of approximately 0.50 (Daniels et al., 2001; Nam et al., 2005; Westberg et al., 2005; Hepworth et al., 2006; Neilson et al., 2007). Parks et al. (2004) reported sensitivity and specificity values of 0.54 and 0.99, respectively, for self-reported exposure to crystalline silica versus assessment made by experts. Hardt et al. (2014) reported poor agreement between self-reported and JEM-assessed exposures to asbestos, with $\varkappa$ values ranging from 0.06 to 0.30, with a median of 0.19. Adegoke et al. (2004), however, reported better agreement ($\varkappa$ range 0.48–0.84, median 0.78) between self-reported and JEM-assessed exposures for benzene, organic solvents, pesticides, and electromagnetic fields. Duell et al. (2001) reported a $\varkappa$ range from 0.63 to 0.84 (median 0.75) for interview responses made 14 months apart by the same study subjects on exposure and use of pesticides.

Eight reliability studies compared exposures obtained from applying JEMs to exposures assessed by expert raters (Daniels et al., 2001; Parks et al., 2004; Semple et al., 2004; Nam et al., 2005; Peters et al., 2011a; Offermans et al., 2012) or other JEMs (Lavoué et al., 2012; Offermans et al., 2012; Table 3). Offermans et al. (2012) compared exposure to asbestos, polycyclic aromatic hydrocarbons (PAHs), and welding fumes determined with three different population JEMs with each other and with case-by-case assessment by experts. Weighted $\varkappa$ agreement between JEM and expert-assessed cumulative exposures ranged from 0.10 for asbestos to 0.70 for welding fumes, with a median of ~0.36. Weighted $\varkappa$ agreement between JEMs ranged from 0.25 to 0.51, with a median of ~0.46. In a multi-centre European lung cancer study, Peters et al. (2011a) reported $\varkappa$ agreement ranging from 0.04 to 0.54 (median~0.38) between a population JEM and case-by-case expert assessment for presence of exposure to asbestos, diesel engine exhaust, and crystalline silica in eight different countries. In another comparison between different population JEMs, Lavoué et al. (2012) reported weighted $\varkappa$ ranging from 0.07 to 0.89 (median: 0.39) for exposure prevalence of 27 different agents. Using case-by-case expert assessment as the standard, Parks et al. (2004) reported a sensitivity of 0.44 and specificity of 0.97 for a general population JEM in assessing exposure to crystalline silica.

Ten reviewed studies tested the reliability of exposures estimated by other methods, such as the use of expert-derived algorithms (Pronk et al., 2012; Bourgkard et al., 2013; Friesen et al., 2013, 2014; Peters et al., 2014) or learning/clustering models that predict exposure based on questionnaire responses (Black et al., 2004; Friesen et al., 2015b; Wheeler et al., 2013, 2015; Friesen et al., 2016b; Table 4). Weighted $\varkappa$ values reported by studies comparing exposure probabilities estimated with algorithms versus expert raters ranged between 0.49 and 0.82 in three different studies, with a median of 0.81 (Pronk et al., 2012; Bourgkard et al., 2013; Friesen et al., 2013). Another study reported a median $\varkappa$ agreement of 0.73 in dichotomous measures of exposure between algorithmic and expert assessment (Peters et al., 2014). Performance of tree-based statistical learning models to predict diesel engine exhaust exposure ratings based on expert assessment from patterns in questionnaire responses was reported in two studies (Wheeler et al., 2013, 2015; Friesen et al., 2016b). When tested against validation subsets, tree-based assessment models created by Wheeler et al. (2013, 2015) had 92–94% agreement with experts in identifying exposed versus non-exposed jobs. When applied in a Spanish bladder cancer study, the same tree-based models predicted expert-assessed exposure probability, intensity, and frequency correctly in 90%, 91%, and 57% of 1442 jobs, respectively (Friesen et al., 2016b).

Table 2. Reliability of self-reported exposures in estimating past occupational exposures in case–control studies in the population.

Authors, year	Exposure	Assessment method	Comparison method	Reliability test	Results
Daniels et al., 2001	Pesticides	Self-reported exposure via telephone interview	Case-by-case expert assessment	κ for presence of exposure; sensitivity and specificity	κ = 0.3–0.7; sensitivity = 100%; specificity = 96.2–97.3%
Duell et al., 2001	Pesticides	Self-reported exposure in telephone interview	Self-reported exposure in re-interview after 14 months	κ for ever/never pesticide application	κ = 0.63–0.78, median = 0.75
Adegoke et al., 2004	Benzene, EMF, pesticides, and other organic solvents	Self-reported exposure by subjects or next of kin during in-person interview	Population JEM developed by authors	Percent agreement, sensitivity, specificity, and κ for presence of exposure	Percent agreement = 91.6–98.5 (median 94.1); sensitivity = 0.83–0.97 (median 0.91); specificity = 0.90–0.99 (median 0.95), κ = 0.48–0.84 (median 0.78)
Parks et al., 2004	Silica	Self-reported exposure with checklist during in-person interview	Case-by-case expert assessment based on questionnaire data plus follow-up telephone interview data	Sensitivity and specificity	Sensitivity = 0.54 for long-term exposures (>12 months) and 0.73 for shorter-term exposures (>2 weeks);specificity=0.99 for all exposures
Westberg et al., 2005	PVC	Self-reported exposure in paper-based questionnaire	Case-by-case assessment by two experts by consensus	κ for presence of exposure; odds ratios for cancer	κ = 0.56; odds ratio for cancer was 1.1 (95% CI 0.8–1.6) based on self-reported exposure and 1.3 (95% CI 1.1–1.7) based on expert-assessed exposure
Nam et al., 2005	Asbestos	Next-of-kin-reported exposure	Case-by-case assessment by an occupational hygienist	κ for presence of exposure; odds ratio for cancer	κ = 0.47 for cases and 0.19 for controls; odds ratios for mesothelioma was 10.7 (95% CI 7.3–16.0) based on self-reported exposure and 4.7 (95% CI 3.2–6.8) based on expert-assessed exposure
Hepworth et al., 2006	Pesticides and solvents	Self-reported exposure in computer-assisted interview	Assessment by two experts for presence/absence of exposure based on job title alone	κ for presence of exposure; sensitivity and specificity	κ = 0.22 for solvents and 0.50 for pesticides; sensitivity = 45.8–53.6%; specificity = 90.3–99.3%
Neilson et al., 2007	PAHs	Self-reported exposure during longest held job	Case-by-case expert assessment	κ for presence of exposure; sensitivity and specificity	κ = 0.54; sensitivity and specificity were both 0.79
Hardt et al., 2014	Asbestos	Self-reported asbestos exposure	DOM JEM (Peters et al., 2011a)	κ for presence of exposure; odds ratio for lung cancer	κ = 0.19; odds ratio was 0.9 (95% CI 0.5–1.6) based on self-reported exposure and 1.9 (95% CI 1.3–2.7) based on JEM-assessed exposure

CI = confidence interval; DOM JEM = Domtoren job-exposure matrix; EMF = electromagnetic field; PVC = polyvinyl chloride.

Table 3. Reliability of JEMs in estimating past occupational exposures in case–control studies in the population.

Authors, year	Exposure	Assessment method	Comparison method	Reliability test	Results
Daniels et al., 2001	Pesticides	Occupation-industry JEM developed by authors	Referent case-by-case expert assessment	κ for presence of exposure; sensitivity and specificity	κ = 0.4–0.6; sensitivity = 57.1–71.4%; specificity = 97.7–99.1%
Parks et al., 2004	Silica	JEM developed by authors	Case-by-case expert assessment based on questionnaire data plus follow-up telephone interview data	Sensitivity and specificity	Sensitivity = 0.44 for long-term exposures (>12 months) and 0.32 for shorter-term exposures (>2 weeks); specificity = 0.97 for all exposures
Semple et al., 2004	Solvents, pesticides, and metals	JEM created by authors, plus exposure modifiers based on questionnaire responses	Case-by-case assessment by experts	Spearman's ρ for cumulative exposure	Spearman's ρ = 0.89 for a validation sample of 30 jobs
Nam et al., 2005	Asbestos	Assessment by population JEM (Sieber et al., 1991)	Case-by-case assessment by an occupational hygienist	κ for presence of exposure; odds ratio for cancer	κ = 0.24 for cases and 0.34 for controls. Odds ratios for mesothelioma was 2.1 (95%CI 1.5–2.9) based JEM-assessed exposure and 4.7 (95% CI 3.2–6.8) based on expert-assessed exposure
Orsi et al., 2010	Solvents	Matgéné JEM (Févotte et al., 2011)	Case-by-case assessment by a chemical engineer	Percent agreement and κ for presence of exposure	Percent agreement = 73–87 (median 82); κ = 0.46– 0.54 (median 0.50)
Peters et al., 2011a	Diesel engine exhaust, crystalline silica, asbestos	Assessment by population-specific JEM developed by authors; population-based DOM JEM	Case-by-case assessment performed by experts in eight research centres	κ for presence of exposure between all methods	κ between population-specific JEM and expert assessment = 0.28–0.91 (median = 0.63); κ between DOM JEM and expert assessment = 0.04–0.54 (median = 0.38); κ between two JEMs = 0.07–0.73 (median = 0.34)
Offermans et al., 2012	Asbestos, PAHs, welding fumes	Dutch Asbestos JEM, DOM JEM, FINJEM	Case-by-case expert assessment by consensus by two experts	Weighted κ on tertiles of cumulative exposure	κ = 0.29 for asbestos and 0.42 for PAHs for DOM JEM; κ = 0.70 for welding fume for FINJEM; κ = 0.10 for asbestos for asbestos JEM.
Lavoué et al., 2012	27 agents	FINJEM-assessed exposure prevalence and intensity	Exposure likelihood, frequency, and intensity assessed by Montreal JEM, developed by authors	Weighted κ for exposure prevalence; Spearman correlation for exposure intensity	Weighted κ = 0.07–0.89; Spearman correlation = −0.35 to 0.89

CI = confidence interval; DOM JEM = Domtoren job-exposure matrix; FINJEM = Finnish Information System on Occupational Exposure.

Table 4. Reliability of other assessment methods in estimating past occupational exposures in case–control studies in the population.

Authors, year	Exposure	Assessment method	Comparison method	Reliability test	Results
Pronk et al., 2012	Diesel engine exhaust	Use of expert-derived algorithms to assess exposure probability, intensity, and frequency based on occupational histories with specific task information	Case-by-case assessment by an occupational hygienist	Weighted κ for ordinal exposure measures; Spearman correlation for continuous exposure measures	Weighted κ = 0.68–0.81 for ordinal exposure probability, frequency, and intensity; Spearman ρ = 0.70–0.72 for continuous exposure frequency and intensity
Bourgkard et al., 2013	Asbestos and PAHs	Algorithmic assessment based on task-based questionnaire data	Reference case-by-case assessment by two experts by consensus based on full interview data; population-based asbestos JEM (Févotte et al., 2011)	Weighted κ for ordinal exposure levels; OR for lung cancer and asbestos exposure	κ = 0.61 for asbestos and 0.36 for PAHs against referent expert assessment; κ = 0.26 against asbestos JEM; lung cancer OR = 1.18 (95% CI 1.06–1.31) based on algorithm derived exposures and 1.02 (95% CI 0.91–1.16) based on JEM-assessed exposures
Friesen et al., 2013	Diesel engine exhaust	Algorithm-based assessment (Pronk et al., 2012) to assess exposure probability, intensity, and frequency based on questionnaire responses	Case-by-case assessment by three experts individually and by aggregate	Weighted κ for exposure probability, intensity, and frequency	κ = 0.58–0.81 (median = 0.70) between individual expert rating and algorithmic assessment; κ = 0.82 for aggregated expert assessment versus algorithmic assessment
Wheeler et al., 2013	Diesel engine exhaust	Use of tree-based statistical learning models to predict exposure probability, frequency, and intensity using previous expert assessments as training data	Case-by-case assessment by an occupational hygienist	Percent agreement for presence of exposure, and ordinal exposure probability, frequency, and intensity	Percent agreement = 92–94 for presence of exposure; percent agreement = 7–90 for ordinal exposure probability, frequency, and intensity
Peters et al., 2014	Diesel engine exhaust, pesticides, and solvents	Expert-derived algorithms were used to assess presence/absence of exposure from information obtained from questionnaires	Case-by-case assessment by an occupational hygienist	κ agreement on presence of exposure	κ = 0.51–0.84 (median 0.73)
Friesen et al., 2014	TCE	A systematic process was developed to extract free-text responses in occupational histories by identifying keywords and phrases associated with exposure	Case-by-case expert assessment	Percent agreement on presence of exposure	Percent agreement = 98.7
Friesen et al., 2015b	Diesel engine exhaust	Hierarchical clustering model grouped jobs with similar exposures based on questionnaire responses	Algorithmic assessment of exposure probability, intensity, and frequency (Pronk et al., 2012)	ICCs within job title clusters	ICC > 80% for exposure probability with >500 clusters w in model; ICC > 70% for exposure frequency and intensity with >200 model clusters
Wheeler et al., 2015	Diesel engine exhaust	Use of ordinal and nominal classification tree models to predict exposure probability, frequency, and intensity using expert assessment information	Case-by-case assessment by an occupational hygienist	Somer's *d* for nominal and ordinal exposure metrics (probability, frequency, and intensity)	Somer's *d* = 0.61–0.66
Friesen et al., 2016b	Diesel engine exhaust	Application of classification tree models (Wheeler et al., 2013)	Case-by-case assessment by two experts independently	Weighted κ for ordinal measures of exposure probability, intensity, and frequency	Weighted κ = 0.09–0.91; model performance was better for unexposed and highly exposed jobs, and for predicting exposure probability and intensity

CI = confidence interval; ICC = intraclass correlation coefficient; OR = odds ratio; TCE = trichloroethylene.

DISCUSSION

We surveyed general population occupational cancer case–control studies published over the last four decades to examine the trends of use for various assessment methods. Case-by-case expert assessment, population JEMs, and self-reported exposure were by far the most frequently used assessment methods in all periods reviewed. Notable trends were also observed in the increasing use of specific questionnaires starting in the 1990s, and the use of exposure algorithms and models starting in the 2000s. We have focused on cancer studies for investigating exposure assessment method trends because it is an active area of chronic occupational diseases research.

In the absence of true gold standards, case-by-case expert assessment is often regarded as the 'alloyed gold standard' and 'best practice' for retrospective occupational exposure assessment (Siemiatycki et al., 1989; Bouyer and Hémon, 1993a, 1993b; Fritschi et al., 1996; Teschke et al., 2002). In the 34 assessment method reliability publications we reviewed, authors in 27 studies selected expert assessment as the standard method of comparison. Given the same work history information, assessment experts, who may be industrial hygienists, chemists, engineers, occupational physicians, or experienced workers, are believed to have better knowledge on occupational exposures than workers and be able to produce individualized exposure estimates. If expert-assessed presence of exposure was used as a comparison standard, our results show slightly higher median κ agreement between different expert assessors (~0.6) than between experts and estimates from self-reports (~0.5) or JEMs (~0.4). However, it is important to note that reliability studies reviewed reported highly variable agreement results for different exposures assessed using various methods (reported unweighted κ values by substance available in Supplementary Figures 3–8). Assessment reliability may be impacted by a number of factors, including the type and number of exposures, study design, quality of occupational history information, number and experience of expert assessors, as well as the comparison standard (i.e. other experts or JEMs). No single assessment approach is likely to outperform others in all study settings. For instance, a European multicentre case–control study compared exposures assigned by eight teams of expert raters and observed κ agreement ranging from −0.04 for PAHs to 0.93 for arc welding fumes (Mannetje et al., 2003). Further, agreement with a selected comparison standard does not necessarily mean higher validity. For example, in a lung cancer study by Peters et al. (2011a), no significant relationship between occupational asbestos exposure and lung cancer was found when exposure was assessed by expert assessors across eight European study centres. However, when asbestos exposure was assessed using a general population JEM, a significant exposure–disease relationship was found among the same study subjects. Use of different local expert assessors in the study resulted in higher inter-rater differences in exposure estimates for asbestos, which likely reduced study power and diminished the observed risk between exposure and disease.

There are additional important limitations with assessing exposures case-by-case with experts. Expert review is labour and resource intensive. The number of expert decisions for assessment increases multiplicatively with the number of study subjects, jobs per subject, exposure agents, exposure metrics, and multiple experts. Authors of a prostate cancer study estimated over 1000 h of expert time was used to assess exposure to six groups of exposure agents for over 13 000 jobs reported by 1479 study subjects (Fritschi et al., 2009, 2007). Efficiencies are higher for experienced exposure assessment experts, who are likely to have developed intrinsic decision rules in order to rate exposures consistently and quickly, but these rules are often not explicitly described, resulting in the frequent criticism of the expert decision process as a 'black box' that lacks transparency (Teschke et al., 2002; Pronk et al., 2012; Peters et al., 2014; Wheeler et al., 2013). Although there is evidence that hidden decision rules within experts may be calibrated by training and measurement data to assess exposures with better agreement (Mannetje et al., 2003; Rocheleau et al., 2011), the lack of explicit documentation leads to challenges in results comparison across different experts, critical evaluation, and reproduction within or across studies.

Use of algorithms and tree-based statistical models to assess exposure represent a direct effort to standardize and address some shortcomings of case-by-case expert assessment while maintaining capability to assess exposure on an individual level. In algorithmic assessment, decision rules are explicitly described, allowing for review, revision, and adaptation in other studies. Initial evidence in comparison studies suggests good reliability for algorithmic assessment when compared to case-by-case expert assessment. In a reliability comparison by Friesen et al. (2013), weighted κ agreement for estimated diesel exhaust exposure between exposure algorithms and individual expert assessors (κ range: 0.58–0.81) was similar to agreement observed between different expert assessors (κ range: 0.50–0.76).

Expert decisions in case-by-case assessment have also been used in a few recent studies to train statistical models to either cluster reported jobs with similar exposures (Friesen et al., 2015b) or directly predict exposure probability, intensity, and frequency (Wheeler et al., 2013, 2015; Friesen et al., 2016b). In general, these models seemed excellent in identifying non-exposed and highly exposed jobs, but had lower performance in predicting jobs with low or medium exposures (Wheeler et al., 2013; Friesen et al., 2015b). Therefore a tiered approach involving initial application of statistical models to first identify difficult-to-assess jobs, followed by expert review of highlighted jobs, has been proposed to increase assessment transparency and reduce expert burden. Friesen et al (2016b) recently applied this hybrid approach in a population-based bladder cancer study in Spain by combining model-derived assessment algorithms from a US study with job-by-job expert assessment. The algorithms showed high agreement with experts in assessing exposure for jobs that were non-exposed and identified only 14% of jobs requiring further expert review, demonstrating good reliability, reproducibility, and efficiency may be achieved with hybrid approaches.

Asking subjects to report past occupational exposures is a direct and convenient method of collecting exposure information at an individual level. Although expertise in exposure assessment is unlikely, workers may have important insight on their work environment, tasks performed, equipment used, materials handled, as well as the intensity and frequency of contact with different exposure agents. The most concerning limitation of self-assessed exposures is the potential for differential recall bias in case–control studies. Increased likelihood for cases to report past exposures and to report higher exposures results in inflated risk estimates (Tielemans et al., 1999; Teschke et al., 2000; de Vocht et al., 2005). At the same time, workers may also under-report exposure to agents that are invisible, cannot be felt, or when their exposure is indirect (bystander exposure), which also diminishes the observed relationship between exposure and disease (Kromhout et al., 1987; Teschke et al., 2002). For instance, Hardt et al (2014) found a significant relationship between lung cancer and asbestos exposure when exposure was assessed using a generic JEM, but not when exposure was assessed using subject self-reports. Authors of the study reported that cases were more likely than controls to either over- or under-report asbestos exposure, leading to misclassifications that affect the observation of the true exposure–disease relationship (Hardt et al., 2014).

Compared to directly asking subjects to report exposures, specific questionnaires are less prone to recall bias, as subjects are typically better able to accurately report work tasks and other exposure determinants versus exposures to specific agents (Teschke et al., 2002). The relationship between exposure determinants and disease is also more indirect and thus less likely for study subjects to uncover. Work task information obtained from specific questionnaires may be used to identify within-job differences in exposure, and work environment information may be used in determining bystander exposures. Detailed exposure determinants information from specific questionnaires may be utilized in various ways in subsequent exposure assessment work, such as helping expert assessors develop more confident and accurate estimates (Segnan et al., 1996; Tielemans et al., 1999; Lillienberg et al., 2008), or supporting the development of exposure assessment algorithms (Wheeler et al., 2013). There are, however, challenges in implementing specific questionnaires. Because specific questionnaires are typically triggered by reports of key occupations and tasks, it is important for the trigger keywords to be adequately sensitive and specific to identify potential exposure scenarios. If specific questionnaires are administered by interviewers, they must be trained to correctly and immediately apply different modules based on key jobs or tasks reported by subjects (Colt et al., 2011; Carey et al., 2014). If an automated system is responsible for administering specific questionnaires, the list of keywords must be sensitive and specific to avoid asking irrelevant modules and missing potential exposures (Friesen et al., 2016a). Finally, although use of specific questionnaires reduces interview burden for subjects who report few or no relevant exposure-related tasks, burden for those reporting multiple relevant tasks increases. Therefore it is important to identify and include only key task and environmental determinants that predict exposures and exclude less predictive determinants as much as possible in order to limit interview burden on exposed subjects.

General population JEMs were first developed in the 1980s mainly for assessing exposure to carcinogens (Hoar et al., 1980; Pannett et al., 1985). Since then, a number of general population JEMs have been developed for various exposures in different countries (Kromhout and Vermeulen, 2001). Dimensions of a JEM typically include occupation/industry and at least one measure of exposure (e.g. intensity, probability), but may include additional axes such as region or period. Once a JEM is developed, application is straightforward, virtually cost-free, and generates immediate assessment results for multiple exposure agents. Although expert judgement is involved in the creation of JEMs, assessment rating for each cross-tabulation JEM cell is available, which allows for comparison, review, and amendment of individual decisions in the matrix. One limitation for JEMs is their overall lower sensitivity compared to other assessment methods (Bouyer and Hémon, 1993b; Teschke et al., 2002; Parks et al., 2004). This lower sensitivity is necessary by design for generic JEMs, where exposures are assigned broadly by job, yet overall specificity must be maximized to limit exposure misclassification in the largely unexposed general population. In fact, many studies further reduce JEM sensitivity for higher specificity by dichotomizing ordinal or semi-quantitative exposure metrics when assessing exposure (Peters et al., 2011a). Though such trade-off between sensitivity and specificity usually increases the positive predictive ability of a JEM, further reduction in sensitivity must be remedied by increasing sample size, adding cost and challenges for the study. Another criticism of JEMs is their inability to account for between-worker variability for subjects with the same job title, missing important determinants of exposure such as differences in tasks, material use, work environments, and period. Because a JEM typically categorizes subjects and assigns exposures using a set of standardized job codes, performance of the JEM is dependent on the ability of the job codes in separating the population in terms of exposure contrast. However, standardized job classification systems, such as the International Standard Classification of Occupations (ISCO) from the International Labour Organisation (ILO), were not designed for use in JEMs, and certain job categorizations that make sense in economic or demographic terms may perform poorly in occupational exposure assessment. For instance, both underground and surface miners are coded as one job in ISCO—a major problem as exposures in confined spaces underground are often much higher. In general, a coding system with more granularity is better in separating jobs with different exposures. However, recent updates to many job classifications systems have decreased the level of detail by consolidating jobs. As an example, the ISCO version from 1968 contains 1506 distinct jobs, whereas versions from 1988 and 2008 contain 390 and 425 jobs, respectively (International Labour Organization (ILO), 2010). The assignment of exposure by job group rather than individuals in JEMs introduces Berkson-type error in exposure assessment, which is a non-differential misclassification leading to a reduction in power and wider confidence intervals around unbiased risk estimates (Armstrong, 1990; Heid et al., 2004).

In the last 5 years, studies have used historical exposure data to calibrate generic JEMs to improve their performance. For instance, existing population JEMs for benzene and lead exposure were calibrated using mixed-effect models based on exposure measurements to produce quantitative estimates for a cohort of women in Shanghai, China (Friesen et al., 2012; Koh et al., 2014). Period and original JEM ordinal exposure rating were included as fixed effects, and occupation and industry were incorporated as random effects in the model. Using similar modelling techniques, Peters et al. (2016) created a measurement-calibrated lung carcinogen JEM (SYNJEM) with 102 306 personal samples for asbestos, chromium, nickel, PAHs, and crystalline silica for use in different European regions and Canada. For SYNJEM, both modelled exposure intensity levels and cumulative exposures were consistent with reported values by other studies (Peters et al., 2011b, 2016), and sensitivity analyses with different model assumptions showed that the quantitative estimates were robust (Peters et al., 2013). Compared to traditional, semi-quantitative expert-derived JEMs, exposure data-calibrated JEMs offer fully quantitative exposure estimates that may be adjusted for period and geographical region (Peters et al., 2016; Olsson et al., 2017). The main challenge of this particular approach is that extensive monitoring data are only available for few exposure agents, so it is not feasible for less-monitored exposures

or in regions lacking substantial existing exposure monitoring data. Even when such data exist, there is considerable costs and difficulty in obtaining and digitizing large quantities of historical exposure data. In addition, within-job variations in exposure are still difficult to assess with data-calibrated JEMs, because historical measurements seldom have detailed descriptions of work tasks, environmental conditions, and other exposure determinants. Finally, a number of non-occupational factors may introduce bias and variance in historical measurement data, such as sampling and analytic methods, reason of data collection (e.g. routine monitoring versus complaint investigation), sampling duration, and sampling strategy (e.g. representative versus worst case). Documentation of these important variables in historical datasets is typically incomplete, which makes model adjustments and results interpretation more challenging.

Although we have so far analysed and discussed various retrospective occupational exposure assessment methods as distinct approaches, they are deeply interdependent and share many similarities. Expert judgement is central in compilation of occupational history and task-based questionnaires, case-by-case expert assessment, development of JEMs, and generation of exposure assessment algorithms. Subject-reported information, including reported job histories, tasks, and exposures, informs case-by-case expert assessment and JEM application. From a broad perspective, the central challenges in retrospective occupational assessment are the following:

1. How to obtain reliable subject job history and exposure determinant information?
2. How to apply expert judgement to subject-reported information systematically, transparently, and effectively to produce exposure estimates that are reliable and reproducible?
3. If available, how may exposure measurements be used to further improve assessment accuracy and reliability?

In the field of occupational disease epidemiology, overall progress to meet these important challenges has been slow. The majority of current research in the field assesses exposure in the population with similar methods as studies did 30 years ago. Although the use of specific questionnaires began more than two decades ago, they were mostly designed and used to support case-by-case expert assessment. In the last 10 years, there are clear efforts in standardizing different elements of exposure assessment and in increasing overall transparency and reproducibility. However, strong reliance on case-by-case expert assessment as the 'alloyed gold standard' for assessing past work exposures results in many study components being designed around job-by-job expert review, making it difficult to apply and test alternative methods. For instance, in two recent studies decision rules for algorithmic assessment and key variables for tree-based statistical models had to be extracted either manually or programmatically from free-text questionnaire responses intended for use by expert assessors (Pronk et al., 2012; Friesen et al., 2014).

Bottom-up study designs tailored for alternative assessments will generate a more positive environment for implementation of systematic assessment approaches. The use of specific questionnaires in general population case–control studies should ideally be a standard practice similar to the collection of full occupational histories. Responses to specific questionnaires may then be a foundation for the application of exposure algorithms, learning models, or, if necessary or desired, any classical assessment methods. Reliance on expert judgement remains in this new paradigm, as identifying key exposure determinants to include in specific questionnaires and exposure algorithms must involve exposure assessment experts. However, details of expert decisions will be more transparent and their application will be systematic. Early evidence already indicates that some alternative methods perform at least as well as the classical assessment methods, or could serve as their compliment for more efficient assessment. The incorporation of historical measurement data to support current assessment efforts should also be encouraged whenever data is available. A more concerted effort in further improvement of these new approaches may enable the creation of assessment methods (or hybrid methods) that are as efficient and transparent as JEMs, while as sensitive and precise as case-by-case expert assessment.

REFERENCES

Adegoke, O. J., Blair, A., Ou Shu, X., Sanderson, M., Addy, C. L., Dosemeci, M., & Zheng, W. (2004). Agreement of job-exposure matrix (JEM) assessed exposure and self-reported exposure among adult leukemia patients and controls in Shanghai. *American Journal of Industrial Medicine*, *45*(3), 281–288. https://doi.org/10.1002/ajim.10351

Armstrong, B. G. (1990). THE EFFECTS OF MEASUREMENT ERRORS ON RELATWE RISK REGRESSIONS. *American Journal of Epidemiology*, *132*(6), 1176–1184. https://doi.org/10.1093/oxfordjournals.aje.a115761

Bhatti, P., Newcomer, L., Onstad, L., Teschke, K., Camp, J., Morgan, M., & Vaughan, T. L. (2011). Wood dust exposure and risk of lung cancer. *Occupational and Environmental Medicine*, *68*(8), 599–604. https://doi.org/10.1136/oem.2010.060004

Black, J., Benke, G., Smith, K., & Fritschi, L. (2004). Artificial neural networks and job-specific modules to assess occupational exposure. *The Annals of Occupational Hygiene*, *48*(7), 595–600. https://doi.org/10.1093/annhyg/meh064

Bourgkard, E., Wild, P., Gonzalez, M., Févotte, J., Penven, E., & Paris, C. (2013). Comparison of exposure assessment methods in a lung cancer case-control study: Performance of a lifelong task-based questionnaire for asbestos and PAHs. *Occupational and Environmental Medicine*, *70*(12), 884–891. https://doi.org/10.1136/oemed-2013-101467

Bouyer, J., & Hémon, D. (1993a). Retrospective evaluation of occupational exposures in population-based case-control studies: General overview with special attention to job exposure matrices. *International Journal of Epidemiology*, *22 Suppl 2*, S57-64.

Bouyer, J., & Hémon, D. (1993b). Studying the performance of a job exposure matrix. *International Journal of Epidemiology*, *22 Suppl 2*, S65-71.

Carey, R. N., Driscoll, T. R., Peters, S., Glass, D. C., Reid, A., Benke, G., & Fritschi, L. (2014). Estimated prevalence of exposure to occupational carcinogens in Australia (2011–2012). *Occupational and Environmental Medicine*, *71*(1), 55–62. https://doi.org/10.1136/oemed-2013-101651

Colt, J. S., Karagas, M. R., Schwenn, M., Baris, D., Johnson, A., Stewart, P., Verrill, C., Moore, L. E., Lubin, J., Ward, M. H., Samanic, C., Rothman, N., Cantor, K. P., Beane Freeman, L. E., Schned, A., Cherala, S., & Silverman, D. T. (2011). Occupation and bladder cancer in a population-based case-control study in Northern New England. *Occupational and Environmental Medicine*, *68*(4), 239–249. https://doi.org/10.1136/oem.2009.052571

Correa, A., Min, Y.-I., Stewart, P. A., Lees, P. S. J., Breysse, P., Dosemeci, M., & Jackson, L. W. (2006). Inter-rater agreement of assessed prenatal maternal occupational exposures to lead. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *76*(11), 811–824. https://doi.org/10.1002/bdra.20311

Daniels, J. L., Olshan, A. F., Teschke, K., Hertz-Picciotto, I., Savitz, D. A., & Blatt, J. (2001). Comparison of Assessment Methods for Pesticide Exposure in a Case-Control Interview Study. *American Journal of Epidemiology*, *153*(12), 1227–1232. https://doi.org/10.1093/aje/153.12.1227

de Vocht, F., Zock, J.-P., Kromhout, H., Sunyer, J., Antó, J. M., Burney, P., & Kogevinas, M. (2005). Comparison of self-reported occupational exposure with a job exposure matrix in an international community-based study on asthma. *American Journal of Industrial Medicine*, *47*(5), 434–442. https://doi.org/10.1002/ajim.20154

DellaValle, C. T., Purdue, M. P., Ward, M. H., Locke, S. J., Stewart, P. A., De Roos, A. J., Hartge, P., Rothman, N., & Friesen, M. C. (2015). Validity of expert assigned retrospective estimates of occupational polychlorinated biphenyl exposure. *The Annals of Occupational Hygiene*, *59*(5), 609–615. https://doi.org/10.1093/annhyg/mev001

Duell, E. J., Millikan, R. C., Savitz, D. A., Schell, M. J., Newman, B., Tse, C. J., & Sandler, D. P. (2001). Reproducibility of reported farming activities and pesticide use among breast cancer cases and controls. A comparison of two modes of data collection. *Annals of Epidemiology*, *11*(3), 178–185.

Févotte, J., Dananché, B., Delabre, L., Ducamp, S., Garras, L., Houot, M., Luce, D., Orlowski, E., Pilorget, C., Lacourt, A., Brochard, P., Goldberg, M., & Imbernon, E. (2011). Matgéné: A Program to Develop Job-Exposure Matrices in the General Population in France. *Annals of Occupational Hygiene*, *55*(8), 865–878. https://doi.org/10.1093/annhyg/mer067

Friesen, M. C., Coble, J. B., Lu, W., Shu, X.-O., Ji, B.-T., Xue, S., Portengen, L., Chow, W.-H., Gao, Y.-T., Yang, G., Rothman, N., & Vermeulen, R. (2012). Combining a Job-Exposure Matrix with Exposure Measurements to Assess Occupational Exposure to Benzene in a Population Cohort in Shanghai, China. *Annals of Occupational Hygiene*, *56*(1), 80–91. https://doi.org/10.1093/annhyg/mer080

Friesen, M. C., Lan, Q., Ge, C., Locke, S. J., Hosgood, D., Fritschi, L., Sadkowsky, T., Chen, Y.-C., Wei, H., Xu, J., Lam, T. H., Kwong, Y. L., Chen, K., Xu, C., Su, Y.-C., Chiu, B. C. H., Ip, K. M. D., Purdue, M. P., Bassig, B. A., … Vermeulen, R. (2016). Evaluation of Automatically Assigned Job-Specific Interview Modules. *The Annals of Occupational Hygiene*, *60*(7), 885–899. https://doi.org/10.1093/annhyg/mew029

Friesen, M. C., Lavoué, J., Teschke, K., & Van Tongeren, M. (2015). Occupational exposure assessment in industry- and population-based epidemiologic studies. In M. J. Nieuwenhuijsen (Ed.), *Exposure Assessment in Environmental Epidemiology* (2nd ed.). Oxford University Press.

Friesen, M. C., Locke, S. J., Tornow, C., Chen, Y.-C., Koh, D.-H., Stewart, P. A., Purdue, M., & Colt, J. S. (2014). Systematically extracting metal- and solvent-related occupational information from free-text responses to lifetime occupational history questionnaires. *The Annals of Occupational Hygiene*, *58*(5), 612–624. https://doi.org/10.1093/annhyg/meu012

Friesen, M. C., Pronk, A., Wheeler, D. C., Chen, Y.-C., Locke, S. J., Zaebst, D. D., Schwenn, M., Johnson, A., Waddell, R., Baris, D., Colt, J. S., Silverman, D. T., Stewart, P. A., & Katki, H. A. (2013). Comparison of Algorithm-based Estimates of Occupational Diesel Exhaust Exposure to Those of Multiple Independent Raters in a Population-based Case–Control Study. *Annals of Occupational Hygiene*, *57*(4), 470–481. https://doi.org/10.1093/annhyg/mes082

Friesen, M. C., Shortreed, S. M., Wheeler, D. C., Burstyn, I., Vermeulen, R., Pronk, A., Colt, J. S., Baris, D., Karagas, M. R., Schwenn, M., Johnson, A., Armenti, K. R., Silverman, D. T., & Yu, K. (2015). Using hierarchical cluster models to systematically identify groups of jobs with similar occupational questionnaire response patterns to assist rule-based expert exposure assessment in population-based studies. *The Annals of Occupational Hygiene*, *59*(4), 455–466. https://doi.org/10.1093/annhyg/meu101

Friesen, M. C., Wheeler, D. C., Vermeulen, R., Locke, S. J., Zaebst, D. D., Koutros, S., Pronk, A., Colt, J. S., Baris, D., Karagas, M. R., Malats, N., Schwenn, M., Johnson, A., Armenti, K. R., Rothman, N., Stewart, P. A., Kogevinas, M., & Silverman, D. T. (2016). Combining Decision Rules from Classification Tree Models and Expert Assessment to Estimate Occupational Exposure to Diesel Exhaust for a Case-Control Study. *The Annals of Occupational Hygiene*, *60*(4), 467–478. https://doi.org/10.1093/annhyg/mev095

Fritschi, L., Glass, D. C., Tabrizi, J. S., Leavy, J. E., & Ambrosini, G. L. (2007). Occupational risk factors for prostate cancer and benign prostatic hyperplasia: A case-control study in Western Australia. *Occupational and Environmental Medicine*, *64*(1), 60–65. https://doi.org/10.1136/oem.2006.027706

Fritschi, L., Siemiatycki, J., & Richardson, L. (1996). Self-assessed versus expert-assessed occupational exposures. *American Journal of Epidemiology*, *144*(5), 521–527.

Fritschi, Lin, Friesen, M. C., Glass, D., Benke, G., Girschik, J., & Sadkowsky, T. (2009). OccIDEAS: Retrospective Occupational Exposure Assessment in Community-Based Studies Made Easier. *Journal of Environmental and Public Health*, *2009*, e957023. https://doi.org/10.1155/2009/957023

Fritschi, Lin, Nadon, L., Benke, G., Lakhani, R., Latreille, B., Parent, M.-E., & Siemiatycki, J. (2003). Validation of expert assessment of occupational exposures. *American Journal of Industrial Medicine*, *43*(5), 519–522. https://doi.org/10.1002/ajim.10208

Gramond, C., Rolland, P., Lacourt, A., Ducamp, S., Chamming's, S., Creau, Y., Hery, M., Laureillard, J., Mohammed-Brahim, B., Orlowski, E., Paris, C., Pairon, J.-C., Goldberg, M., Brochard, P., & for the PNSM Study Group. (2012). Choice of rating method for assessing occupational asbestos exposure: Study for compensation purposes in France. *American Journal of Industrial Medicine*, *55*(5), 440–449. https://doi.org/10.1002/ajim.22008

Hardt, J. S., Vermeulen, R., Peters, S., Kromhout, H., McLaughlin, J. R., & Demers, P. A. (2014). A comparison of exposure assessment approaches: Lung cancer and occupational asbestos exposure in a population-based case-control study. *Occupational and Environmental Medicine*, *71*(4), 282–288. https://doi.org/10.1136/oemed-2013-101735

Heid, I. M., Küchenhoff, H., Miles, J., Kreienbrock, L., & Wichmann, H. E. (2004). Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. *Journal of Exposure Science and Environmental Epidemiology*, *14*(5), 365–377. https://doi.org/10.1038/sj.jea.7500332

Hepworth, S. J., Bolton, A., Parslow, R. C., Tongeren, M. van, Muir, K. R., & McKinney, P. A. (2006). Assigning exposure to pesticides and solvents from self-reports collected by a computer assisted personal interview and expert assessment of job codes: The UK Adult Brain Tumour Study. *Occupational and Environmental Medicine*, *63*(4), 267–272. https://doi.org/10.1136/oem.2005.021022

Hoar, S. K., Morrison, A. S., Cole, P., & Silverman, D. T. (1980). An occupation and exposure linkage system for the study of occupational carcinogenesis. *Journal of Occupational Medicine.: Official Publication of the Industrial Medical Association*, *22*(11), 722–726.

ILO. (2010, June 10). *ISCO-International Standard Classification of Occupations: Brief History*. http://www.ilo.org/public/english/bureau/stat/isco/intro2.htm

Koh, D.-H., Bhatti, P., Coble, J. B., Stewart, P. A., Lu, W., Shu, X.-O., Ji, B.-T., Xue, S., Locke, S. J., Portengen, L., Yang, G., Chow, W.-H., Gao, Y.-T., Rothman, N., Vermeulen, R., & Friesen, M. C. (2014). Calibrating a population-based job-exposure matrix using inspection measurements to estimate historical occupational exposure to lead for a population-based cohort in Shanghai, China. *Journal of Exposure Science and Environmental Epidemiology*, *24*(1), 9–16. https://doi.org/10.1038/jes.2012.86

Kromhout, H., Oostendorp, Y., Heederik, D., & Boleij, J. S. (1987). Agreement between qualitative exposure estimates and quantitative exposure measurements. *American Journal of Industrial Medicine*, *12*(5), 551–562.

Kromhout, Hans, & Vermeulen, R. (2001). Application of job-exposure matrices in studies of the general population: Some clues to their performance. *European Respiratory Review*, *11*(80), 80–90.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lavoué, J., Pintos, J., Tongeren, M. V., Kincl, L., Richardson, L., Kauppinen, T., Cardis, E., & Siemiatycki, J. (2012). Comparison of exposure estimates in the Finnish job-exposure matrix FINJEM with a JEM derived from expert assessments performed in Montreal. *Occup Environ Med*, *69*(7), 465–471. https://doi.org/10.1136/oemed-2011-100154

Lee, D. G., Lavoué, J., Spinelli, J. J., & Burstyn, I. (2015). Statistical Modeling of Occupational Exposure to Polycyclic Aromatic Hydrocarbons Using OSHA Data. *Journal of Occupational and Environmental Hygiene*, *12*(10), 729–742. https://doi.org/10.1080/15459624.2015.1043049

Lillienberg, L., Zock, J.-P., Kromhout, H., Plana, E., Jarvis, D., Torén, K., & Kogevinas, M. (2008). A Population-Based Study on Welding Exposures at Work and Respiratory Symptoms. *The Annals of Occupational Hygiene*, *52*(2), 107–115. https://doi.org/10.1093/annhyg/mem063

Lynge, E., Andersen, A., Rylander, L., Tinnerberg, H., Lindbohm, M.-L., Pukkala, E., Romundstad, P., Jensen, P., Clausen, L. B., & Johansen, K. (2006). Cancer in Persons Working in Dry Cleaning in the Nordic Countries. *Environmental Health Perspectives*, *114*(2), 213–219.

Mannetje, A. 't, Fevotte, J., Fletcher, T., Brennan, P., Legoza, J., Szeremi, M., Paldy, A., Brzeznicki, S., Gromiec, J., Ruxanda-Artenie, C., Stanescu-Dumitru, R., Ivanov, N., Shterengorz, R., Hettychova, L., Krizanova, D., Cassidy, A., van Tongeren, M., & Boffetta, P. (2003). Assessing exposure misclassification by expert assessment in multicenter occupational studies. *Epidemiology (Cambridge, Mass.)*, *14*(5), 585–592. https://doi.org/10.1097/01.ede.0000072108.66723.0f

Nam, J., Rice, C., & Gail, M. H. (2005). Comparison of asbestos exposure assessments by next-of-kin respondents, by an occupational hygienist, and by a job-exposure matrix from the National Occupational Hazard Survey. *American Journal of Industrial Medicine*, *47*(5), 443–450. https://doi.org/10.1002/ajim.20168

Neilson, H. K., Sass-Kortsak, A., Lou, W. Y. W., & Purdham, J. T. (2007). Personal factors influencing agreement between expert and self-reported assessments of an occupational exposure. *Chronic Diseases in Canada*, *28*(1–2), 1–9.

Offermans, N. S. M., Vermeulen, R., Burdorf, A., Peters, S., Goldbohm, R. A., Koeman, T., Tongeren, M. van, Kauppinen, T., Kant, I., Kromhout, H., & Brandt, P. A. van den. (2012). Comparison of expert and job-exposure matrix-based retrospective exposure assessment of occupational carcinogens in the Netherlands Cohort Study. *Occupational and Environmental Medicine*, *69*(10), 745–751. https://doi.org/10.1136/oemed-2011-100556

Olsson, A. C., Vermeulen, R., Schüz, J., Kromhout, H., Pesch, B., Peters, S., Behrens, T., Portengen, L., Mirabelli, D., Gustavsson, P., Kendzia, B., Almansa, J., Luzon, V., Vlaanderen, J., Stücker, I., Guida, F., Consonni, D., Caporaso, N., Landi, M. T., … Straif, K. (2017). Exposure–Response Analyses of Asbestos and Lung Cancer Subtypes in a Pooled Analysis of Case–Control Studies. *Epidemiology (Cambridge, Mass.)*, *28*(2), 288–299. https://doi.org/10.1097/EDE.0000000000000604

Orsi, L., Monnereau, A., Dananche, B., Berthou, C., Fenaux, P., Marit, G., Soubeyran, P., Huguet, F., Milpied, N., Leporrier, M., Hemon, D., Troussard, X., & Clavel, J. (2010). Occupational exposure to organic solvents and lymphoid neoplasms in men: Results of a French case-control study. *Occupational and Environmental Medicine*, *67*(10), 664–672. https://doi.org/10.1136/oem.2009.049460

Pannett, B., Coggon, D., & Acheson, E. D. (1985). A job-exposure matrix for use in population based studies in England and Wales. *British Journal of Industrial Medicine*, *42*(11), 777–783.

Parks, C. G., Cooper, G. S., Nylander-French, L. A., Hoppin, J. A., Sanderson, W. T., & Dement, J. M. (2004). Comparing questionnaire-based methods to assess occupational silica exposure. *Epidemiology (Cambridge, Mass.)*, *15*(4), 433–441.

Peters, S., Glass, D. C., Milne, E., Fritschi, L., & Aus-ALL consortium. (2014). Rule-based exposure assessment versus case-by-case expert assessment using the same information in a community-based study. *Occupational and Environmental Medicine*, *71*(3), 215–219. https://doi.org/10.1136/oemed-2013-101699

Peters, S., Kromhout, H., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Vermeulen, R. (2013). Sensitivity Analyses of Exposure Estimates from a Quantitative Job-exposure Matrix (SYN-JEM) for Use in Community-based Studies. *Annals of Occupational Hygiene*, *57*(1), 98–106. https://doi.org/10.1093/annhyg/mes045

Peters, S., Vermeulen, R., Cassidy, A., Mannetje, A. 't, Tongeren, M. van, Boffetta, P., Straif, K., & Kromhout, H. (2011). Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case–control study. *Occupational and Environmental Medicine*, *68*(2), 148–153. https://doi.org/10.1136/oem.2010.055608

Peters, S., Vermeulen, R., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Kromhout, H. (2011). *Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies*. *13*(11), 3262–3268. https://doi.org/10.1039/C1EM10628G

Peters, S., Vermeulen, R., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Kromhout, H. (2016). SYN-JEM: A Quantitative Job-Exposure Matrix for Five Lung Carcinogens. *Annals of Occupational Hygiene*, *60*(7), 795–811. https://doi.org/10.1093/annhyg/mew034

Pronk, A., Stewart, P. A., Coble, J. B., Katki, H. A., Wheeler, D. C., Colt, J. S., Baris, D., Schwenn, M., Karagas, M. R., Johnson, A., Waddell, R., Verrill, C., Cherala, S., Silverman, D. T., & Friesen, M. C. (2012). Comparison of two expert-based assessments of diesel exhaust exposure in a case–control study: Programmable decision rules versus expert review of individual jobs. *Occupational and Environmental Medicine*, *69*(10), 752–758. https://doi.org/10.1136/oemed-2011-100524

Richiardi, L., Mirabelli, D., Calisti, R., Ottino, A., Ferrando, A., Boffetta, P., & Merletti, F. (2006). Occupational exposure to diesel exhausts and risk for lung cancer in a population-based case-control study in Italy. *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO*, *17*(12), 1842–1847. https://doi.org/10.1093/annonc/mdl307

Rocheleau, C. M., Lawson, C. C., Waters, M. A., Hein, M. J., Stewart, P. A., Correa, A., Echeverria, D., & Reefhuis, J. (2011). Inter-Rater Reliability of Assessed Prenatal Maternal Occupational Exposures to Solvents, Polycyclic Aromatic Hydrocarbons, and Heavy Metals. *Journal of Occupational and Environmental Hygiene*, *8*(12), 718–728. https://doi.org/10.1080/15459624.2011.627293

Segnan, N., Ponti, A., Ronco, G., Kromhout, H., Heederik, D., Cock, J. de, Bosia, S., Luccolo, L., Piccioni, P., Seniori Costantini, A., Miligi, L., Scarpelli, A., Mariotti, M., Scarnato, C., & Morisi, L. (1996). Comparison of methods for assessing the probability of exposure in metal plating, shoe and leather goods manufacture and vine growing. *Occupational Hygiene*. http://agris.fao.org/agris-search/search.do?recordID=NL2012079562

Semple, S. E., Dick, F., & Cherrie, J. W. (2004). Exposure assessment for a population-based case-control study combining a job-exposure matrix with interview data. *Scandinavian Journal of Work, Environment & Health*, *30*(3), 241–248. https://doi.org/10.5271/sjweh.785

Sieber, W. K., Sundin, D. S., Frazier, T. M., & Robinson, C. F. (1991). Development, use, and availability of a job exposure matrix based on national occupational hazard survey data. *American Journal of Industrial Medicine*, *20*(2), 163–174.

Siemiatycki, J., Dewar, R., & Richardson, L. (1989). Costs and statistical power associated with five methods of collecting occupation exposure information for population-based case-control studies. *American Journal of Epidemiology*, *130*(6), 1236–1246.

Stewart, P. A., Lees, P. S., & Francis, M. (1996). Quantification of historical exposures in occupational cohort studies. *Scandinavian Journal of Work, Environment & Health*, *22*(6), 405–414. https://doi.org/10.5271/sjweh.161

Teschke, K., Olshan, A. F., Daniels, J. L., Roos, A. J. D., Parks, C. G., Schulz, M., & Vaughan, T. L. (2002). Occupational exposure assessment in case–control studies: Opportunities for improvement. *Occupational and Environmental Medicine*, *59*(9), 575–594. https://doi.org/10.1136/oem.59.9.575

Teschke, K., Smith, J. C., & Olshan, A. F. (2000). Evidence of recall bias in volunteered vs. Prompted responses about occupational exposures. *American Journal of Industrial Medicine*, *38*(4), 385–388.

Tielemans, E., Heederik, D., Burdorf, A., Vermeulen, R., Veulemans, H., Kromhout, H., & Hartog, K. (1999). Assessment of occupational exposures in a general population: Comparison of different methods. *Occupational and Environmental Medicine*, *56*(3), 145–151.

Tinnerberg, H., Björk, J., & Welinder, H. (2001). Evaluation of occupational and leisure time exposure assessment in a population-based case control study on leukaemia. *International Archives of Occupational and Environmental Health*, *74*(8), 533–540.

Tinnerberg, H., Heikkilä, P., Huici-Montagud, A., Bernal, F., Forni, A., Wanders, S., Welinder, H., Wilhardt, P., Strömberg, U., Norppa, H., Knudsen, L., Bonassi, S., & Hagmar, L. (2003). Retrospective exposure assessment and quality control in an international multi-centre case-control study. *The Annals of Occupational Hygiene*, *47*(1), 37–47.

Vlaanderen, J., Straif, K., Pukkala, E., Kauppinen, T., Kyyrönen, P., Martinsen, J. I., Kjaerheim, K., Tryggvadottir, L., Hansen, J., Sparén, P., & Weiderpass, E. (2013). Occupational exposure to trichloroethylene and perchloroethylene and the risk of lymphoma, liver, and kidney cancer in four Nordic countries. *Occupational and Environmental Medicine*, *70*(6), 393–401. https://doi.org/10.1136/oemed-2012-101188

Westberg, H. B. T., Hardell, L. O., Malmqvist, N., Ohlson, C.-G., & Axelson, O. (2005). On the Use of Different Measures of Exposure—Experiences from a Case-Control Study on Testicular Cancer and PVC Exposure. *Journal of Occupational and Environmental Hygiene*, *2*(7), 351–356. https://doi.org/10.1080/15459620590969046

Wheeler, D. C., Archer, K. J., Burstyn, I., Yu, K., Stewart, P. A., Colt, J. S., Baris, D., Karagas, M. R., Schwenn, M., Johnson, A., Armenti, K., Silverman, D. T., & Friesen, M. C. (2015). Comparison of Ordinal and Nominal Classification Trees to Predict Ordinal Expert-Based Occupational Exposure Estimates in a Case–Control Study. *The Annals of Occupational Hygiene*, *59*(3), 324–335. https://doi.org/10.1093/annhyg/meu098

Wheeler, D. C., Burstyn, I., Vermeulen, R., Yu, K., Shortreed, S. M., Pronk, A., Stewart, P. A., Colt, J. S., Baris, D., Karagas, M. R., Schwenn, M., Johnson, A., Silverman, D. T., & Friesen, M. C. (2013). Inside the black box: Starting to uncover the underlying decision rules used in a one-by-one expert assessment of occupational exposure in case-control studies. *Occupational and Environmental Medicine*, *70*(3), 203–210. https://doi.org/10.1136/oemed-2012-100918

Supplementary Table 1: List of included case-control occupational cancer studies in the general population

Lead Author	Year	Country /Regions	Title	Journal Abbreviation	Pages	Issue	Volume
Adegoke	2004	China	Agreement of job-exposure matrix (JEM) assessed exposure and self-reported exposure among adult leukemia patients and controls in Shanghai	Am. J. Ind. Med.	281-288	3	45
Adegoke	2003	China	Occupational history and exposure and the risk of adult leukemia in Shanghai	Ann Epidemiol	485-494	7	13
Agudo	2000	Spain	Occupation and risk of malignant pleural mesothelioma: A case-control study in Spain	Am. J. Ind. Med.	159-168	2	37
Aguilar- Madrid	2010	Mexico	Case-control study of pleural mesothelioma in workers with social security in Mexico	Am. J. Ind. Med.	241-251	3	53
Ahrens	1993	Germany	Retrospective assessment of asbestos exposure--I. Case-control analysis in a study of lung cancer: efficiency of job-specific questionnaires and job exposure matrices	Int J Epidemiol	S83-95		22 Suppl 2
Ahrens	2007	EU	Occupational exposure to endocrine-disrupting compounds and biliary tract cancer among men	Scand J Work Environ Health	387-396	5	33
Ajani	1992	US	Occupation and risk of uveal melanoma. An exploratory study	Cancer	2891-2900	12	70
Arbman	1993	Sweden	Do occupational factors influence the risk of colon and rectal cancer in different ways?	Cancer	2543-2549	9	72
Aschengrau	1998	US	Occupational exposure to estrogenic chemicals and the occurrence of breast cancer: an exploratory analysis	Am. J. Ind. Med.	6-14	1	34
Band	2011	Canada	Prostate cancer risk and exposure to pesticides in British Columbia farmers	Prostate	168-183	2	71
Baris	2004	US	Occupation, pesticide exposure and risk of multiple myeloma	Scand J Work Environ Health	215-222	3	30
Becher	1993	Poland	Effect of occupational air pollutants on various histological types of lung cancer: a population based case-control study	Br J Ind Med	136-142	2	50
Becher	2005	Germany	Occupation, exposure to polycyclic aromatic hydrocarbons and laryngeal cancer risk	Int. J. Cancer	451-457	3	116
Behrens	2010	9 EU countries	Hormonal exposures and the risk of uveal melanoma	Cancer Causes Control	1625-1634	10	21
Behrens	2012	Germany	Pesticide exposure in farming and forestry and the risk of uveal melanoma	Cancer Causes Control	141-151	1	23
Benke	1997	Australia	Retrospective assessment of occupational exposure to chemicals in community-based studies: validity and repeatability of industrial hygiene panel ratings	Int J Epidemiol	635-642	3	26
Berrino	2003	EU	Occupation and larynx and hypopharynx cancer: a job-exposure matrix approach in an international case-control study in France, Italy, Spain and Switzerland	Cancer Causes Control	213-223	3	14
Beveridge	2010	Canada	Lung cancer risk associated with occupational exposure to nickel, chromium VI, and cadmium in two population-based case-control studies in Montreal	Am. J. Ind. Med.	476-485	5	53
Bhatti	2011	Canada	Wood dust exposure and risk of lung cancer	Occup Environ Med	599-604	8	68
Black	2004	Australia	Artificial neural networks and job-specific modules to assess occupational exposure	Ann Occup Hyg	595-600	7	48

Blair	1993	US	Evaluation of risks for non-Hodgkin's lymphoma by occupation and industry exposures from a case-control study	Am. J. Ind. Med.	301-312	2	23
Blair	2001	US	Occupation and leukemia: a population-based case-control study in Iowa and Minnesota	Am. J. Ind. Med.	3-14	1	40
Boffetta	1999	7 EU countries	Exposure to environmental tobacco smoke and risk of adenocarcinoma of the lung	Int. J. Cancer	635-639	5	83
Boffetta	2001	Canada	Exposure to titanium dioxide and risk of lung cancer in a population-based study from Montreal	Scand J Work Environ Health	227-232	4	27
Bonassi	1989	Italy	Bladder cancer and occupational exposure to polycyclic aromatic hydrocarbons	Int. J. Cancer	648-651	4	44
Bourgkard	2013	France	Comparison of exposure assessment methods in a lung cancer case-control study: performance of a lifelong task-based questionnaire for asbestos and PAHs	Occup Environ Med	884-891	12	70
Brandi	2013	Italy	Asbestos: a hidden player behind the cholangiocarcinoma increase? Findings from a case-control analysis	Cancer Causes Control	911-918	5	24
Briggs	2003	US	Occupational risk factors for selected cancers among African American and White men in the United States	Am J Public Health	1748-1752	10	93
Brown	1990	US	Pesticide exposures and other agricultural risk factors for leukemia among men in Iowa and Minnesota	Cancer Res.	6585-6591	20	50
Brown	1993	US	Pesticide exposures and multiple myeloma in Iowa men	Cancer Causes Control	153-156	2	4

Brownson	1993	US	Occupational risk factors for lung cancer among nonsmoking women: a case-control study in Missouri (United States)	Cancer Causes Control	449-454	5	4
Bruske-Hohlfeld	1999	Germany	Lung cancer risk in male workers occupationally exposed to diesel motor emissions in Germany	Am. J. Ind. Med.	405-414	4	36
Bruske-Hohlfeld	2000	Germany	Occupational lung cancer risk for men in Germany: results from a pooled case-control study	Am. J. Epidemiol.	384-395	4	151
Cantor	1995	US	Occupational exposures and female breast cancer mortality in the United States	J. Occup. Environ. Med.	336-348	3	37
Carreon	2005	US	Gliomas and farm pesticide exposure in women: the Upper Midwest Health Study	Environ. Health Perspect.	546-551	5	113
Charbotel	2006	France	Case-control study on renal cell cancer and occupational exposure to trichloroethylene. Part II: Epidemiological aspects	Ann Occup Hyg	777-787	8	50
Chatzis	1999	Greece	Lung cancer and occupational risk factors in Greece	J. Occup. Environ. Med.	29-35	1	41
Chiu	2004	US	Agricultural pesticide use, familial cancer, and risk of non- Hodgkin lymphoma	Cancer Epidemiol. Biomarkers Prev.	525-531	4	13
Christensen	2015	Canada	Lack of a protective effect of cotton dust on risk of lung cancer: evidence from two population-based case-control studies	BMC Cancer	212		15
Christensen	2013	Canada	Risk of selected cancers due to occupational exposure to chlorinated solvents in a case-control study in Montreal	J. Occup. Environ. Med.	198-208	2	55
Clavel	1996	France	Hairy cell leukaemia and occupational exposure to benzene	Occup Environ Med	533-539	8	53

Coble	2003	Puerto Rico	Sugarcane farming, occupational solvent exposures, and the risk of oral cancer in Puerto Rico	J. Occup. Environ. Med.	869-874	8	45
Cocco	1998	US	Occupational risk factors for cancer of the central nervous system: a case-control study on death certificates from 24 U.S. states	Am. J. Ind. Med.	247-255	3	33
Cocco	2013	US	Occupational exposure to trichloroethylene and risk of non-Hodgkin lymphoma and its major subtypes: a pooled InterLymph [correction of IinterLlymph] analysis	Occup Environ Med	795-802	11	70
Cocco	1998	International	Occupational risk factors for cancer of the gastric cardia. Analysis of death certificates from 24 US states	J. Occup. Environ. Med.	855-861	10	40
Colt	2014	US	A case-control study of occupational exposure to metalworking fluids and bladder cancer risk among men	Occup Environ Med	667-674	10	71
Colt	2011	US	Occupation and bladder cancer in a population-based case-control study in Northern New England	Occup Environ Med	239-249	4	68
Cordier	1993	France	Occupational risks of bladder cancer in France: a multicentre case-control study	Int J Epidemiol	403-411	3	22
Costantini	2008	Italy	Risk of leukemia and multiple myeloma associated with exposure to benzene and other organic solvents: evidence from the Italian Multicenter Case-control study	Am. J. Ind. Med.	803-811	11	51
De Matteis	2012	Italy	Impact of occupational carcinogens on lung cancer risk in a general population	Int J Epidemiol	711-721	3	41
DellaValle	2015		Validity of expert assigned retrospective estimates of occupational polychlorinated biphenyl exposure	Ann Occup Hyg	609-615	5	59
Demers	1994	US	Construction occupations, asbestos exposure, and cancer of the colon and rectum	J Occup Med	1027-1031	9	36
Deng	2013	US	Occupational solvent exposure, genetic variation in immune genes, and the risk for non-Hodgkin lymphoma	Eur. J. Cancer Prev.	77-82	1	22
Dietz	2004	Germany	Exposure to cement dust, related occupational groups and laryngeal cancer risk: results of a population based case-control study	Int. J. Cancer	907-911	6	108
Dosemeci	1999	US	Gender differences in risk of renal cell carcinoma and occupational exposures to chlorinated aliphatic hydrocarbons	Am. J. Ind. Med.	54-59	1	36
Duell	2001	US	Reproducibility of reported farming activities and pesticide use among breast cancer cases and controls. A comparison of two modes of data collection	Ann Epidemiol	178-185	3	11
Dumas	2000	Canada	Rectal cancer and occupational risk factors: a hypothesis-generating, exposure-based case-control study	Int. J. Cancer	874-879	6	87
Eheman	1999	US	Estimating occupational radiation doses when individual dosimetry information is not available: a job exposure matrix	Am. J. Ind. Med.	348-359	3	36

Ekburanawat	2010	Thailand	Evaluation of non-viral risk factors for nasopharyngeal carcinoma in Thailand: results from a case-control study	Asian Pac. J. Cancer Prev.	929-932	4	11
Ekpanyaskul	2015	Thailand	Semi-quantitative exposure assessment of occupational exposure to wood dust and nasopharyngeal cancer risk	Asian Pac. J. Cancer Prev.	4339-4345	10	16
Ekstrom	1999	Sweden	Occupational exposures and risk of gastric cancer in a population-based case-control study	Cancer Res.	5932-5937	23	59
Elghany	1990	US	Occupation, cadmium exposure, and prostate cancer	Epidemiology	107-115	2	1
El-Zaemey	2014	Australia	Household and occupational exposure to pesticides and risk of breast cancer	Int J Environ Health Res	91-102	2	24
Eriksson	1992	Sweden	Occupational and other environmental factors and multiple myeloma: a population based case-control study	Br J Ind Med	95-103	2	49
Fabbro- Peray	2001	France	Environmental risk factors for non-Hodgkin's lymphoma: a population-based case-control study in Languedoc-Roussillon, France	Cancer Causes Control	201-212	3	12
Ferrante	2016	Italy	Pleural mesothelioma and occupational and non-occupational asbestos exposure: a case-control study with quantitative risk assessment	Occup Environ Med	147-153	3	73
Fevotte	2006	France	Case-control study on renal cell cancer and occupational exposure to trichloroethylene. Part I: Exposure assessment	Ann Occup Hyg	765-775	8	50
Friesen	2014	US	Systematically extracting metal- and solvent-related occupational information from free-text responses to lifetime occupational history questionnaires	Ann Occup Hyg	612-624	5	58
Friesen	2014	US	Developing estimates of frequency and intensity of exposure to three types of metalworking fluids in a population-based case- control study of bladder cancer	Am. J. Ind. Med.	915-927	8	57
Friesen	2013	US	Comparison of algorithm-based estimates of occupational diesel exhaust exposure to those of multiple independent raters in a population-based case-control study	Ann Occup Hyg	470-481	4	57
Friesen	2015	US	Using hierarchical cluster models to systematically identify groups of jobs with similar occupational questionnaire response patterns to assist rule-based expert exposure assessment in population-based studies	Ann Occup Hyg	455-466	4	59
Fritschi	2005	Australia	Occupational exposure to pesticides and risk of non-Hodgkin's lymphoma	Am. J. Epidemiol.	849-857	9	162
Fritschi	2007	Australia	Occupational risk factors for prostate cancer and benign prostatic hyperplasia: a case-control study in Western Australia	Occup Environ Med	60-65	1	64
Fritschi	1996	Canada	Lymphoma, myeloma and occupation: results of a case-control study	Int. J. Cancer	498-503	4	67
Fritschi	1996	Canada	Melanoma and occupation: results of a case-control study	Occup Environ Med	168-173	3	53
Fritschi	1996	Canada	Self-assessed versus expert-assessed occupational exposures	Am. J. Epidemiol.	521-527	5	144

Fritschi	2005	Australia	Risk of non-Hodgkin lymphoma associated with occupational exposure to solvents, metals, organic dusts and PCBs (Australia)	Cancer Causes Control	599-607	5	16
Fritschi	2015	Australia	Occupational exposure to N-nitrosamines and pesticides and risk of pancreatic cancer	Occup Environ Med	678-683	9	72
Fritschi	2003	Australia	Validation of expert assessment of occupational exposures	Am. J. Ind. Med.	519-522	5	43
Fryzek	1997	US	A case-control study of self-reported exposures to pesticides and pancreas cancer in southeastern Michigan	Int. J. Cancer	62-67	1	72
Gago-Dominguez	2001	US	Use of permanent hair dyes and bladder-cancer risk	Int. J. Cancer	575-579	4	91
Gallagher	1996	Canada	Chemical exposures, medical history, and risk of squamous and basal cell carcinoma of the skin	Cancer Epidemiol. Biomarkers Prev.	419-424	6	5
Garabedian	1999	US	Occupational chlorophenol exposure and non-Hodgkin's lymphoma	J. Occup. Environ. Med.	267-272	4	41
Garabrant	1992	US	Asbestos and colon cancer: lack of association in a large case- control study	Am. J. Epidemiol.	843-853	8	135
Gerin	1998	Canada	Associations between several sites of cancer and occupational exposure to benzene, toluene, xylene, and styrene: results of a case-control study in Montreal	Am. J. Ind. Med.	144-156	2	34
Gerin	1989	Canada	Cancer risks due to occupational exposure to formaldehyde: results of a multi-site case-control study in Montreal	Int. J. Cancer	53-58	1	44
Glass	2015	Australia	Occupational exposure to solvents and risk of breast cancer	Am. J. Ind. Med.	915-922	9	58
Goldberg	2001	Canada	A case-control study of the relationship between the risk of colon cancer in men and exposures to occupational agents	Am. J. Ind. Med.	531-546	6	39
Goldberg	1986	Canada	Inter-rater agreement in assessing occupational exposure in a case-control study	Br J Ind Med	667-676	10	43
Grimsrud	1998	Norway	Lung and bladder cancer in a Norwegian municipality with iron and steel producing industry: population based case-control studies	Occup Environ Med	387-392	6	55
Guida	2013	France	Risk of lung cancer associated with occupational exposure to mineral wools: updating knowledge from a french population-based case-control study, the ICARE study	J. Occup. Environ. Med.	786-795	7	55
Guo	2009	China	Evaluation of nonviral risk factors for nasopharyngeal carcinoma in a high-risk population of Southern China	Int. J. Cancer	2942-2947	12	124
Gustavsson	2002	Sweden	Low-dose exposure to asbestos and lung cancer: dose-response relations and interaction with smoking in a population-based case-referent study in Stockholm, Sweden	Am. J. Epidemiol.	1016-1022	11	155
Hardell	2004	Sweden	Testicular cancer and occupational exposure to polyvinyl chloride plastics: a case-control study	Int. J. Cancer	425-429	3	109
Hardt	2014	Canada	A comparison of exposure assessment approaches: lung cancer and occupational asbestos exposure in a population-based case-control study	Occup Environ Med	282-288	4	71

Hartge	1994	US	Occupation and ovarian cancer: a case-control study in the Washington, DC, metropolitan area, 1978-1981	J Occup Med	924-927	8	36
Hayes	1990	US	Occupation and risk for testicular cancer: a case-control study	Int J Epidemiol	825-831	4	19
Hayes	1993	US	Are the known bladder cancer risk-factors associated with more advanced bladder cancer?	Cancer Causes Control	157-162	2	4
Heavner	2015	US	Working environment and myeloproliferative neoplasm: A population-based case-control study following a cluster investigation	Am. J. Ind. Med.	595-604	6	58
Heineman	1992	Denmark	Occupational risk factors for multiple myeloma among Danish men	Cancer Causes Control	555-568	6	3
Hinds	1985	US	Application of a job-exposure matrix to a case-control study of lung cancer	J. Natl. Cancer Inst.	193-197	2	75
Hoffmann	2008	Germany	Population-based research on occupational and environmental factors for leukemia and non-Hodgkin's lymphoma: the Northern Germany Leukemia and Lymphoma Study (NLL)	Am. J. Ind. Med.	246-257	4	51
Holly	1997	US	Non-Hodgkin's lymphoma in homosexual men in the San Francisco Bay Area: occupational, chemical, and environmental exposures	J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.	223-231	3	15
Hoppin	1999	US	Occupational risk factors for sarcoma subtypes	Epidemiology	300-306	3	10
Horn-Ross	1997	US	Environmental factors and the risk of salivary gland cancer	Epidemiology	414-419	4	8
Hours	1994	France	Bladder cancer and occupational exposures	Scand J Work Environ Health	322-330	5	20
Howe	1980	Canada	Tobacco use, occupation, coffee, various nutrients, and bladder cancer	J. Natl. Cancer Inst.	701-713	4	64
Hu	1999	Canada	Risk factors for meningioma in adults: a case-control study in northeast China	Int. J. Cancer	299-304	3	83
Hu	2002	Canada	Risk factors for lung cancer among Canadian women who have never smoked	Cancer Detect. Prev.	129-138	2	26
Hu	1999	China	When to be skeptical of negative studies: pitfalls in evaluating occupational risks using population-based case-control studies	Can J Public Health	138-142	2	90
Iwatsubo	1998	France	Pleural mesothelioma: dose-response relation at low levels of asbestos exposure in a French population-based case-control study	Am. J. Epidemiol.	133-142	2	148
Jansson	2006	Sweden	Airborne occupational exposures and risk of oesophageal and cardia adenocarcinoma	Occup Environ Med	107-112	2	63
Jarvholm	1993	Sweden	Quantitative importance of asbestos as a cause of lung cancer in a Swedish industrial city: a case-referent study	Eur. Respir. J.	1271-1275	9	6
Jayaprakash	2008	US	Wood dust exposure and the risk of upper aero-digestive and respiratory cancers in males	Occup Environ Med	647-654	10	65
Ji	1999	China	Occupation and pancreatic cancer risk in Shanghai, China	Am. J. Ind. Med.	76-81	1	35
Ji	2001	US	Occupational exposure to pesticides and pancreatic cancer	Am. J. Ind. Med.	92-99	1	39
Jiao	2012	US	Occupational solvent exposure, genetic variation of DNA repair genes, and the risk of non-Hodgkin's lymphoma	Eur. J. Cancer Prev.	580-584	6	21

Jockel	1998	Germany	Occupational risk factors for lung cancer: a case-control study in West Germany	Int J Epidemiol	549-560	4	27
Jockel	1998	Germany	Lung cancer risk and welding: results from a case-control study in Germany	Am. J. Ind. Med.	313-320	4	33
Johnson	2001	US	Lifetime residential and workplace exposure to environmental tobacco smoke and lung cancer in never-smoking women, Canada 1994-97	Int. J. Cancer	902-906	6	93
Johnson	1993	Canada	Data on prior pesticide use collected from self- and proxy respondents	Epidemiology	157-164	2	4
Kachuri	2014	Canada	Occupational exposure to crystalline silica and the risk of lung cancer in Canadian men	Int. J. Cancer	138-148	1	135
Kachuri	2016	Canada	Workplace exposure to diesel and gasoline engine exhausts and the risk of colorectal cancer in Canadian men	Environ Health	4	1	15
Kaerlev	2002	EU	Occupational risk factors for small bowel carcinoid tumor: a European population-based case-control study	J. Occup. Environ. Med.	516-522	6	44
Karunanayake	2008	Canada	Occupational exposures and non-Hodgkin's lymphoma: Canadian case-control study	Environ Health	44		7
Karunanayake	2012	Canada	Hodgkin lymphoma and pesticides exposure in men: a Canadian case-control study	J Agromedicine	30-39	1	17
Kasim	2005	Canada	Environmental tobacco smoke and risk of adult leukemia	Epidemiology	672-680	5	16
Kato	2005	US	Personal and occupational exposure to organic solvents and risk of non-Hodgkin's lymphoma (NHL) in women (United States)	Cancer Causes Control	1215-1224	10	16
Kato	2004	US	Pesticide product use and risk of non-Hodgkin lymphoma in women	Environ. Health Perspect.	1275-1281	13	112

Kauppinen	1992	Finland	Magnitude of misclassification bias when using a job-exposure matrix	Scand J Work Environ Health	105-112	2	18
Kauppinen	1992	Finland	Primary liver cancer and occupational exposure	Scand J Work Environ Health	18-25	1	18
Kellen	2007	Belgium	Does occupational exposure to PAHs, diesel and aromatic amines interact with smoking and metabolic genetic polymorphisms to increase the risk on bladder cancer?; The Belgian case control study on bladder cancer risk	Cancer Lett.	51-60	1-2	245
Kernan	1999	US	Occupational risk factors for pancreatic cancer: a case-control study based on death certificates from 24 U.S. states	Am. J. Ind. Med.	260-270	2	36
Kiran	2010	EU	Occupational exposure to ethylene oxide and risk of lymphoma	Epidemiology	905-910	6	21
Kogevinas	2003	EU	Occupation and bladder cancer among men in Western Europe	Cancer Causes Control	907-914	10	14
Koh	2014	US	Calibrating a population-based job-exposure matrix using inspection measurements to estimate historical occupational exposure to lead for a population-based cohort in Shanghai, China	J Expo Sci Environ Epidemiol	9-16	1	24
Kokouva	2013	Greece	Relationship between the paraoxonase 1 (PON1) M55L and Q192R polymorphisms and lymphohaematopoietic cancers in a Greek agricultural population	Toxicology	12-16		307
Kreuzer	1999	Germany	Occupational risk factors for lung cancer among young men	Scand J Work Environ Health	422-429	5	25

Krstev	2005	Poland	Occupation and risk of stomach cancer in Poland	Occup Environ Med	318-324	5	62
Lacourt	2014	France	Occupational and non-occupational attributable risk of asbestos exposure for malignant pleural mesothelioma	Thorax	532-539	6	69
Lacourt	2013	13 countries	INTEROCC case-control study: lack of association between glioma tumors and occupational exposure to selected combustion products, dusts and other chemical agents	BMC Public Health	340		13
Lacourt	2012	France	Temporal patterns of occupational asbestos exposure and risk of pleural mesothelioma	Eur. Respir. J.	1304-1312	6	39
Lacourt	2015	Canada	Lung cancer risk among workers in the construction industry: results from two case-control studies in Montreal	BMC Public Health	941		15
Lacourt	2010	France	Attributable risk in men in two French case-control studies on mesothelioma and asbestos	Eur. J. Epidemiol.	799-806	11	25
Langevin	2013	US	Occupational dust exposure and head and neck squamous cell carcinoma risk in a population-based case-control study conducted in the greater Boston area	Cancer Med	978-986	6	2
Langevin	2013	US	Occupational asbestos exposure is associated with pharyngeal squamous cell carcinoma in men from the greater Boston area	Occup Environ Med	858-863	12	70
Latifovic	2015	Canada	Bladder cancer and occupational exposure to diesel and gasoline engine emissions among Canadian men	Cancer Med	1948-1962	12	4

Lee	2015	Canada	Statistical Modeling of Occupational Exposure to Polycyclic Aromatic Hydrocarbons Using OSHA Data	J Occup Environ Hyg	729-742	10	12
Lee	2005	US	Agricultural pesticide use and risk of glioma in Nebraska, United States	Occup Environ Med	786-792	11	62
Lee	2006	US	Pesticide exposure and lung cancer mortality in Leningrad province in Russia	Environ Int	412-416	3	32
Lee	2006	Russia	Asthma history, occupational exposure to pesticides and the risk of non-Hodgkin's lymphoma	Int. J. Cancer	3174-3176	12	118
Linet	1987	US	A case-control study of multiple myeloma in whites: chronic antigenic stimulation, occupation, and drug use	Cancer Res.	2978-2981	11	47
Linet	1987	US	Comparison of methods for determining occupational exposure in a case-control interview study of chronic lymphocytic leukemia	J Occup Med	136-141	2	29
Luce	1993	France	Sources of discrepancies between a job exposure matrix and a case by case expert assessment for occupational exposure to formaldehyde and wood-dust	Int J Epidemiol	S113-120		22 Suppl 2
Luce	2011	France	Investigation of occupational and environmental causes of respiratory cancers (ICARE): a multicenter, population-based case-control study in France	BMC Public Health	928		11
Luqman	2014	Pakistan	Risk factors for lung cancer in the Pakistani population	Asian Pac. J. Cancer Prev.	3035-3039	7	15

Mahboubi	2013	Canada	Assessment of the effect of occupational exposure to formaldehyde on the risk of lung cancer in two Canadian population-based case-control studies	Scand J Work Environ Health	401-410	4	39
't Mannetje	2003	EU	Assessing exposure misclassification by expert assessment in multicenter occupational studies	Epidemiology	585-592	5	14
Mao	2000	Canada	Non-Hodgkin's lymphoma and occupational exposure to chemicals in Canada. Canadian Cancer Registries Epidemiology Research Group	Ann. Oncol.	69-73		11 Suppl 1
Marsh	1998	US	A case-control study of lung cancer mortality in four rural Arizona smelter towns	Arch. Environ. Health	15-28	1	53
Matos	2000	Argentina	Occupational exposures and lung cancer in Buenos Aires, Argentina	J. Occup. Environ. Med.	653-659	6	42
Matrat	2015	France	Occupational Exposure to Diesel Motor Exhaust and Lung Cancer: A Dose-Response Relationship Hidden by Asbestos Exposure Adjustment? The ICARE Study	J Cancer Epidemiol	879302		2015
Matrat	2016	France	Welding, a risk factor of lung cancer: the ICARE study	Occup Environ Med			
Mattei	2014	France	Exposure to chlorinated solvents and lung cancer: results of the ICARE study	Occup Environ Med	681-689	10	71
Maule	2007	Italy	Modeling mesothelioma risk associated with environmental asbestos exposure	Environ. Health Perspect.	1066-1071	7	115
McClean	2011	US	A case-control study of asphalt and tar exposure and lung cancer in minorities	Am. J. Ind. Med.	811-818	11	54
McHugh	2010	US	Assessing environmental and occupational risk factors for lung cancer in Mexican-Americans	Cancer Causes Control	2157-2164	12	21
Menvielle	2003	New Caledonia	Occupational exposures and lung cancer in New Caledonia	Occup Environ Med	584-589	8	60
Menvielle	2016	France	The joint effect of asbestos exposure, tobacco smoking and alcohol drinking on laryngeal cancer risk: evidence from the French population-based case-control study, ICARE	Occup Environ Med	28-33	1	73
Merler	1986	Italy	On the causal association between exposure to leather dust and nasal cancer: further evidence from a case-control study	Br J Ind Med	91-95	2	43
Merletti	1991	Italy	Occupation and cancer of the oral cavity or oropharynx in Turin, Italy	Scand J Work Environ Health	248-254	4	17
Merletti	2006	EU	Occupational factors and risk of adult bone sarcomas: a multicentric case-control study in Europe	Int. J. Cancer	721-727	3	118
Miligi	2006	Italy	Occupational exposure to solvents and the risk of lymphomas	Epidemiology	552-561	5	17
Miligi	2003	Italy	Non-Hodgkin's lymphoma, leukemia, and exposures in agriculture: results from the Italian multicenter case-control study	Am. J. Ind. Med.	627-636	6	44
Miligi	2006	Italy	Cancer and pesticides: an overview and some results of the Italian multicenter case-control study on hematolymphopoietic malignancies	Ann. N. Y. Acad. Sci.	366-377		1076

Mirabelli	2000	US	Occupational exposure to chlorophenol and the risk of nasal and nasopharyngeal cancers among U.S. men aged 30 to 60	Am. J. Ind. Med.	532-541	5	37
Mirabelli	2009	Italy	Occupational exposure to high molecular weight allergens and lymphoma risk among Italian adults	Cancer Epidemiol. Biomarkers Prev.	2650-2654	10	18
Mommsen	1984	Denmark	Occupational exposures as risk indicator of male bladder carcinoma in a predominantly rural area	Acta Radiol Oncol	147-152	2-3	23
Morabia	1992	US	Lung cancer and occupation: results of a multicentre case-control study	Br J Ind Med	721-727	10	49
Morales-Suarez- Varela	2005	EU	Occupational exposures and mycosis fungoides. A European multicentre case-control study (Europe)	Cancer Causes Control	1253-1259	10	16
Muscat	1998	US	Lung cancer risk and workplace exposures in black men and women	Environ. Res.	78-84	2	76
Nanni	1996	Italy	Chronic lymphocytic leukaemias and non-Hodgkin's lymphomas by histological type in farming-animal breeding workers: a population case-control study based on a priori exposure matrices	Occup Environ Med	652-657	10	53
Navaranjan	2013	Canada	Exposures to multiple pesticides and the risk of Hodgkin lymphoma in Canadian men	Cancer Causes Control	1661-1673	9	24
Nordstrm	1998	Sweden	Occupational exposures, animal exposure and smoking as risk factors for hairy cell leukaemia evaluated in a case-control study	Br. J. Cancer	2048-2052	11	77
Nyberg	2000	Sweden	Urban air pollution and lung cancer in Stockholm	Epidemiology	487-495	5	11

Ohlson	2000	Sweden	Testicular cancer and occupational exposures with a focus on xenoestrogens in polyvinyl chloride plastics	Chemosphere	1277-1282	9-11	40
Olsson	2010	EU	Occupational exposure to polycyclic aromatic hydrocarbons and lung cancer risk: a multicenter study in Europe	Occup Environ Med	98-103	2	67
Olsson	2011	EU and Canada	Exposure to diesel motor exhaust and lung cancer risk in a pooled analysis from case-control studies in Europe and Canada	Am. J. Respir. Crit. Care Med.	941-948	7	183
Orlowski	1993	Germany	Retrospective assessment of asbestos exposure--II. At the job level: complementarity of job-specific questionnaire and job exposure matrices	Int J Epidemiol	S96-105		22 Suppl 2
Orsi	2010	France	Occupational exposure to organic solvents and lymphoid neoplasms in men: results of a French case-control study	Occup Environ Med	664-672	10	67
Pahwa	2011	Canada	Soft-tissue sarcoma and pesticides exposure in men: results of a Canadian case-control study	J. Occup. Environ. Med.	1279-1286	11	53
Pahwa	2006	Canada	Hodgkin lymphoma, multiple myeloma, soft tissue sarcomas, insect repellents, and phenoxyherbicides	J. Occup. Environ. Med.	264-274	3	48
Pannett	1985	UK	A job-exposure matrix for use in population based studies in England and Wales	Br J Ind Med	777-783	11	42
Parent	2000	Canada	Occupational risk factors for renal cell carcinoma in Montreal	Am. J. Ind. Med.	609-618	6	38
Parent	1998	Canada	Occupational exposures and gastric cancer	Epidemiology	48-55	1	9
Parent	2000	Canada	Workplace exposures and oesophageal cancer	Occup Environ Med	325-334	5	57

Parent	1996	Canada	Case-control study of exposure to carbon black in the occupational setting and risk of lung cancer	Am. J. Ind. Med.	285-292	3	30
Parent	2007	Canada	Exposure to diesel and gasoline engine emissions and the risk of lung cancer	Am. J. Epidemiol.	53-62	1	165
Paris	2010	France	Relationships between lung adenocarcinoma and gender, age, smoking and occupational risk factors: A case-case study	Lung Cancer	146-153	2	68
Park	2014	US	Estimation of the probability of exposure to machining fluids in a population-based case-control study	J Occup Environ Hyg	757-770	11	11
Partanen	1991	Finland	Renal cell cancer and occupational exposure to chemical agents	Scand J Work Environ Health	231-239	4	17
Pastorino	1984	Italy	Proportion of lung cancers due to occupational exposure	Int. J. Cancer	231-237	2	33
Pearce	1986	US	Non-Hodgkin's lymphoma and exposure to phenoxyherbicides, chlorophenols, fencing work, and meat works employment: a case-control study	Br J Ind Med	75-83	2	43
Peplonska	2010	Poland	Occupational exposure to organic solvents and breast cancer in women	Occup Environ Med	722-729	11	67
Pesch	2000	Germany	Occupational risk factors for urothelial carcinoma: agent-specific results from a case-control study in Germany. MURC Study Group. Multicenter Urothelial and Renal Cancer	Int J Epidemiol	238-247	2	29
Pesch	2013	EU	N-acetyltransferase 2 phenotype, occupation, and bladder cancer risk: results from the EPIC cohort	Cancer Epidemiol. Biomarkers Prev.	2055-2065	11	22
Peters	2014	Australia	Rule-based exposure assessment versus case-by-case expert assessment using the same information in a community-based study	Occup Environ Med	215-219	3	71
Peters	2012	EU and Canada	Occupational exposure to organic dust increases lung cancer risk in the general population	Thorax	111-116	2	67
Peters	2011	EU	Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case-control study	Occup Environ Med	148-153	2	68
Peters	2011	EU and Canada	Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies	J. Environ. Monit.	3262-3268	11	13
Peters	2016	EU and Canada	SYN-JEM: A Quantitative Job-Exposure Matrix for Five Lung Carcinogens	Ann Occup Hyg	mew034		
Pintos	2009	Canada	Risk of mesothelioma and occupational exposure to asbestos and man-made vitreous fibers: evidence from two case-control studies in Montreal, Canada	J. Occup. Environ. Med.	1177-1184	10	51
Pintos	2012	Canada	Occupational exposure to diesel engine emissions and risk of lung cancer: evidence from two case-control studies in Montreal, Canada	Occup Environ Med	787-792	11	69
Pintos	2008	Canada	Occupational exposure to asbestos and man-made vitreous fibers, and risk of lung cancer: evidence from two case-control studies in Montreal, Canada	J. Occup. Environ. Med.	1273-1281	11	50

Pohlabeln	2000	Germany	Lung cancer and exposure to man-made vitreous fibers: results from a pooled case-control study in Germany	Am. J. Ind. Med.	469-477	5	37
Pohlabeln	2002	Germany	Asbestos fibreyears and lung cancer: a two phase case-control study with expert exposure assessment	Occup Environ Med	410-414	6	59
Preller	2010	Netherlands	Occupational exposure to silica and lung cancer risk in the Netherlands	Occup Environ Med	657-663	10	67
Pronk	2012	US	Comparison of two expert-based assessments of diesel exhaust exposure in a case-control study: programmable decision rules versus expert review of individual jobs	Occup Environ Med	752-758	10	69
Provost	2007	France	Brain tumours and exposure to pesticides: a case-control study in southwestern France	Occup Environ Med	509-514	8	64
Purdue	2009	US	Degreasing and risk of non-Hodgkin lymphoma	Occup Environ Med	557-560	8	66
Purdue	2011	US	A case-control study of occupational exposure to trichloroethylene and non-Hodgkin lymphoma	Environ. Health Perspect.	232-238	2	119
Rabstein	2010	Germany	N-acetyltransferase 2, exposure to aromatic and heterocyclic amines, and receptor-defined breast cancer	Eur. J. Cancer Prev.	100-109	2	19
Ramanakumar	2008	Canada	Risk of lung cancer following exposure to carbon black, titanium dioxide and talc: results from two case-control studies in Montreal	Int. J. Cancer	183-189	1	122
Ramanakumar	2011	Canada	Exposures in painting-related occupations and risk of lung cancer among men: results from two case-control studies in Montreal	Occup Environ Med	44-51	1	68
Ramroth	2011	Germany	Occupational asbestos exposure as a risk factor for laryngeal carcinoma in a population-based case-control study from Germany	Am. J. Ind. Med.	510-514	7	54
Ramroth	2008	Germany	Occupational wood dust exposure and the risk of laryngeal cancer: a population based case-control study in Germany	Am. J. Ind. Med.	648-655	9	51
Rauscher	2003	US, Canada	Is family history of breast cancer a marker of susceptibility to exposures in the incidence of de novo adult acute leukemia?	Cancer Epidemiol. Biomarkers Prev.	289-294	4	12
Richardson	2008	Germany	Occupational risk factors for non-Hodgkin's lymphoma: a population-based case-control study in Northern Germany	Am. J. Ind. Med.	258-268	4	51
Richiardi	2006	Italy	Occupational exposure to diesel exhausts and risk for lung cancer in a population-based case-control study in Italy	Ann. Oncol.	1842-1847	12	17
Risch	1988	Canada	Occupational factors and the incidence of cancer of the bladder in Canada	Br J Ind Med	361-367	6	45
Rodelsperger	2001	Germany	Asbestos and man-made vitreous fibers as risk factors for diffuse malignant mesothelioma: results from a German hospital-based case-control study	Am. J. Ind. Med.	262-275	3	39
Rodvall	1996	Sweden	Glioma and occupational exposure in Sweden, a case-control study	Occup Environ Med	526-532	8	53
Rousseau	2007	Canada	Occupational exposure to lead compounds and risk of cancer among men: a population-based case-control study	Am. J. Epidemiol.	1005-1014	9	166

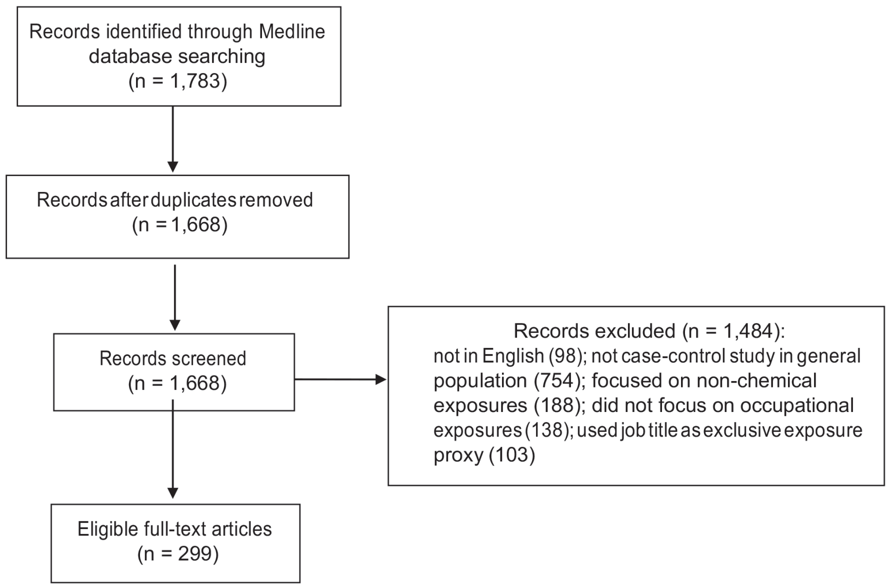
Ruder	2004	US	Gliomas and farm pesticide exposure in men: the upper midwest health study	Arch. Environ. Health	650-657	12	59
Ruder	2013	US	The Upper Midwest Health Study: gliomas and occupational exposure to chlorinated solvents	Occup Environ Med	73-80	2	70
Russi	1997	US	Occupational exposure to machining fluids and laryngeal cancer risk: contrasting results using two separate control groups	Am. J. Ind. Med.	166-171	2	31
Rybicki	2006	US	Prostate cancer risk from occupational exposure to polycyclic aromatic hydrocarbons interacting with the GSTP1 Ile105Val polymorphism	Cancer Detect. Prev.	412-422	5	30
Sadetzki	2000	Israel	Selected risk factors for transitional cell bladder cancer	Med. Oncol.	179-182	3	17
Samanic	2006	Spain	Smoking and bladder cancer in Spain: effects of tobacco type, timing, environmental tobacco smoke, and gender	Cancer Epidemiol. Biomarkers Prev.	1348-1354	7	15
Sasco	2002	Morocco	A case-control study of lung cancer in Casablanca, Morocco	Cancer Causes Control	609-616	7	13
Schlehofer	1995	Germany	Occupation, smoking and demographic factors, and renal cell carcinoma in Germany	Int J Epidemiol	51-57	1	24
Schlehofer	2005	International	Occupational risk factors for low grade and high grade glioma: results from an international case control study of adult brain tumours	Int. J. Cancer	116-125	1	113
Schmeisser	2010	EU	Occupational exposure to pesticides and bile tract carcinoma in men: results from a European multicenter case-control study	Cancer Causes Control	1493-1502	9	21

Schmidt-Pokrzywniak	2010	Germany	A case-control study: occupational cooking and the risk of uveal melanoma	BMC Ophthalmol	26		10
Schoenberg	1987	US	Occupation and lung cancer risk among New Jersey white males	J. Natl. Cancer Inst.	13-21	1	79
Schoenberg	1984	US	Case-control study of bladder cancer in New Jersey. I. Occupational exposures in white males	J. Natl. Cancer Inst.	973-981	5	72
Seidler	2010	Germany, Italy	Asbestos exposure and malignant lymphoma: a multicenter case-control study in Germany and Italy	Int Arch Occup Environ Health	563-570	5	83
Siemiatycki	1989	Canada	Cancer risks associated with 10 inorganic dusts: results from a case-control study in Montreal	Am. J. Ind. Med.	547-567	5	16
Siemiatycki	1994	Canada	Occupational risk factors for bladder cancer: results from a case-control study in Montreal, Quebec, Canada	Am. J. Epidemiol.	1061-1080	12	140
Siemiatycki	1997	Canada	Reliability of an expert rating procedure for retrospective assessment of occupational exposures in community-based case-control studies	Am. J. Ind. Med.	280-286	3	31
Simpson	1998	US	Wood-dust exposures and cancer of the colon	Int J Occup Environ Health	179-183	3	4
Smith	1992	Australia	Phenoxy herbicides and chlorophenols: a case control study on soft tissue sarcoma and malignant lymphoma	Br. J. Cancer	442-448	3	65
Soskolne	2011	Canada	A population-based case-control study of occupational exposure to acids and the risk of lung cancer: evidence for specificity of association	Int J Occup Environ Health	1-8	1	17

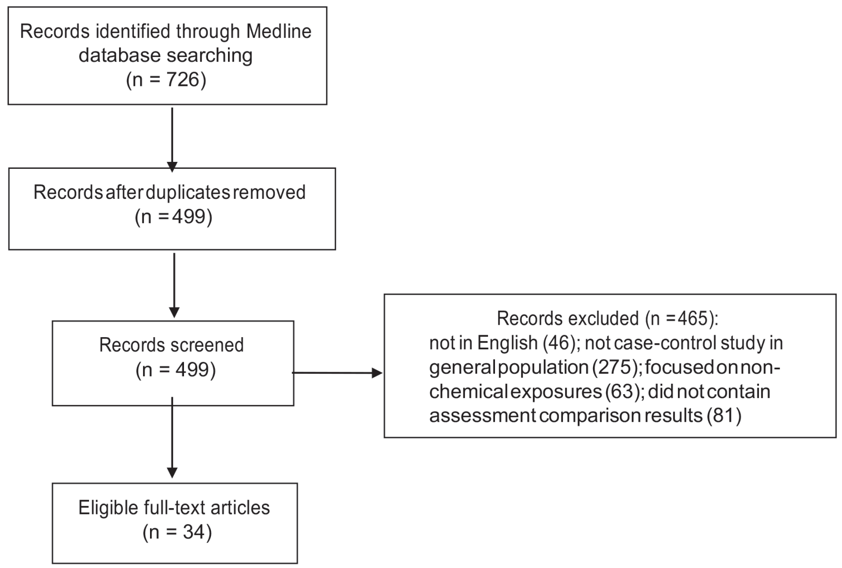
Spinelli	2010	France	Occupational and environmental risk factors for brain cancer: a pilot case-control study in France	Presse Med	e35-44	2	39
Steineck	1990	Sweden	Increased risk of urothelial cancer in Stockholm during 1985-87 after exposure to benzene and exhausts	Int. J. Cancer	1012-1017	6	45
Stengel	1993	France	Retrospective evaluation of occupational exposure to organic solvents: questionnaire and job exposure matrix	Int J Epidemiol	S72-82		22 Suppl 2
Strom	2008	US	Prostate cancer in Mexican-Americans: identification of risk factors	Prostate	563-570	5	68
Swerdlow	1991	England	Cancer of the testis, socioeconomic status, and occupation	Br J Ind Med	670-674	10	48
't Mannetje	2011	EU	Occupational exposure to metal compounds and lung cancer. Results from a multi-center case-control study in Central/Eastern Europe and UK	Cancer Causes Control	1669-1680	12	22
't Mannetje	2012	EU	Welding and lung cancer in Central and Eastern Europe and the United Kingdom	Am. J. Epidemiol.	706-714	7	175
Talibov	2014	Nordic	Occupational exposure to solvents and acute myeloid leukemia: a population-based, case-control study in four Nordic countries	Scand J Work Environ Health	511-517	5	40
Tatham	1997	US	Occupational risk factors for subgroups of non-Hodgkin's lymphoma	Epidemiology	551-558	5	8
Theis	2008	US	Smoking, environmental tobacco smoke, and risk of renal cell cancer: a population-based case-control study	BMC Cancer	387		8
Tinnerberg	2001	Sweden	Evaluation of occupational and leisure time exposure assessment in a population-based case control study on leukaemia	Int Arch Occup Environ Health	533-540	8	74
Tinnerberg	2003	EU	Retrospective exposure assessment and quality control in an international multi-centre case-control study	Ann Occup Hyg	37-47	1	47
Tranah	2009	US	Solvent exposure and non-Hodgkin lymphoma: no risk in a population-based study in the San Francisco Bay Area	Cancer Epidemiol. Biomarkers Prev.	3130-3132	11	18
Tsuda	2001	Japan	A case-control study of the relationships among silica exposure, gastric cancer, and esophageal cancer	Am. J. Ind. Med.	52-57	1	39
Ugnat	2004	Canada	Occupational exposure to chemical and petrochemical industries and bladder cancer risk in four western Canadian provinces	Chronic Dis Can	7-15	2	25
Vajdic	2007	Australia	Atopy, exposure to pesticides and risk of non-Hodgkin lymphoma	Int. J. Cancer	2271-2274	10	120
Vallieres	2012	Canada	Exposure to welding fumes increases lung cancer risk among light smokers but not among heavy smokers: evidence from two case-control studies in Montreal	Cancer Med	47-58	1	1
Vallieres	2015	Canada	Occupational exposure to wood dust and risk of lung cancer in two population-based case-control studies in Montreal, Canada	Environ Health	1		14

van Loon	1997	Netherlands	Occupational exposure to carcinogens and risk of lung cancer: results from The Netherlands cohort study	Occup Environ Med	817-824	11	54
Vaughan	1997	US	Work in dry cleaning and the incidence of cancer of the oral cavity, larynx, and oesophagus	Occup Environ Med	692-695	9	54
Vaughan	2000	US	Occupational exposure to formaldehyde and wood dust and nasopharyngeal carcinoma	Occup Environ Med	376-384	6	57
Vaughan	1986	US	Formaldehyde and cancers of the pharynx, sinus and nasal cavity: I. Occupational exposures	Int. J. Cancer	677-683	5	38
Vida	2010	Canada	Occupational exposure to silica and lung cancer: pooled analysis of two case-control studies in Montreal, Canada	Cancer Epidemiol. Biomarkers Prev.	1602-1611	6	19
Villeneuve	2004	Canada	Environmental tobacco smoke and the risk of pancreatic cancer: findings from a Canadian population-based case-control study	Can J Public Health	32-37	1	95
Villeneuve	2012	Canada	Occupational exposure to asbestos and lung cancer in men: evidence from a population-based case-control study in eight Canadian provinces	BMC Cancer	595		12
Villeneuve	2011	Canada	Occupational exposure to diesel and gasoline emissions and lung cancer in Canadian men	Environ. Res.	727-735	5	111
Vineis	1985	Italy	Occupation and bladder cancer in males: a case-control study	Int. J. Cancer	599-606	5	35
Vizcaya	2013	Canada	Risk of lung cancer associated with six types of chlorinated solvents: results from two case-control studies in Montreal, Canada	Occup Environ Med	81-85	2	70
Wang	2009	US	Occupational exposure to solvents and risk of non-Hodgkin lymphoma in Connecticut women	Am. J. Epidemiol.	176-185	2	169
Wheeler	2013	US	Inside the black box: starting to uncover the underlying decision rules used in a one-by-one expert assessment of occupational exposure in case-control studies	Occup Environ Med	203-210	3	70
Wild	2012	France	Occupational risk factors have to be considered in the definition of high-risk lung cancer populations		1346-1352	7	106
Wild	2016	France	The 2-phase case-control design: an efficient way to use expert- time	Scand J Work Environ Health			
Wilson	2004	US	Occupational exposures and salivary gland cancer mortality among African American and white workers in the United States	J. Occup. Environ. Med.	287-297	3	46
Wortley	1992	US	A case-control study of occupational risk factors for laryngeal cancer	Br J Ind Med	837-844	12	49
Wu	1995	US	A case-control study of wood dust exposure, mutagen sensitivity, and lung cancer risk	Cancer Epidemiol. Biomarkers Prev.	583-588	6	4
Wynant	2013	Canada	Occupational exposure to lead and lung cancer: results from two case-control studies in Montreal, Canada	Occup Environ Med	164-170	3	70
Yiin	2012	US	The Upper Midwest Health Study: a case-control study of pesticide applicators and risk of glioma	Environ Health	39		11
Young	2005	US	Triazine herbicides and epithelial ovarian cancer risk in central California	J. Occup. Environ. Med.	1148-1156	11	47

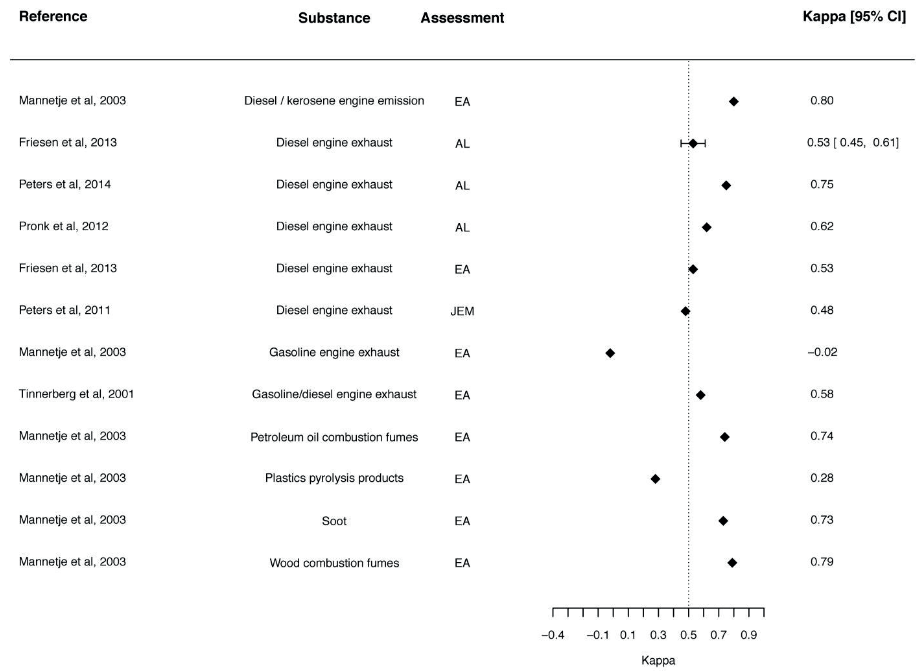
Yu	1990	China	Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China	Int. J. Cancer	1033-1039	6	45
Zahm	1993	US	The role of agricultural pesticide use in the development of non-Hodgkin's lymphoma in women	Arch. Environ. Health	353-358	5	48
Zheng	2001	China	Agricultural exposure to carbamate pesticides and risk of non-Hodgkin lymphoma	J. Occup. Environ. Med.	641-649	7	43
Zheng	1992	China	A population-based case-control study of cancers of the nasal cavity and paranasal sinuses in Shanghai	Int. J. Cancer	557-561	4	52
Zheng	1992	China	Diet and other risk factors for laryngeal cancer in Shanghai, China	Am. J. Epidemiol.	178-191	2	136
Zheng	1996	US	Diet and other risk factors for cancer of the salivary glands:a population-based case-control study	Int. J. Cancer	194-198	2	67
Zhong	1999	China	A case-control study of lung cancer and environmental tobacco smoke among nonsmoking women living in Shanghai, China	Cancer Causes Control	607-616	6	10
Zhu	2002	US	Case-control study evaluating the homogeneity and heterogeneity of risk factors between sinonasal and nasopharyngeal cancers	Int. J. Cancer	119-123	1	99



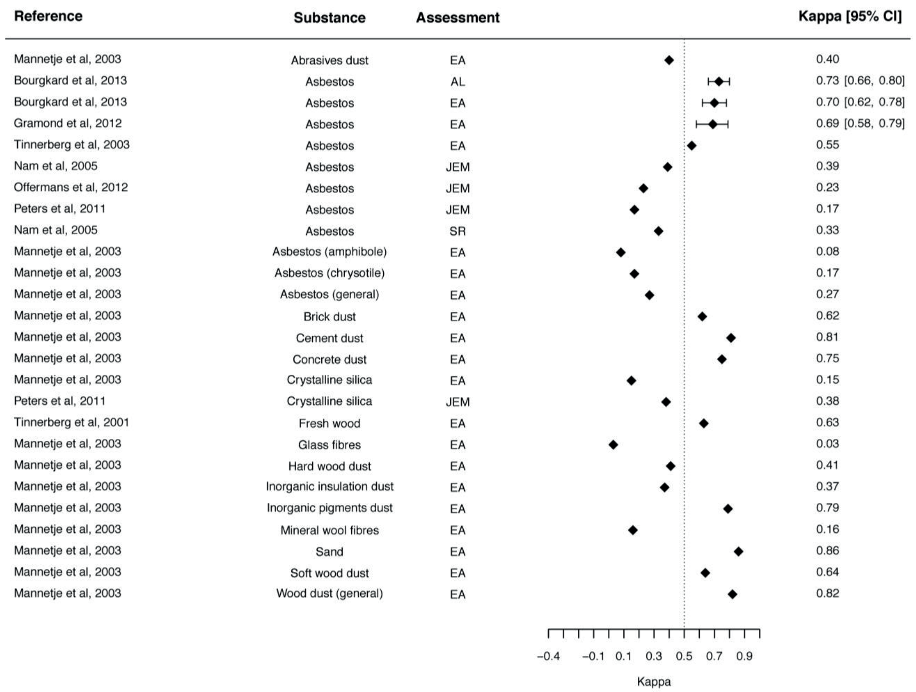
Supplementary Figure 1: Prisma diagram for systematic review of occupational cancer case-control studies of chemical agents in the general population



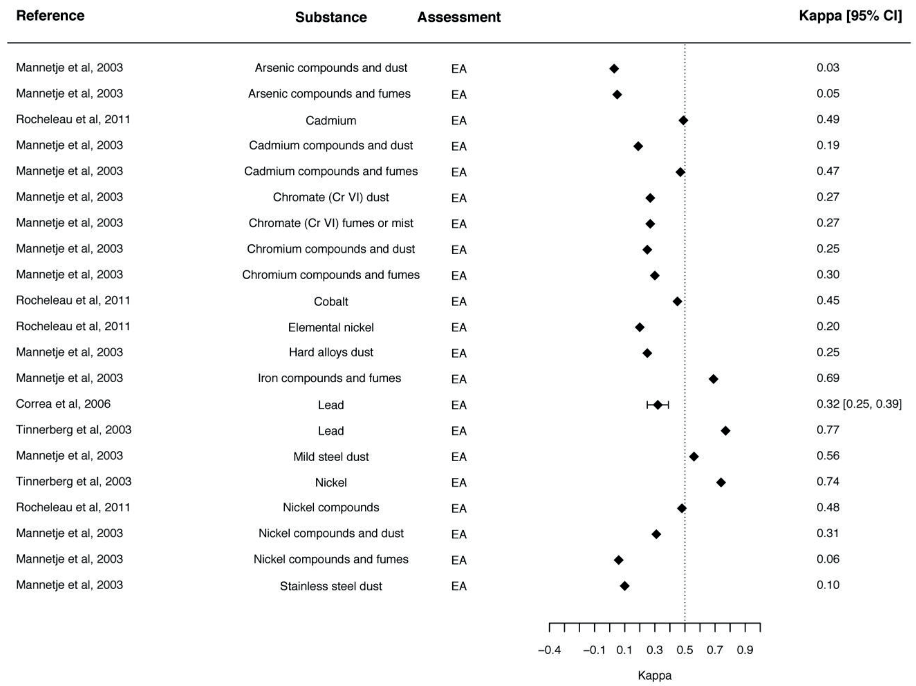
Supplementary Figure 2: Prisma diagram for systematic review of retrospective occupational exposure assessment reliability comparison studies, where assessment was performed for chemicals in case-control studies in the general population



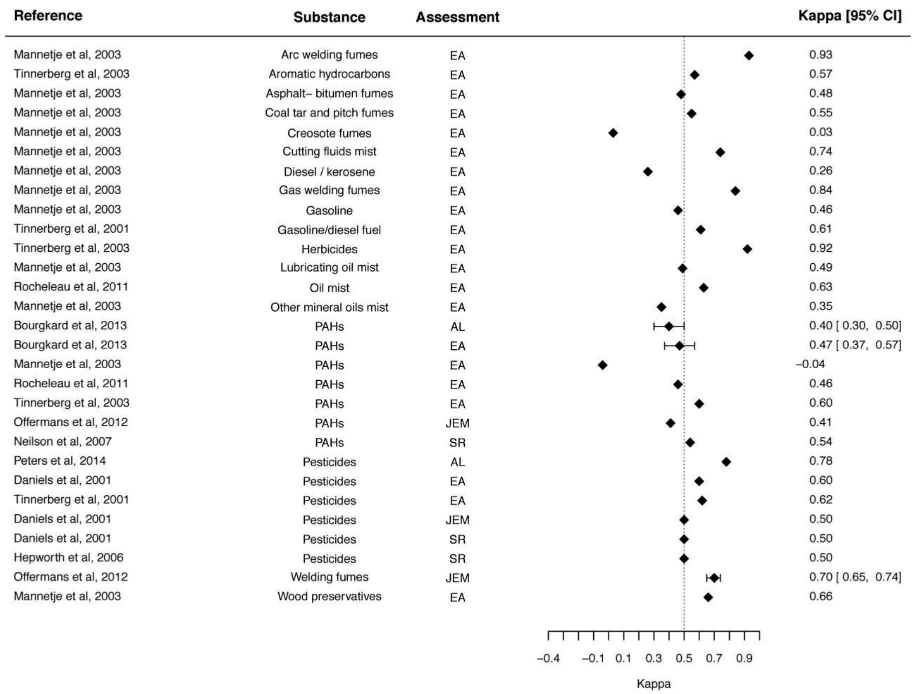
Supplementary Figure 3: Forest plot of unweighted kappa values and 95% confidence intervals (95% CI) by substance (combustion products) and assessment method (AL = algorithmic assessment; EA = case-by-case expert assessment; JEM = job-exposure matrix) with case-by-case expert assessment as comparison reference



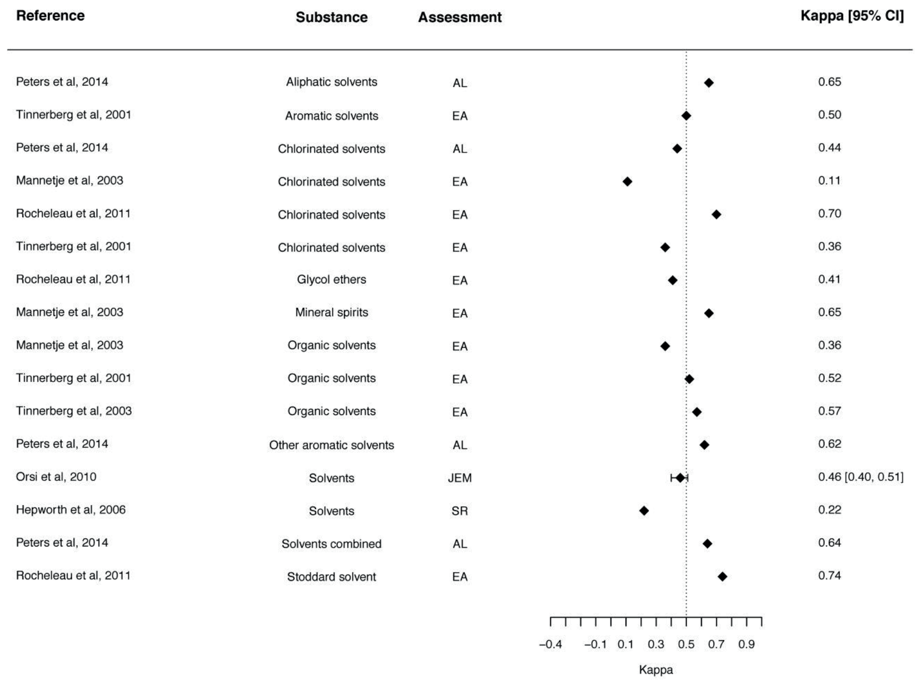
Supplementary Figure 4: Forest plot of unweighted kappa values and 95% confidence intervals (95% CI) by substance (fibres and dusts) and assessment method (AL = algorithmic assessment; EA = case-by-case expert assessment; JEM = job-exposure matrix; SR = self- report) with case-by-case expert assessment as comparison reference



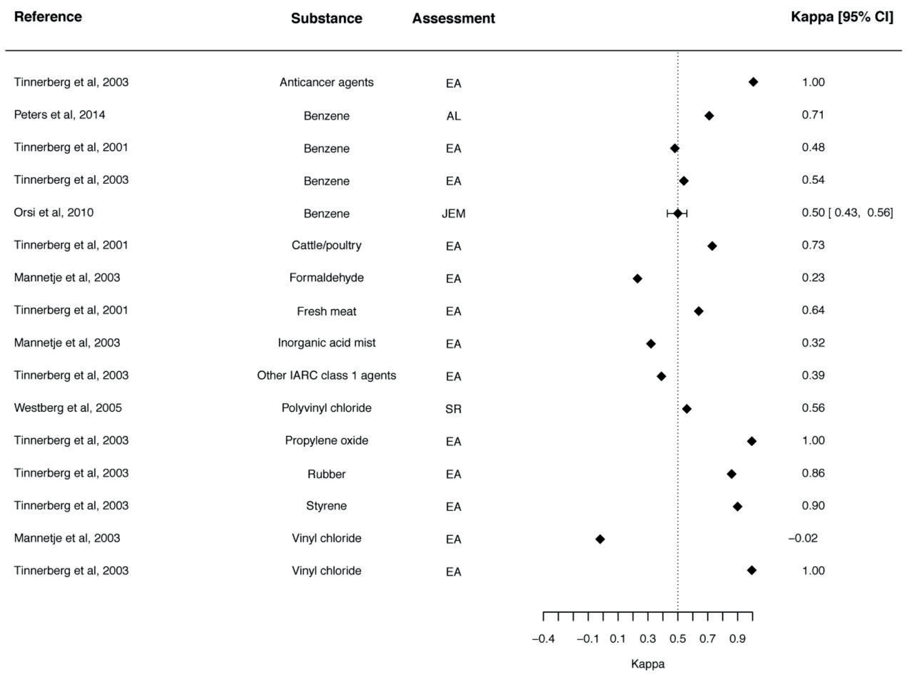
Supplementary Figure 5: Forest plot of unweighted kappa values and 95% confidence intervals (95% CI) by substance (metals) and assessment method (EA = case-by-case expert assessment) with case-by-case expert assessment as comparison reference



Supplementary 6: Forest plot of unweighted kappa values and 95% confidence intervals (95% CI) by substance (mixtures) and assessment method (EA = case-by-case expert assessment; JEM = job-exposure matrix; SR = self-report) with case-by-case expert assessment as comparison reference



Supplementary Figure 7: Forest plot of unweighted kappa values and 95% confidence intervals (95% CI) by substance (solvents) and assessment method (AL = algorithmic assessment; EA = case-by-case expert assessment; JEM = job-exposure matrix; SR = self- report) with case-by-case expert assessment as comparison reference



Supplementary Figure 8: Forest plot of unweighted kappa values and 95% confidence intervals (95% CI) by substance (others) and assessment method (AL = algorithmic assessment; EA = case-by-case expert assessment; JEM = job-exposure matrix; SR = self- report) with case-by-case expert assessment as comparison reference

CHAPTER 3

Evaluation of Automatically Assigned Job-Specific Interview Modules

M Friesen[1*], and Q Lan[1†], C Ge[2], S Locke[1], D Hosgood[3], L Fritschi[4], T Sadkowsky[5], YC Chen[1,6], H Wei[1], J Xu[1,7], TH Lam[7], YL Kwong[8,9], K Chen[10], C Xu[11], YC Su[12,13], B Chiu[14], KM Ip[7], M Purdue[1], B Bassig[1], N Rothman[1‡] and R Vermeulen[2]

1. Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, North Bethesda, MD 20980, USA;
2. University of Utrecht, Utrecht, The Netherlands;
3. Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA;
4. School of Public Health, Curtin University, Perth, Australia;
5. Data Scientists Pty Ltd, Sunshine Coast, Queensland, Australia;
6. Environmental Health Research Center, National Health Research Institutes, Zhunan, Taiwan;
7. Division of Community Medicine and Public Health Practice, School of Public Health, The University of Hong Kong, Hong Kong;
8. Bone Marrow Transplant Unit, Queen Mary Hospital, Hong Kong;
9. Division of Haematology, Oncology and Bone Marrow Transplantation, Department of Medicine, The University of Hong Kong, Hong Kong;
10. Department of Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China;
11. Department of Hematology, Hematology Research Laboratory and Pathology, West China Hospital of Sichuan University, Chengdu, Sichuan,China;
12. Division of Hematology-Oncology, Department of Internal Medicine, Buddhist Dalin Tzu Chi General Hospital, Chiayi, Taiwan;
13. School of Medicine, Tzu Chi University, Hualian, Taiwan;
14. Department of Public Health Sciences, University of Chicago, Chicago, IL, USA

†Co-first author.
‡Co-senior author.

Published: *Annals of Occupational Hygiene,* 2016, Vol. 60, No. 7, 885–899.
doi: doi:10.1093/annhyg/mew029

ABSTRACT

Objective: In community-based epidemiological studies, job- and industry-specific 'modules' are often used to systematically obtain details about the subject's work tasks. The module assignment is often made by the interviewer, who may have insufficient occupational hygiene knowledge to assign the correct module. We evaluated, in the context of a case–control study of lymphoid neoplasms in Asia ('AsiaLymph'), the performance of an algorithm that provided automatic, real-time module assignment during a computer-assisted personal interview.

Methods: AsiaLymph's occupational component began with a lifetime occupational history questionnaire with free-text responses and three solvent exposure screening questions. To assign each job to one of 23 study-specific modules, an algorithm automatically searched the free-text responses to the questions 'job title' and 'product made or services provided by employer' using a list of module-specific keywords, comprising over 5800 keywords in English, Traditional and Simplified Chinese. Hierarchical decision rules were used when the keyword match triggered multiple modules. If no keyword match was identified, a generic solvent module was assigned if the subject responded 'yes' to any of the three solvent screening questions. If these question responses were all 'no', a work location module was assigned, which redirected the subject to the farming, teaching, health professional, solvent, or industry solvent modules or ended the questions for that job, depending on the location response. We conducted a reliability assessment that compared the algorithm-assigned modules to consensus module assignments made by two industrial hygienists for a subset of 1251 (of 11 409) jobs selected using a stratified random selection procedure using module-specific strata. Discordant assignments between the algorithm and consensus assignments (483 jobs) were qualitatively reviewed by the hygienists to evaluate the potential information lost from missed questions with using the algorithm-assigned module (none, low, medium, high).

Results: The most frequently assigned modules were the work location (33%), solvent (20%), farming and food industry (19%), and dry cleaning and textile industry (6.4%) modules. In the reliability subset, the algorithm assignment had an exact match to the expert consensus-assigned module for 722 (57.7%) of the 1251 jobs. Overall, adjusted for the proportion of jobs in each stratum, we estimated that 86% of the algorithm-assigned modules would result in no information loss, 2% would have low information loss, and 12% would have medium to high information loss. Medium to high information loss occurred for <10% of the jobs assigned the generic solvent module and for 21, 32, and 31% of the jobs assigned the work location module with location responses of 'someplace else', 'factory', and 'don't know', respectively. Other work location responses had ≤8% with medium to high information loss because of redirections to other modules. Medium to high information loss occurred more frequently when a job description matched with multiple keywords pointing to different modules (29–69%, depending on the triggered assignment rule).

Conclusions: These evaluations demonstrated that automatically assigned modules can reliably reproduce an expert's module assignment without the direct involvement of an industrial hygienist or interviewer. The feasibility of adapting this framework to other studies will be language- and exposure-specific.

INTRODUCTION

Case–control studies that aim to evaluate health effects related to occupational risk factors typically collect occupational information using a lifetime occupational history questionnaire and, in many studies, supplemental occupation- and industry-specific job 'modules' that ask additional exposure-oriented questions. For example, a participant who reported having worked as a textile worker would be asked additional questions related to the textile industry using a textile module and a welder would be asked additional welding-related questions using a welder module. Use of these modules can reduce exposure misclassification by capturing important within-job exposure differences that occur both between- and within-subjects across time that would not be captured using occupation alone (Gérin et al., 1985; P. A. Stewart et al., 1998). The module responses can then be used to develop exposure decision rules to efficiently and transparently obtain exposure estimates for the study participants (Behrens et al., 2012; Carey et al., 2014; Friesen et al., 2013; Fritschi et al., 2009; Peters et al., 2014; Pronk et al., 2012).

The assignment of the appropriate module can be challenging during interviews because many interviewers have insufficient occupational hygiene knowledge with which to choose the appropriate module. As a result, use of these modules may be a two-step process: first, the participant is interviewed to obtain general occupational information, such as job title and task; second, after an occupational hygienist reviews the first interview's occupation information, the participant is re-interviewed with expert-assigned modules targeted to the reported jobs (e.g. Fritschi et al., 2005; Gérin et al., 1985; Macfarlane et al., 2012; P. A. Stewart et al., 1998; Patricia A. Stewart et al., 1996). To remove the burden of re-contacting and re-interviewing the subject, some studies have provided training to the interviewer to select the most appropriate module based on the occupational history information (e.g. Carey et al., 2014). In some cases, the interviewer's selection has been aided by searching the responses entered during a computer-assisted interview with expert-derived keyword lists based on job title and task to provide the interviewer with a real-time, short list of modules from which to choose (e.g. Colt et al., 2011). However, the error rate from interviewer module assignment is unknown and the training required to select the modules may be impractical in studies involving multiple sites and using a large team of interviewers.

To facilitate module assignment during a computer-assisted personal interview, we developed a computerized algorithm—NCI OccMATES: Occupational Modules Automatically Triggered in Epidemiologic Studies—that used free-text questionnaire responses to provide an automated, real-time assignment of each job to one of 23 modules. NCI OccMATES was implemented on study tablet computers to search the free-text entry of responses to a lifetime occupational history questionnaire against extensive lists of over 5800 module-specific keywords to identify keyword matches. Based on the keyword match(es), a single module was assigned for each job using a set of expert-derived hierarchical decision rules. This module was incorporated into the interviews immediately following the lifetime occupational history questions using OccIDEAS, a software application that provided the framework for storage and delivery of our exposure-oriented modules (Fritschi et al., 2009). This occupa-tional data collection and algorithm-module assignment were conducted within the Multi-Center Study of Lymphoid Neoplasms in Asia (hereafter, 'AsiaLymph'), a hospital-based case–control study that enrolled cases and controls in four study centers: Hong Kong, Chengdu, and Tianjin, China, and Kaohsiung, Taiwan. In this article, we describe the algorithm and the results of a reliability assessment that compared the algorithm-assigned modules to those assigned by two industrial hygienists using English translations of participants' responses to the occupational history interview.

METHODS

Occupational questionnaires

The occupational questionnaires used in AsiaLymph comprised a lifetime occupational history questionnaire and 23 modules with exposure-oriented questions focused on potential solvent exposure, including benzene, trichloroethylene, and formaldehyde. All questions were developed in English (M.C.F., S.J.L., R.V.) and translated into Traditional and Simplified Chinese (Y.C.C., H.W., J.X.). The translations of all occupational questions were reviewed by industrial hygienists from each study center to ensure location-specific nomenclature was incorporated. English-language modules are provided in the supplement; Chinese language versions are available from the corresponding authors.

The occupational history questionnaire comprised, for each job reported by the subject, open-ended questions on 'what was the name of the employer or workplace', 'what was your job title', 'what did the employer make, or what service did they provide', and 'what were your main activities or duties', as well as questions on job start and stop years, and days per week and months per year worked in each job. In addition, three solvent exposure screening questions were asked:

- In this job, did you ever use paints, stains or varnishes or work in an area where they were used?
- In this job, did you ever use solvents, glues, degreasing agents (to clean metal parts), gasoline or other fuels, or work in an area where they were used?
- In this job, did you ever use particle board, plywood, or veneered woods or work in an area where they were used?

The exposure-oriented modules comprised 20 modules focused on specific occupations and industries (e.g. chemist module, healthcare module), two solvent modules that captured information on solvent-related tasks (solvent module, industry solvent module), and one work location module (Table 1). The modules were adapted (by M.C.F., S.J.L., R.V.) from previously used modules in NCI case–control studies and incorporated additional knowledge obtained from previous studies conducted in China.

The solvent and industry solvent modules asked solvent task-related questions on degreasing, painting, paint stripping, gluing, fueling, hand contact with solvents, and working with particle board. The solvent module asked all of the solvent task-related questions, whereas in the industry solvent module the tasks that were queried depended on the occupation (e.g. administrative, production, management, quality control and engineering, maintenance, and material handling). For example, if the occupation was 'administrative or management', no solvent task questions were asked; if the occupation was 'material handling', only the paint, glue, board, and fuel questions were asked; and if the occupation was 'quality control, engineers, or other technical positions', questions on the collection and testing of production line samples were asked. These solvent task-related questions were also asked within many of the occupation- and industry-specific modules, which are identified as cross-module questions in Table 1 (for more detail see Supplementary Table 1). In particular, these cross-module solvent task-related questions were triggered when the subject indicated that they did maintenance or utility work or were involved in shipping, receiving, or storage work.

The work location module was assigned whenever a job was not assigned to any of the occupation-, industry-, or solvent-specific modules, based on the algorithm described in the next section, to redirect jobs in specific work locations to appropriate modules. The module stated: 'The computer has not been able to accurately assess what kind of job you reported. We would like to verify if the current job is mostly: [participant asked to identify the best fit from 8 categorical work locations]'. Jobs with work locations of farm, hospital, school, factory, or construction site work locations were redirected to farming, health professional, teaching, industry solvent, or solvent modules, respectively. If the work was in any other location (e.g. office, store, restaurant, someplace else) no more questions were asked for that job.

Keyword development

We developed lists of over 5800 occupation and industry keywords in English (n = 1580), traditional Chinese (Hong Kong, Kaohsiung, n = 2422) and simplified Chinese (Chengdu, Tianjin, n = 1892) that were specific to one of the 23 modules. One set of occupational keywords was developed to search the subject's responses to the questions 'job title' to identify keywords linked to occupations associated with each module (21 occupation sets). For example 'seamstress' was a keyword linked to dry cleaning and textile industry occupations. Another set of industrial keywords was developed to search the responses to 'product made or services provided by employer' to identify keywords linked to industries associated with each module (17 industry sets). For example 'smelt- ing' was a keyword linked to the foundry industry.

All keyword sets were first developed in English (M.C.F., S.J.L., R.V.). This team supervised the trans- lation of the English-language lists by three native Chinese speakers (Y.C.C., H.W., J.X.) into the local languages of the four study centers (simplified Chinese for Chengdu and Tianjin; traditional Chinese for Kaohsiung and Hong Kong). The translation process included generation of additional, similar local words and phrases that may be used to describe job title or employer activity. The translations were then reviewed by occupational hygienists from each of the four study centers, with additional words and revisions incorporated by the develop- ment team.

Table 1. List of study modules, including an overview of within-module redirections to other modules and cross-module questions. See Supplementary Table 1 for more detail on what subsets of responses result in these redirections and cross-module questions.

Module, abbreviation	Includes redirection to	Cross-module questions												Proportion of all jobs (%)
		Degrease	Paint	Strip Paint	Glue	Particle board	Handle fuel	Hands	Stain	Job type	QC sample collection	Molding plastic	Pesticide use	
Occupation & Industry-specific modules														
Chemist, CH								x			x			0.9%
Chemical industry, CHI		x	x	x	x	x	x	x		x	x			2.3%
Dry cleaning & textile industries, DLI_TXI	LEI	x	x	x	x	x	x	x		x	x			6.4%
Embalming, EM														<0.1%
Foundry industry, FOI		x	x	x	x	x	x	x		x	x			2.7%
Furniture industry, FUI		x	x	x	x	x	x	x	x			x		1.3%
Farming and food industry, GF_FDI		x	x	x	x	x	x	x		x			x	18.8%
Health professional, HP		x	x	x	x	x	x	x						2.6%
Janitor, JA		x	x		x		x	x						2.2%
Leather industry, LEI		x	x	x	x	x	x	x		x		x		0.5%
Lumber industry, LUI	FUI	x	x	x	x	x	x	x	x					0.9%
Oil refinery industry, ORI		x	x	x	x	x	x	x		x	x			0.9%
Pesticide applicators, PE								x					x	<0.1%
Photographic industry, PHI		x	x	x	x	x	x	x		x	x			0.2%
Plastic industry, PLI		x	x	x	x	x	x	x		x	x	x		1.2%
Pulp and paper industry, PPI		x	x	x	x	x	x	x		x	x			0.4%
Printing industry, PRI		x	x	x	x	x	x	x		x	x			1.0%
Rubber industry, RUI		x	x	x	x	x	x	x		x	x			0.2%
Shoe industry, SHI		x	x	x	x	x	x	x		x				0.8%
Teaching, TE		x	x					x						3.4%

Table 1. Continued

Generic modules														
Solvent, SOL	GF[b]	x	x	x	x	x	x	x	x					21.5%
Industrysolvent,INDSOL	GF[b]	x	x	x	x	x	x	x		x	x			0.3%
Work location, BUP	GF[b], HP, INDSOL, SOL, TE													32.7%

[a]Responses to certain questions redirected that job to questions in other modules. For example textile workers working with leather were asked questions from the leather industry module.
[b]Gardening and farming-related questions from the GF_FDI module.

Each keyword set contained three types of keywords that were used alone because the word or word string uniquely identified relevant occupation or industry (Type 1) or were used in combination with another keyword to identify the relevant occupation or industry (Types 2 and 3). Examples of Type 1 keywords were 'teacher' and 'dry cleaner', which were considered sufficient information with which to assign the 'Teacher' module and 'Dry Cleaning and Textile Industry' module, respectively. Type 2 and 3 keywords were designed because not all processes and activities could be described succinctly or completely by specific words or word strings or because the word order may vary. In general, Type 2 keywords described the material or service being processed or provided (e.g. clothes, dry-cleaning), and Type 3 keywords described the process, place or person related to the product or service (e.g. bleaching, workshop, worker). A Type 2 or Type 3 keyword could each appear in multiple keyword sets to capture similar work activities (e.g. Type 3 keyword 'worker'). However, unique combinations of Type 2 and Type 3 keywords were designed to occur only within one keyword set. To match a keyword set, the response had to include a Type 1 match or both a Type 2 and a Type 3 match, in any order (hereafter, Type 2/3 match). For example, the responses 'bleaching clothes' and 'dry-cleaning worker' would both have a positive match to the occupation keyword set for the 'Dry Cleaning and Textile Industry' module.

Algorithm description

The occupational and industry keyword sets were used to search the occupational history responses to assign the most appropriate module. All jobs received a module. The occupational keyword sets were used to search the job title responses from the occupational history; this search either identified no matches to any keyword set, single match (Type 1 or Type 2/3), or multiple matches (of any match type). Similarly, the industrial keywords were used to search the employer activity responses to identify no match, single match, or multiple industrial keyword set matches.

The varying combinations of match results were processed using a set of hierarchical decision rules that assigned a single module for each reported job according to the flow chart shown in Figure 1. See Supplementary Table 2 for the individual rules and actions. If no module-specific keyword was identified for a job, the module was assigned based on the solvent screening questions: if 'yes' to any of these questions, the solvent module was assigned (rule #1); and if 'no' to all these questions, the work location module was assigned (rule #2). If the keyword(s) matched a single module based on the occupation *or* industry keyword sets, the corresponding module was assigned (single industry module: rule #3; single occupation module: rule #4).

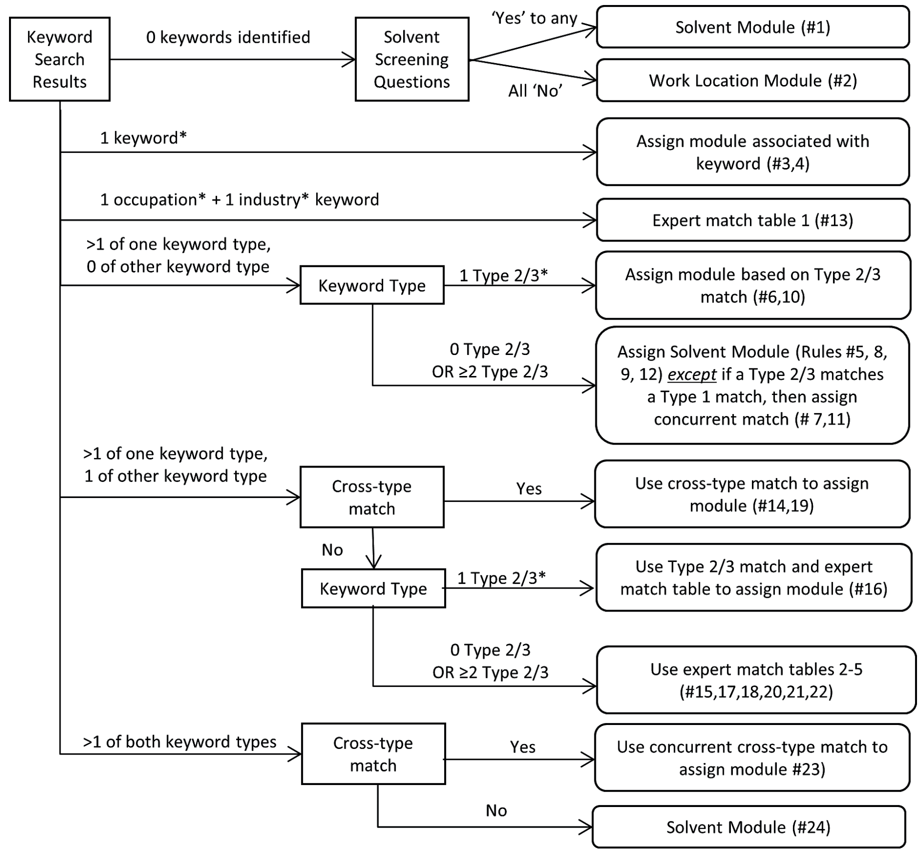


Figure 1: Hierarchical algorithm decision rules to assign modules. The rules were based on the number of keyword sets identified in the response (occupation and/or industry), the keyword type (Type 1, Type 2/3), and whether the occupation and industry keywords assigned the same or different modules ('cross-type match' = yes or no, respectively). Flowchart steps for single sets or single keyword types marked with * also include the rare scenario where multiple matches were found but all matches were within the same occupation or industry keyword set.

If multiple keywords were identified for a job, the module was assigned using expert-derived rules based on the number of keyword sets identified in the response (occupation and/or industry), the keyword type (Type 1, Type 2/3), and whether the occupation and industry keywords assigned the same or different modules ('cross-type match' = yes or no, respectively). These rules fall into three distinct groups: multiple occupation keywords but no industry keywords identified (rules #5–8), multiple industry keywords but no occupation keywords identified (rules #9–12), or combinations of both occupation and industry keywords identified (rules #13–24). Within each group, the rules designated what to do when matches to multiple occupation and/or industry sets occurred. When matches to multiple occupation or industry keyword sets occurred, the rules prioritized the cross-type module match (rules #14, 19, 23). For example, if multiple occupation keyword matches were found, the rules prioritized the match that assigned the same module as the industry keyword match. Otherwise, the rules prioritized Type 2/3 matches, which used detailed adjectives and were not order specific, over Type 1 matches (rules #6, 7, 10, 11, 16). When the occupation and industry matches directed to different modules (cross-type match = no), the module assignment was made using expert-derived match tables (rules #13, 15–23). In the expert match tables, the industry solvent module was assigned when the participant reported working in a plant or factory and did not identify the type of factory, or when the keywords identified multiple industries. When the match combination was deemed too ambiguous, the generic solvent module was assigned (rule #24 and within expert match tables).

Study implementation of questionnaires and algorithm

The occupational history component was administered during comprehensive interviews conducted by trained interviewers using tablet computers with study-specific software. The interviewers were trained to prompt the participants for detailed responses, which were entered into the software. The software used the algorithm to automatically search the 'job title' and 'product made or services provided by employer' responses to assign one of the 23 modules to each job reported by the participant. The study-specific software launched the assigned modules for that participant immediately after the occupational history section, using a stand-alone version of OccIDEAS that incorporated the study-specific modules (Fritschi et al., 2009).

Job selection for reliability assessment

Overall, 11 409 jobs were reported between 1 September 2013, when final versions of the modules and algorithm were in place and interviewer training was completed, and 31 January 2015, when the evaluations reported here began. To compare the algorithm-assigned module to the module that an industrial hygienist would have assigned in a reliability assessment, we selected a subset of these jobs using a stratified randomized selection (without replacement) procedure using strata defined by the algorithm-assigned module. Most modules were infrequently assigned, except the solvent and work location modules (Table 1). For each module except the solvent and work location modules, we selected all jobs if the module was assigned to ≤25 jobs and randomly selected 25 jobs if the module was assigned to 26–249 jobs, 50 jobs if the module was assigned to 250–499 jobs, and 100 jobs if the module was assigned to ≥500 jobs. For the solvent module, we randomly selected 100 jobs that were assigned the solvent module based on keyword-module linkage (55% of solvent module assignments, rules #3–24) and 100 jobs that were assigned that module because no keywords were identified but the participant responded 'yes' to any of the three solvent screening questions (44% of all solvent module assignments, rule #1). For the work location module, job selection was stratified by work location responses. We selected all jobs in the location category if the location category was reported for ≤25 jobs and randomly selected 25 jobs if the category was reported for 26–249 jobs, 50 jobs if the category was reported for 250–499 jobs, and 100 jobs if the category was reported for ≥500 jobs. In total, 1251 jobs were selected.

Expert module assignment

Two industrial hygienists (S.J.L., C.G.) independently reviewed English translations of the 1251 occupational history questionnaire responses in the reliability subset and assigned the most appropriate module to each job. S.J.L. had been involved in the original keyword and algorithm development and C.G. had no prior involvement. The experts were blind to the algorithm module assignments and to the module question responses. Jobs in which the two experts disagreed on the module assignment were re-reviewed by the same two experts to obtain a consensus assignment. Jobs where the assignments still differed were reviewed by M.C.F. (blind to the algorithm module assignment), who had been involved in all aspects of the keyword and algorithm development.

Expert review of assigned module coverage of pertinent questions

The two experts (S.J.L., C.G.) were provided with the subset of jobs where the expert consensus and algorithm assignments differed (n = 483). This subset excluded jobs where the discordance was between the solvent and the industry solvent module because these modules were nearly identical (n = 46). The experts independently reviewed the job description, the expert consensus module assignment, the algorithm module assignment, and the work location category for those assigned the work location module. Each expert considered whether the assigned module included the questions that were likely to be most relevant to that job (e.g. did they paint?). Relevant questions that were missed were considered to be a potential 'information loss' for the exposure assessment process. Each expert provided a qualitative estimate of the degree of potential exposure information lost by the assigned module from the missed questions for each job, using four subjective categories:

1. *All* pertinent questions covered. No loss of information.
2. *Most* pertinent questions covered. Low loss of information.
3. *Some* pertinent questions covered. Medium loss of information.
4. *Few* (or no) pertinent questions covered. High loss of information.

Potential information loss was more likely to occur when pertinent industry-specific questions were missed because the cross-module questions were included in most modules. For example, if the most relevant questions for a job were about painting and gluing and the assigned module included those questions, the expert would have provided a rating of '1', no loss of information. In a second example, if the expert consensus module was the furniture module and the assigned module was the solvent module, the expert would have provided a rating of '1' (no information loss) if the expert considered the most relevant questions were covered by the cross-module questions, or a rating of '3' (medium information loss) if the most relevant questions were about the participant's role in manufacturing furniture. The experts were asked to not consider the expected answer (e.g. yes or no to the paint question) in the evaluation of information loss. The impact of the information loss on exposure decisions will be evaluated in future analyses. The two experts' degree of loss ratings had very good agreement (% agreement = 76%; kappa = 0.61; weighted kappa = 0.84); thus, consensus 'degree of loss' ratings were not obtained.

Statistical analyses

We calculated the proportion of jobs in the reliability subset where the two experts agreed on the module assignments. Similarly, we calculated the proportion of jobs where the expert consensus and algorithm module assignments agreed, as raw agreement in the reliability subset and as estimated agreement extrapolated to all 11409 jobs. The extrapolated agreement was obtained by first calculating strata-specific agreement and then weighting that agreement by the proportion of jobs in those strata. For example the strata-specific agreement for the chemist module, farming and food industry module, and industry solvent module contributed to 0.9, 18.8, and 0.3% of the extrapolated estimate, respectively (Table 1). Evaluations of the proportion of jobs with varying degrees of information loss were based on the average of the two experts' ratings, categorized as follows: 1 = both experts indicated no loss; >1–2 = low loss; >2–3 = medium loss; and >3 = high loss.

RESULTS

Keyword matches

In the 11 409 jobs reported during the sample period, an average of 1.4 keyword matches per job was identified by the algorithm. The three most frequently identified industry keywords were Chinese translations of 'cultivation' (in 979 jobs), 'grain' (in 447 jobs), and 'cultivating' (in 446 jobs). The three most frequently identified occupation keywords were 'farmer' (in 750 jobs), 'field hand' (in 204 jobs), and 'repair' (in 185 jobs).

Expert agreement

The proportion of the 11 409 jobs that the algorithm assigned to each module is provided in Table 1. The most frequently assigned modules were the work location (32.7%), solvent (21.5%), farming and food industry (18.8%), and dry cleaning and textile industries (6.4%) modules. Other modules accounted for <0.1 to 3.4% of all jobs.

The two experts' independent module assignments matched for 80.3% of the 1251 jobs. Discordance between expert assignments to the solvent and industry solvent modules accounted for an additional 6.2% of the jobs. The remaining 246 jobs were re-reviewed to derive a consensus assignment; 87 of these jobs required review by a third expert.

Algorithm versus expert agreement and information loss: Overall

In comparison to the expert consensus-assigned modules, the algorithm assignment had exact matches for 722 (57.7%) of the 1251 jobs in the reliability assessment (Table 2). An additional 46 (3.7%) jobs differed because one approach assigned the solvent module and the other assigned the industry solvent module. The remaining 483 jobs (38.6%) represented disagreements between the consensus and algorithm assignments and were reviewed to identify the potential information lost. Evaluations based on potential information loss showed that the algorithm's discordant assignment resulted in no information loss for the majority of these jobs (51.1% of discordant jobs, 19.7% of jobs in reliability subset; Table 2). Low, medium, or high information losses were estimated to occur for 2.7, 6.4, and 9.8% of the jobs in the reliability assessment, respectively. Because we oversampled infrequent module assignments, extrapolation to all jobs resulted in a higher estimated proportion of exact matches (67.5% of all jobs) and lower proportions of medium and high information losses (5.3 and 6.2%, respectively) than in the reliability subset. Overall, an estimated 86.3% of the algorithm's assignments were consistent with the consensus assignment or would have no information loss (67.5 + 3.8 + 15.0). An additional 2.3% would be expected to have low information loss and 11.5% would have potentially medium to high information loss.

Table 2. Overall: agreement between algorithm- and expert-assigned modules, with an assessment of the degree of information loss when the two approaches were discordant

	Jobs in reliability study ($n = 1251$) N	%[a]	Extrapolated proportion (%) of all jobs[b]
Exact match	722	57.7	67.5
No match			
Solvent/industry solvent mismatch	46	3.7	3.8
No information loss	247	19.7	15.0
Low information loss	34	2.7	2.3
Medium information loss	80	6.4	5.3
High information loss	122	9.8	6.2

[a]Proportions are unadjusted for sampling weights.
[b]Extrapolation to all jobs was calculated by weighting each stratum-specific agreement by the proportion of all jobs observed in that stratum. For example the strata-specific agreement for the Chemist Module, Farming and Food Industry Module, and Industry Solvent Module accounted for 0.9, 18.8, and 0.3% of the extrapolated estimate, respectively. Proportions of all jobs each stratum represented are shown in Table 1 for most modules and Table 4 for the Solvent and Work Location Modules.

Algorithm versus expert agreement and information loss: By algorithm rule and module

Evaluations stratified based on groups of algorithm rules that applied to each assignment are shown in Table 3. Medium or high potential information losses were ≤15% when no keywords were identified (rule#1, #2) and when only a single occupation and/or industry keyword was identified (rule #3, #4, #13). These scenarios occurred for 94% of all jobs. The algorithm identified multiple matches within a keyword type for the remaining 6% of jobs. For these multiple matches, medium or high information losses ranged from 24 to 65% (rules #5–12, 14–24).

Evaluations conducted within the solvent, industry solvent, and work location modules are shown in Table 4. For jobs assigned the solvent module (20.5% of all jobs), medium or high potential information losses were observed for 7% of the jobs when the module was triggered by the exposure screening questions and 10% when triggered by algorithm rules. The industry solvent module was assigned only 0.3% of the time, but its assignment was estimated to have medium or high information loss for 48% of the jobs. For jobs receiving the work location module (33% of all jobs), medium or high information losses were less than 8% for most work location categories, with three exceptions. A higher proportion with medium or high information loss was observed for locations of 'someplace else' (21%), 'factory/ware-house' (32%), and 'don't know/missing' (31%).

The two most frequently assigned industry-specific modules—farming and food industry and dry cleaning and textile industries modules—had low, medium, or high information loss in less than 2% of the jobs assigned each module (not shown). Module-specific evaluations were not reported for the other modules because of their low prevalence.

Table 3. By algorithm rule: agreement between algorithm- and expert-assigned modules, with an assessment of the degree of information loss when the two approaches were discordant.

Exposure screening question with yes response[a]	No. of occupation keywords identified	No. of industry keywords identified	Algorithm rule numbers	No. of jobs	% of all jobs	Proportion of jobs in strata (%) Exact match	No info. loss or INDSOL/SOL discordance	Low, info. loss	Medium or high info. loss
Yes	0	0	1	100	9.0	58	30	5	7
No	0	0	2	311	32.5	48	32	5	15
—	0	1	3	229	12.4	54	30	1	15
—	1	0	4	265	20.5	67	18	1	14
—	>1	0	5–8	26	1.7	8	23	4	65
—	0	>1	9–12	15	0.9	40	20	0	40
—	1	1	13	241	19.8	73	10	2	15
—	1	>1	14–18	25	1.2	36	24	4	36
—	>1	1	19–22	34	1.4	50	21	6	24
—	>1	>1	23–24	4	0.4	50	0	25	25
—	Missing	Missing	Missing	1	0.3	100	0	0	0
Overall[c]		Reliability subset		1251		58	23	3	16
		All jobs (estimated)		11 409		68	18	2	12

[a]Screening questions were not used except in rules #1 and #2. No indicates that a 'no' response was received to all three exposure screening questions. Yes indicates that at least one screening question had a 'yes' response.
[b]See Supplementary Table 1 for more detail on each rule's criteria and resulting action.
[c]See Table 2. Provided here for comparison purposes.

Table 4. Generic solvent and work location modules: agreement between algorithm- and expert- assigned modules, with an assessment of the degree of information loss when the two approaches were discordant

			Proportion of jobs in strata (%)			
Module, sub-group	**No. of jobs in reliability study**	**% of all jobs**	**Exact match**	**INSOL/SOL discordance or no info. loss**	**Low, info. loss**	**Medium or high info. loss**
Solvent (SOL), ***reason for assignment***						
Assigned based on screening questions, no keywords identified	100	9.0	58	30	5	7
Assigned based on identified keywords	100	11.5	55	30	5	10
Industry solvent(INDSOL)	25	0.3	12	36	4	48
Work location (BUP), ***response category***[a]						
Factory/warehouse	25	3.4	28	28	12	32
School	25	0.6	32	64	0	4
Store/restaurant	25	4.9	48	44	0	8
Office	25	13.9	92	0	0	8
Construction	50	2.8	14	80	2	4
Someplace else	100	6.6	70	2	7	21
Farm	21	0.2	10	86	0	5
Hospital	6	0.1	0	100	0	0
DK/missing/skipped	35	0.3	60	0	9	31

[a] Participant was asked which category best described his or her work location. No additional information was collected on the work location if the participant responded 'Someplace else'.

DISCUSSION

This study demonstrated that an expert-designed automated algorithm provided real-time module assignments during computer-assisted personal interviews that reliably reproduced post hoc module assignments made by industrial hygienists in this study. Overall, we estimated that 86% of all algorithm module assignments would result in no potential exposure information loss, 2% would have low information loss, and 12% would have medium or high information loss. Evaluations in strata based on groups of algorithm rules and the assigned module provided important insights into directions for future improvements of the algorithm. To our knowledge, no similar evaluations comparing interviewer or automated module assignments to those of occupational hygienists have been previously reported.

The two most prevalent modules—the solvent module (assigned to 20.5% of jobs) and the work location module (assigned to 33% of jobs)—generally had ≤10% of the jobs with medium to high information loss. This good performance occurred because the solvent module captured the majority of solvent-exposed tasks and the work location module redirected participants reporting work locations with potential solvent exposure (e.g. healthcare, farming) to the appropriate module. In the work location module, some refinements may reduce information loss for the 'factory' and 'someplace else' responses. Responses of 'factory' resulted in an estimated medium to high information loss for 32% of the jobs reporting that location because these responses were redirected to the industry solvent module, which does not include important industry-specific questions. This could be refined by asking an additional question about whether the factory was in any of the industries of interest for which modules were developed to redirect that job to the appropriate module. Responses of 'someplace else' resulted in medium to high information loss for 21% of the jobs reporting that location, which suggests that an important work location category may have been missed, but the closed-ended design of this response category did not allow us to explicitly evaluate which locations were missed. However, these modest potential information losses are likely overestimates, because to be assigned this module the participants had to respond 'no' to each of the three solvent screening questions and no keyword matches were identified from the occupational history responses.

Potential information loss was more prevalent when multiple keyword matches were identified in the occupational history (29–65% medium/high loss, depending on rule). Fortunately, this occurred in <6% of all jobs. The information loss generally occurred when appropriate industry-specific solvent exposure questions were not asked, whereas solvent task-related questions were generally captured within imperfect module assignments. For example the industry solvent module, with 48% of the jobs with medium/high loss, was assigned when there were multiple keyword matches pointing to different modules or when there was too little information to determine the type of industry. Module-specific evaluations also showed that some keywords thought to be specific to a module were likely not specific enough. For example translations of the keywords 'wood', 'saw', and 'planing' used for the lumber industry resulted in many other jobs working with wood being assigned that module, whereas the more relevant module might be the furniture industry for furniture workers and solvent module for construction or other trade laborers. In another example, translations of the words 'testing' and 'analysis' assigned many workers involved in quality control/quality assurance to the chemist module, whereas an industry-specific module may have been more appropriate. This latter example could be addressed with refinements to the hierarchical set of rules to prioritize the triggered industry rather than the chemist occupation.

Incorporating OccMATES to provide a real-time assignment during the personal interview had module assignment procedure was used for all participants, regardless of the study center and interviewer's occupational hygiene expertise; thus reducing interviewer module selection bias. It also reduced study and respondent burden by conducting both components of the occupational data collection concurrently. The assignment of the modules was also transparent and the approach can be updated based on these evaluations. In addition, exposure assessments for jobs assigned an imperfect module will still be performed, but will require expert review rather than the automated assignment of exposure assessment decision rules (Fritschi et al., 2009). Our findings suggest that expert review in this study should be focused on jobs with multiple matches within a keyword type and those indicating 'factory' or 'someplace else' in the work location module.

These evaluations also have several limitations that should be considered in interpreting these findings. First, the expert consensus module assignments represented only an 'alloyed' gold standard. The expert assignments were made post-interview, using job descriptions translated from Chinese to English. Thus, they were made using a more limited context than if it would have taken place during the interview, although they were made with much greater occupational expertise than the interviewer. We observed a modest degree of variability between the two experts, with only 13% of the jobs requiring a consensus review, of which 30% required a third expert to provide the final assignment. Second, because we excluded interviewer involvement in the module assignment, we were unable to obtain a measure of the error rate that would occur if an interviewer selected a module from a short list for comparison purposes. Third, to provide a conservative test of the algorithm, we provided the experts access to responses to all occupational history responses, whereas the algorithm module assignment was made solely based on two of these questions (job title, products made/services provided). For example the experts could use important information about industry that was reported in the employer's name response (which was often descriptive). Future refinements of the algorithm may include keyword searches of responses to these additional questions. For instance, we could use the job task information reported to develop keyword lists for tasks associated with each module to improve the specificity of the assignments. Finally, our categorization of potential 'information loss' considers only whether questions of interest were asked and not the likely response to the question or whether the job was likely exposed to solvents. The impact of this information loss on exposure decisions will be evaluated in future application of exposure decision rules.

In summary, this computerized algorithm provided a real-time module assignment that reliably reproduced an expert's module assignment, with limited potential information loss. Our findings are study specific. However, the framework could be extended to other studies. The degree of adaptation required will depend on the languages and exposures of interest and may require substantial changes to the modules, keyword lists (including adding common misspellings), and algorithm decision rules.

REFERENCES

Behrens, T., Mester, B., & Fritschi, L. (2012). Sharing the knowledge gained from occupational cohort studies: A call for action. *Occupational and Environmental Medicine*, *69*(6), 444–448. https://doi.org/10.1136/oemed-2011-100305

Carey, R. N., Driscoll, T. R., Peters, S., Glass, D. C., Reid, A., Benke, G., & Fritschi, L. (2014). Estimated prevalence of exposure to occupational carcinogens in Australia (2011–2012). *Occupational and Environmental Medicine*, *71*(1), 55–62. https://doi.org/10.1136/oemed-2013-101651

Colt, J. S., Karagas, M. R., Schwenn, M., Baris, D., Johnson, A., Stewart, P., Verrill, C., Moore, L. E., Lubin, J., Ward, M. H., Samanic, C., Rothman, N., Cantor, K. P., Beane Freeman, L. E., Schned, A., Cherala, S., & Silverman, D. T. (2011). Occupation and bladder cancer in a population-based case-control study in Northern New England. *Occupational and Environmental Medicine*, *68*(4), 239–249. https://doi.org/10.1136/oem.2009.052571

Friesen, M. C., Pronk, A., Wheeler, D. C., Chen, Y.-C., Locke, S. J., Zaebst, D. D., Schwenn, M., Johnson, A., Waddell, R., Baris, D., Colt, J. S., Silverman, D. T., Stewart, P. A., & Katki, H. A. (2013). Comparison of Algorithm-based Estimates of Occupational Diesel Exhaust Exposure to Those of Multiple Independent Raters in a Population-based Case–Control Study. *Annals of Occupational Hygiene*, *57*(4), 470–481. https://doi.org/10.1093/annhyg/mes082

Fritschi, L., Benke, G., Hughes, A. M., Kricker, A., Vajdic, C. M., Grulich, A., Turner, J., Milliken, S., Kaldor, J., & Armstrong, B. K. (2005). Risk of non-Hodgkin lymphoma associated with occupational exposure to solvents, metals, organic dusts and PCBs (Australia). *Cancer Causes & Control: CCC*, *16*(5), 599–607. https://doi.org/10.1007/s10552-004-7845-0

Fritschi, L., Friesen, M. C., Glass, D., Benke, G., Girschik, J., & Sadkowsky, T. (2009). OccIDEAS: Retrospective Occupational Exposure Assessment in Community-Based Studies Made Easier. *Journal of Environmental and Public Health*, *2009*, e957023. https://doi.org/10.1155/2009/957023

Gérin, M., Siemiatycki, J., Kemper, H., & Bégin, D. (1985). Obtaining occupational exposure histories in epidemiologic case-control studies. *Journal of Occupational Medicine.: Official Publication of the Industrial Medical Association*, *27*(6), 420–426.

Macfarlane, E., Benke, G., Sim, M. R., & Fritschi, L. (2012). OccIDEAS: An Innovative Tool to Assess Past Asbestos Exposure in the Australian Mesothelioma Registry. *Safety and Health at Work*, *3*(1), 71–76. https://doi.org/10.5491/SHAW.2012.3.1.71

Peters, S., Glass, D. C., Milne, E., Fritschi, L., & Aus-ALL consortium. (2014). Rule-based exposure assessment versus case-by-case expert assessment using the same information in a community-based study. *Occupational and Environmental Medicine*, *71*(3), 215–219. https://doi.org/10.1136/oemed-2013-101699

Pronk, A., Stewart, P. A., Coble, J. B., Katki, H. A., Wheeler, D. C., Colt, J. S., Baris, D., Schwenn, M., Karagas, M. R., Johnson, A., Waddell, R., Verrill, C., Cherala, S., Silverman, D. T., & Friesen, M. C. (2012). Comparison of two expert-based assessments of diesel exhaust exposure in a case–control study: Programmable decision rules versus expert review of individual jobs. *Occupational and Environmental Medicine*, *69*(10), 752–758. https://doi.org/10.1136/oemed-2011-100524

Stewart, P. A., Stewart, W. F., Siemiatycki, J., Heineman, E. F., & Dosemeci, M. (1998). Questionnaires for collecting detailed occupational information for community-based case control studies. *American Industrial Hygiene Association Journal*, *59*(1), 39–44. https://doi.org/10.1080/15428119891010325

Stewart, Patricia A., Lees, P. S., & Francis, M. (1996). Quantification of historical exposures in occupational cohort studies. *Scandinavian Journal of Work, Environment & Health*, *22*(6), 405–414. https://doi.org/10.5271/sjweh.161

Supplementary Table 1: List of study modules, including subsets of responses that direct the interview to other modules and to cross-module questions

Module	Directed to:	Cross-module questions												
		Degrease	Paint	Strip paint	Glue	Particle board	Handle fuel	Hands	Stain	Job type	QC sample collection	Molding plastic	Pesticide use	Industry specific questions
Work location, BUP														
On a farm	GF													
In a factory or warehouse	INDSOL													
In a school	TE													
In a hospital or medical care center	HP													
On a construction site or other outdoor location (include driving, fishing, and shipyard jobs)	SOL													
In a store, restaurant, or hotel	end													
In an office	end													
Someplace else	end													
Chemist, CH														
All								x			x			x
Chemical industry, CHI								x		x	x			
All								x		x				
Production work														x
Maintenance or utilities work		x	x	x	x	x	x							
Quality control, environmental control, engineering, research work											x			
Shipping, receiving, or storage work			x		x	x	x							
Administration														
Dry cleaning & textile industries, DLI_TXI														
Cleaning service for fabric and cloth	DLI													
Dry cleaning, DLI														
All								x						x
Maintenance or repair work		x	x	x	x	x	x	x						
Mill, factory workshop for yarn, textiles, clothing	TXI													
Textile, TXI										x	x			
Product made of leather, fur, or hair	LEI													
All										x				
Production work								x						x
Maintenance or utilities work		x	x	x	x	x	x	x						
Quality control, environmental control, engineering, or research work											x			
Shipping, receiving, or storage work			x		x	x	x	x						
Administration														
None of the above	end	x	x	x	x	x	x	x						
Embalming, EM														x
Foundry industry, FOI														
All								x		x				x
Production work														
Maintenance or utilities work		x	x	x	x	x	x							
Quality control, environmental control, engineering, or research work											x			

Shipping, receiving, or storage work			x		x	x	x							
Administration														
Furniture industry, FUI														
All		some	some	some	x	detailed		x	some			some		x
Performed maintenance or repair work		x					x							x
Farming and food industry, GF_FDI														
Farm work where crops were grown or livestock raised	GF													
Farming, GF														
All								x					x	
Performed maintenance or repair work		x					x							
Factory, warehouse, or processing center for food products	FDI													
Food industry, FDI														
All								x		x			some	x
Production work														
Maintenance or utilities work		x	x	x	x	x	x							
Quality control, environmental control, engineering, or research work														
Shipping, receiving, or storage work			x		x	x	x							
Administration														
Restaurant or other food service location where food was prepared or sold	end													
None of the above	end	x	x	x	x	x	x	x						

Health professional, HP														
All								x						
Performed maintenance or repair work		x	x	x	x	x	x							
Industry solvent, INDSOL														
All								x		x				
Administration, soles or other office work														
Management														
Production work		x	x	x	x	x	x							
Maintenance or utilities work		x	x	x	x	x	x							
Quality control, environmental control, engineering, or research work											x			
Shipping, receiving, or storage work			x		x	x	x							
Farm work	GF													
Somewhere else														
Janitor, JA														
All		some	some		some			x						x
Performed maintenance or repair work		x					x							
Leather industry, LEI														
All								x		x				
Production work					some							some		
Maintenance or utilities work		x	x	x	x	x	x							
Quality control, environmental control,														
engineering, or research work														
Shipping, receiving, or storage work			x		x	x	x							
Administration														

Lumber industry, LUI														
All								x						
Logging or cut up logs in sawmill														
Make furniture	FUI													
Make composite or veneered wood						some								x
Performed maintenance or repair work		x	x	x	x	some	x		x					x
Oil refinery industry, ORI														
All								x		x				x
Gas station attendant work							x							
Production work														
Maintenance or utilities work		x	x	x	x	x	x							
Quality control, environmental control, engineering, or research work											x			
Shipping, receiving, or storage work			x		x	x	x							
Administration														
Pesticide applicators, PE														
All							x					x		
Photographic industry, PHI														
All							x		x				x	
Production work														
Maintenance or utilities work	x	x	x	x	x	x								
Quality control, environmental control, engineering, or research work										x				
Shipping, receiving, or storage work		x		x	x	x								
Administration														
Plastic industry, PLI														
All							x		x		some		x	
Production work														
Maintenance or utilities work	x	x	x	x	x	x								
Quality control, environmental control, engineering, or research work										x				
Shipping, receiving, or storage work		x		x	x	x								
Administration														
Pulp and paper industry, PPI														
All							x		x				x	
Production work														
Maintenance or utilities work	x	x	x	x	x	x								
Quality control, environmental control, engineering, or research work										x				
Shipping, receiving, or storage work		x		x	x	x								
Administration														
Printing industry, PRI														
All							x		x				x	
Production work	some			some										
Maintenance or utilities work	x	x	x	x	x	x								
Quality control, environmental control, engineering, or research work	some			some						x				
Shipping, receiving, or storage work		x		x	x	x								
Administration	some			some										

Rubber industry, RUI													
All				some			x		x			x	
Production work													
Maintenance or utilities work		x	x	x		x	x						
Quality control, environmental control, engineering, or research work											x		
Shipping, receiving, or storage work			x			x	x						
Administration													
Shoe industry, SHI													
All					some			x		x			
Production work													
Maintenance or utilities work		x	x	x		x	x						
Quality control, environmental control,													
engineering, or research work													
Shipping, receiving, or storage work			x			x	x						
Administration													
Solvent, SOL													
All		x	x	x	x	x	x	x	x				
Farm work	GF												
Teaching, TE													
All													x
Shop teacher		x	x					x					

Supplementary Table 2: Decision rules for the NCI OccMATE algorithm's assignment of modules

Rule No.	Solvent screening question response	Number of module matches to occupation keywords, by keyword type		Number of module matches to industry keyword, by keyword type		Additional criteria	Action
		Type 1	Type 2/3	Type 1	Type 2/3		
1	≥1 yes	0	0	0	0		Assign solvent module
2	All "no's"	0	0	0	0		Assign work location module
3	Any	0		1			Use keyword match to assign module
4	Any	1		0			Use keyword match to assign module
5	Any	≥ 2	0	0	0		Assign solvent module
6	Any	≥ 2	1	0	0		Use T2/3 keyword match to assign module
7	Any	≥ 1	≥ 2	0	0	One T1 and T2/3 keyword directs to same module	Use concurrent T1 and T2/3 keyword match to assign module
8	Any	≥ 1	≥ 2	0	0	Occupation keywords direct to different modules	Assign solvent module
9	Any	0	0	≥ 2	0		Assign solvent module
10	Any	0	0	≥ 2	1		Use T2/3 keyword match to assign module
11	Any	0	0	≥ 1	≥ 2	One T1 and T2/3 keyword directs to same module	Use concurrent T1 and T2/3 keyword match to assign module
12	Any	0	0	≥ 1	≥ 2	Industry keywords direct to different modules	Assign solvent module
13	Any	1		1		Occupation and industry keywords direct to different modules	Use custom match[a] tables to assign module
14	Any	1		≥ 2		One occupation and one industry keyword directs to same module	Use concurrent occupation and industry keyword match to assign module
15	Any	1		≥ 2	0	Occupation and industry keywords direct to different modules	Use custom match[a] tables to assign module

16	Any	1	0	≥ 1	1	Occupation and industry keywords direct to different modules	Use custom match[a] tables using Type 2/3 match to assign module
17	Any	0	1	≥ 1	1	Occupation and industry keywords direct to different modules	Use custom match[a] tables using Type 2/3 matches to assign module
18	Any	1		1	≥ 2	Occupation and industry keywords direct to different modules	Use custom match[a] tables to assign module
19	Any	≥ 2		1		One occupation and one industry keyword directs to same module	Use concurrent occupation and industry keyword match to assign module
20	Any	≥ 2	0	1		Occupation and industry keywords direct to different modules	Use custom match[a] tables to assign module
21	Any	≥ 1	1	1		Occupation and industry keywords direct to different modules	Use custom match[a] tables using Type 2/3 match to assign module
22	Any	0	≥ 2	1		Occupation and industry keywords direct to different modules	Use custom match[a] tables to assign module
23	Any	≥ 2		≥ 2		One occupation and one industry keyword directs to same module	Use concurrent occupation and industry keyword match to assign module
24	Any	≥ 2		≥ 2		Occupation and industry keywords direct to different modules	Assign solvent module

CHAPTER 4

Reliability of Inter-rater Occupation Coding and Potential Impact on Occupational Exposure Assessment

C Ge[1]; MC Friesen[2]; S Locke[2]; CH Chou[3]; HI Hsu[3]; YH Lai[3]; L Liu[4]; Y Liu[4]; WW Shang[5]; J Shi[4]; LM Tan[5]; PJ Tsai[3]; TW Tsin; WH Tsin; CY Zhang[5]; N Rothman[2]; L Qing[2]; R Vermeulen[1,6]

1. Institute for Risk Assessment Sciences, Utrecht University, The Netherlands;
2. Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD, USA;
3. Department of Environmental and Occupational Health, Medical College, National Cheng Kung University, Tainan, Taiwan;
4. Occupational Poisoning Department, Tianjin Occupational Diseases Precaution and Therapeutic Hospital, Tianjin, China;
5. Sichuan Provincial Center for Disease Control and Prevention, Chengdu, China;
6. Julius Centre for Public Health Sciences and Primary Care, University Medical Centre, Utrecht, The Netherlands.

Submitted

ABSTRACT

Objectives: Coding of job descriptions to occupation classification codes is the foundation for exposure assessment in epidemiology studies. We assessed agreement between codes assigned by coder-pairs in a hospital-based case-control study in East Asia and evaluated agreement between job-exposure matrix (JEM) estimates assigned to these codes.

Methods: Each job was assigned an ISCO-88 code independently by two coders. We calculated percent agreement between the codes from each coder pairing (1-Chengdu versus 2-Tianjin; 3-Hong Kong versus 4-Taiwan). The codes were linked to a population-based JEM and the exposure estimates were compared using weighted Cohen's Kappa ($\varkappa$). Stratified analyses were conducted for different occupations, industries, and study phases. Between two study phases, additional training and discussions on coding disagreements from the first phase were conducted.

Results: This analysis included 34,353 jobs from 12,590 subjects. Coders from centres 1 and 2 coded 21,774 jobs from Mainland China; coders from centres 3 and 4 coded 12,579 jobs from Hong Kong and Taiwan. Job coding percentage agreement ranged from 51.0% to 77.1% depending on job code granularity and coder pairing. JEM-based exposure estimates comparisons resulted in $\varkappa$'s ranging from 0.72-0.95 for coders from centres 1/2 (median: 0.86) and 0.62-0.95 for coders from centres 3/4 (median 0.75). Agreement in both measures varied depending on occupation and industry, and improved in the second phase.

Conclusion: Job coding agreement was similar with other occupational studies. Agreement between JEM estimates was higher than between assigned job codes, as not all job coding disagreements led to different exposure assignment.

INTRODUCTION

In occupational epidemiological studies conducted among the general population, occupation information is collected to either directly infer occupational exposure or used in combination with different tools, such as job-exposure matrices (JEMs), to assess exposure (Kromhout and Vermeulen, 2001). Prior to use in exposure assessment, however, this occupation information is typically coded using different classification schemes into groups with similar job titles and industries. These classification systems may be unique to a particular study, but are more often standardized international or national classification systems with a hierarchical structure (Mannetje and Kromhout, 2003).

Because exposure assessment in many studies is based directly on job classification rather than the subjects' actual job descriptions obtained through interview or registries, the assignment of job codes forms the foundation of exposure assessment in occupational epidemiological studies. In recent years, job coding quality and reliability have received more attention for two additional reasons. First, modern population-based epidemiological studies with thousands of subjects typically divide the task of coding tens of thousands jobs among different coders. These coders may have various degrees of job coding experience and knowledge in occupational exposures; therefore job coding quality assurance and control became an important component in modern large-scale epidemiological studies. Second, there are increasing efforts to investigate and explore exposure-disease relationships with relatively low risks, which necessitate better exposure assessment. Improvements in job coding may have a crucial positive effect on a study's overall exposure assessment quality. However, only a few studies have compared inter-coder agreement in job classification, typically with small subsets of selected jobs within larger studies (Kennedy et al., 2000; Koeman et al., 2013; Kromhout and Vermeulen, 2001; Russ et al., 2016). Similarly, very few studies have investigated the reliability of JEM-based exposure assessment following job coding with different coders.

In the current study, we assessed the concordance of job coding between four different job coding study centres and investigated subsequent agreement in JEM-assessed exposures for jobs reported by over 12,000 study subjects in AsiaLymph, a large-scale hospital-based case-control study of lymphoma and leukemia in East Asia.

METHODS

Study Population

The AsiaLymph study enrolled subjects from 18 hospitals from four different regions in East Asia: Chengdu (Sichuan University Huaxi Hospital; Sichuan Province People's Hospital; Chengdu Tumour Hospital); Hong Kong (Queen Elizabeth Hospital; Queen Mary Hospital; Tuen Mun Hospital; Princess Margaret Hospital; Pamela Youde Eastern Hospital); Tianjin (Tianjin First Central Hospital; Tianjin Medical University Cancer Institute and Hospital; Tianjin Medical University General Hospital; Institute of Hematology and Blood Diseases Hospital; The Second Hospital of Tianjin Medical University); Taiwan (China Medical University Hospital; Chang Gung Memorial Hospital; Chia-Yi Christian Hospital; Kaohsiung Medical University Hospital). Informed consent was obtained from each participant and ethical approval for the research was granted by the Institutional Review Board of the US National Cancer Institute.

All eligible subjects were interviewed in person by the local research team within 48 hours of recruitment using a computer assisted personal interview (CAPI) device in the local language. The occupational history section of the interview collected information on all long-term (>1 year) jobs ever held by study participants and consisted of 12 core questions that included employer name, type of products/services provided by the employer, job title, and job tasks plus numeric information for the starting and ending age/year for each job (see Supplementary Figures 1-4 for full questionnaire). Based on the responses given for the core questions, the CAPI system automatically assigned job-specific question modules, which were described in detail elsewhere (Friesen et al., 2016). Free-text information from the core questions was used for occupation coding.

Occupation coding

Standardised coding of occupations was performed separately for jobs reported in regions where Simplified Chinese is used (Chengdu, Tianjin) and regions where Traditional Chinese is used (Hong Kong, Taiwan). Local industrial hygienist teams in Chengdu (centre 1) and Tianjin (centre 2) coded all jobs from the Simplified Chinese regions and industrial hygienist teams in Hong Kong (centre 3) and Taiwan (centre 4) coded all jobs from the Traditional Chinese regions, resulting in two sets of assigned job codes for each job. The industrial hygienist teams consisted of departmental directors (LMT; ZCY) and an industrial hygienist (WWS) from the Sichuan Provincial Centre for Disease Control in centre 1, three occupational medicine physicians (YL, LL, JS) from the Tianjin Occupational Diseases Precaution and Therapeutic Hospital in centre 2, two occupational hygienists (TWT, WHT) who were formerly employed in the Hong Kong Labour Department in centre 3, and a team of two graduate students (YHL, CHC) and one occupational hygienist (HIH) lead by a professor in occupational health (PJT) in the Department of Environmental and Occupational Health in National Cheng Kung University in centre 4. Prior to the coding work, teams from all centres received a one-day face-to-face training on job coding, followed by two rounds of coding exercises consisting of 380 actual jobs reported in the study, with feedback given after each training round. The training jobs were drawn from jobs assigned with different job-specific modules by the CAPI system (Friesen et al., 2016) to ensure inclusion of jobs from different trades and industries.

Occupation coding was performed by the coding teams using the 5th Revision of Standard Occupational Classification System of the Republic of China (SOCS-ROC) published by The Directorate General of Budget, Accounting and Statistics of Executive Yuan of the Republic of China (DGBAS) (DGBAS, 1992), which was selected for its nearly identical structure to the 1988 version of the International Standard Classification of Occupations (ISCO-88) (ILO, 2010) and availability in the Traditional Chinese language with translated English job titles. Jobs were additionally assigned industry codes in Revision 4 of the International Standard Industrial Classification of All Economic Activities (ISIC4) (United Nations, 2008). After the duplicate coding in SOCS-ROC, all job codes were translated to ISCO-88 using official crosswalk documentation from DGBAS. All one to many matches in the crosswalk process from SOCS-ROC to ISCO-88 were reviewed (by CG) in order to assign the appropriate ISCO-88 codes.

Job coding was performed in two phases. Phase 1 began in March 2016, when 11,898 jobs from the Simplified Chinese regions were assigned to centres 1 and 2, and 8,942 jobs from the Traditional Chinese regions were assigned to the study centres 3 and 4. Prior to the second phase of job coding, additional training was provided to each study centre, by phone, to review job coding protocols and coding disagreements in phase 1. Phase 2 began in April 2018, when 9,876 jobs from the Simplified Chinese regions were assigned to centres 1 and 2, and 3,637 jobs from the Traditional Chinese regions were assigned to centres 3 and 4.

Statistical analysis

Percent agreement was calculated between ISCO-88 coding of jobs in Simplified Chinese by centres 1 and 2. Similarly for coding of jobs in Traditional Chinese, percent agreement was calculated for ISCO-88 codes assigned by centres 3 and 4. The comparisons were performed at all digit levels (i.e. 1-digit major groups; 2-digit sub-major groups; 3-digit sub-groups; 4-digit unit groups) to assess agreement in different hierarchical levels of job coding.

To study the potential impact in exposure assessment of different occupational exposures, all jobs were linked by ISCO-88 codes to the ALOHA+ JEM (Skorge et al., 2009), which assigns exposure to various substances using exposure scores of 0 (no exposure), 1 (low exposure), and 2 (high exposure). Cohen's Kappa (κ) values with linear weights were calculated for different exposures assessed using job codes from centres 1 versus 2, and centres 3 versus 4.

Stratified analyses were performed to assess if coding and exposure assessment agreement differed between 1) the two phases of job coding; 2) important trades and professions for occupational exposure (i.e. craft and trade workers, machine operators and assemblers, and labourers as defined by first-digit ISCO-88 codes 7, 8, and 9, respectively as coded by centres 1 and 3); and 3) important industries for occupational exposure (i.e. agriculture, mining, manufacturing, and construction as defined by first-digit ISIC4 codes A, B, C, and F, respectively as coded by centres 1 and 3). To further explore the reliability of JEM-assessed exposures for different occupations, percent agreement was calculated for three ALOHA+ JEM exposures (i.e. metals, aromatic solvents, and the combined category "vapours, dusts, gases, and fumes") for all frequently reported (n ≥50) occupations.

All statistical analyses were conducted in R (version 3.6).

RESULTS

A total of 21,774 jobs were reported by 8,474 subjects in the Simplified Chinese region and 12,579 jobs were reported by 4,116 subjects in the Traditional Chinese region. Percent agreement of independent ISCO-88 coding of jobs in Simplified Chinese between centres 1 and 2 ranged from 56.0% at the 4-digit level to 77.1% at the 1-digit level (Table 1). ISCO-88 coding agreement between centres 3 and 4 ranged from 51.0% at the 4-digit level to 71.0% at the 1-digit level.

Kappa agreement for JEM-assessed occupational exposure to various substances ranged from 0.72 to 0.95 between centres 1 and 2 (median: 0.86) and from 0.62 to 0.95 between centres 3 and 4 (median 0.75) (Table 2). Reliability of assessed exposures for all substances were higher between centres 1 and 2 compared to centres 3 and 4.

Table 1. Percent agreement (%) in various levels of assigned ISCO-88 job codes between different study centres

ISCO-88#	Centres 1 versus 2*	Centres 3 versus 4*
4 digit	56.0	51.0
3 digit	64.4	60.7
2 digit	73.0	68.1
1 digit	77.1	71.0

#: 1-digit: major groups; 2-digit: sub-major groups; 3-digit: sub-groups; 4-digit: unit groups
*: Centre 1: Chengdu; center 2: Tianjin; centre 3: Hong Kong; centre 4: Taiwan

ISCO-88 coding and JEM assessment agreements were both higher in phase 2 versus phase 1 for all study centre comparisons (Table 3). Compared to coding agreement in all jobs, ISCO-88 agreement was higher in agricultural jobs in all centres, as well as for machine operator and product assembler occupations in centres 1/2 and labourer jobs in centres 3/4. All craft and trade occupations, as well as jobs in manufacturing and construction industries had lower ISCO-88 agreement compared to agreement results for all jobs. Correspondingly for JEM-assessment agreement, the lowest median kappa values were found in craft and trade occupations and jobs in construction for all centres (except for mining jobs in centres 3/4 with median kappa of 0.32; however there were only 25 such jobs reported).

Table 2. Agreement (weighted Cohen's Kappa) in various occupational exposures assessed by the ALOHA+ JEM based on 4-digit job codes assigned by different study centres

Substance	Centres 1 versus 2*		Centres 3 versus 4*	
	Kappa	Prevalence %#	Kappa	Prevalence %#
Biological dust	0.85	35.6	0.71	30.8
Mineral dust	0.87	44.0	0.76	32.1
Gas fumes	0.81	55.1	0.74	50.7
Vapours, gases, dust, and fumes	0.87	61.5	0.77	58.7
Pesticides, all	0.93	19.6	0.87	6.5
Herbicides	0.95	16.5	0.95	4.7
Insecticides	0.95	17.6	0.91	6.1
Fungicides	0.93	18.4	0.90	5.2
Solvent, aromatic	0.82	26.2	0.71	17.6
Solvent, chlorinated	0.73	10.0	0.62	13.0
Solvent, other	0.72	17.3	0.69	18.4
Metals	0.78	11.3	0.68	11.3

*: Centre 1: Chengdu; center 2: Tianjin; centre 3: Hong Kong; centre 4: Taiwan
#: Exposure prevalence percentages based on job coding from Centre 1 (Chengdu) and 3 (Hong Kong)

Table 3. Percent agreement (%) in 4-digit ISCO-88 job codes assigned by different study centres and subsequent agreement (weighted Cohen's Kappa, κ) in various occupational exposures assessed by the ALOHA+ JEM[5]

	Centres 1 versus 2*			Centres 3 versus 4*		
	N	ISCO-88 agreement %[1]	Exposure median κ (range)[2]	N	ISCO-88 agreement %[1]	Exposure median κ (range)[2]
Phase 1	11,898	53.0	0.85 (0.73, 0.93)	8,942	50.5	0.72 (0.54, 0.94)
Phase 2	9,876	59.7	0.87 (0.70, 0.98)	3,637	57.0	0.82 (0.76, 0.96)
Craft & trade[3]	3,090	52.1	0.66 (0, 0.78)	1,632	36.5	0.53 (0, 0.69)
Machine op & assembler[3]	2,848	56.7	0.73 (0.55, 0.80)	1,898	42.7	0.58 (0.24, 0.65)
Labourer[3]	1,382	48.5	0.67 (0.46, 0.78)	1,137	62.1	0.64 (0, 0.72)
Agriculture[4]	3,928	67.2	0.76 (0.67, 0.90)	705	54.5	0.80 (0.25, 0.84)
Mining[4]	368	44.6	0.68 (0.42, 0.81)	25	52.0	0.32 (0, 0.84)
Manufacture[4]	5,800	51.1	0.73 (0.46, 0.79)	3,407	39.4	0.61 (0, 0.68)
Construction[4]	1,671	50.0	0.64 (0.40, 0.81)	818	39.0	0.50 (-0.01, 0.66)

*: Centre 1: Chengdu; center 2: Tianjin; centre 3: Hong Kong; centre 4: Taiwan
1. ISCO-88 4-digit unit group agreement;
2. Median and range of weighted kappa values (κ) for all substances in the ALOHA+ JEM;
3. Craft and trade workers, machines operators and assemblers, and labourers were as defined by first-digit ISCO-88 codes 7, 8, and 9, respectively, as coded by centres 1 and 3;
4. Agriculture, mining and quarrying, manufacturing, and construction industries were as indicated by first-digit ISIC4 codes A, B, C, and F, respectively, as coded by centres 1 and 3.
5. JEM ratings: exposure scores of 0 (no exposure), 1 (low exposure), and 2 (high exposure).

Agreement in JEM-based exposure was highly variable when evaluated at the job level. For frequently reported jobs in the Simplified Chinese region (n=81), percent agreement for the selected JEM exposures ranged from around 45 to 100%, though one outlier job (ISCO-88: 61 Market-oriented skilled agricultural and fishery workers) had percent agreement of 4.2% (Figure 1). The corresponding percent agreement results for frequently reported jobs in the Traditional Chinese region (n=66) ranged from around 25 to 100% (Figure 2). Full results of these analyses including each job code and agreement value for all three selected exposures are available in Supplementary Tables 1-6. Agreement results for exposures to the exposure category "vapours, gases, dusts and fumes" were lower in all centres than those observed for exposure to metals and aromatic solvents for all regions.

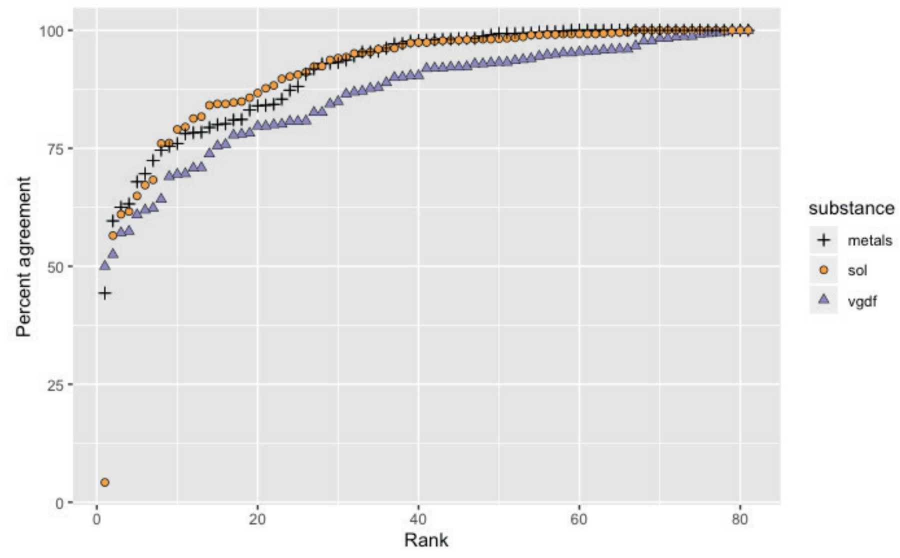


Figure 1: Percent agreement of JEM-assessed exposures for commonly reported jobs (n≥50) using job codes from Centre 1 (Chengdu) versus Centre 2 (Tianjin) (sol=aromatic solvents; vgdf= vapours, gases, dust, and fumes).

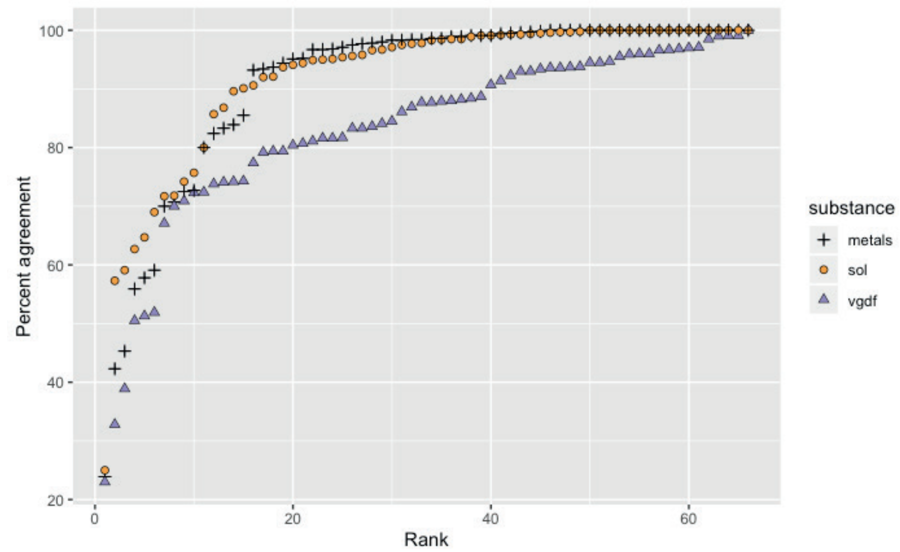


Figure 2: Percent agreement of JEM-assessed exposures for commonly reported jobs (n≥50) using job codes from Centre 3 (Hong Kong) versus Centre 4 (Taiwan) (sol=aromatic solvents; vgdf= vapours, gases, dust, and fumes).

DISCUSSION

In a large, multi-centred hospital-based case-control study, we compared differences in occupational coding performed independently by industrial hygienist teams in different study centres. Our results show 51% and 56% agreement for the two sets of jobs in 4-digit ISCO-88 coding, with higher agreement when coding granularity is decreased (up to 77% agreement in 1-digit ISCO-88). These results are in similar ranges as those reported in other studies comparing inter-rater job coding agreement. In a Canadian asthma case-control study with 680 reported jobs, the 4-digit ISCO-88 coding agreement was 60% between the original coders and an independent recoder (Kennedy et al., 2000). When four job coders coded 112 jobs in a study in the Netherlands, percent agreement for 5-digit ISCO-68 coding ranged from 56-68% (Kromhout and Vermeulen, 2001). Another Dutch study reported percent agreement of 36% for 220 jobs individually coded in 5-digit ISCO-68 with minimal training and few instructions (Koeman et al., 2013). Two coders in a US study had agreements of 50.4%, 56.8%, 68.6%, and 86.2% for a set of 1,000 jobs coded in six-, five-, three-, and two-digit Standard Occupational Classification (SOC) 2010 codes (Russ et al., 2016). Coding agreement in our study improved as coders became more experienced with the classification system and received feedback on coding errors. This observation is in accordance with previous works noting improvements in coding reliability with clearer coding instructions, better training, and more experience (Kromhout and Vermeulen, 2001; Mannetje and Kromhout, 2003).

Observed kappa agreements on JEM-assessed occupational exposures to various agents using job codes from different centres varied depending on exposure substance, but overall our results were similar with those from other studies with similar design. For instance, Koeman and colleagues reported weighted kappa agreements ranging from 0.66 to 0.84 (median 0.72) for various JEM-assessed exposures using job codes from different coders (Koeman et al., 2013). Similarly, Russ and colleagues reported a weighted kappa of 0.72 for JEM-based lead exposure probabilities obtained using job codes from two coders (Russ et al., 2016). Although higher coding agreement generally led to higher exposure assessment agreement, our results are also compatible with reports from others suggesting that the effect of coding disagreement tends to diminish when the job codes are used to assign exposure in general population JEMs (Kennedy et al., 2000; Koeman et al., 2013; Kromhout and Vermeulen, 2001). For instance, in our study agricultural jobs had coding agreements of 67.2% for centres 1 versus 2 and 54.5% for centres 3 versus 4, but the median kappa values for JEM-based exposures were very similar (0.76 and 0.80, respectively). Coding disagreements in jobs with no exposure (e.g. administrative versus managerial workers) or similar exposures (e.g. grain versus fruit farmers) do not necessarily lead to disagreements in JEM-assessed exposures.

Our stratified results suggest that coding and assessment reliability differed between different major occupation and industry groups. For instance, both coding and JEM-assessment agreement metrics for jobs in construction and manufacturing were lower compared to jobs in agriculture in all study centres, presumably because it is more difficult to code less detailed job descriptions in construction and manufacturing (e.g. construction/manufacturing worker) versus those in agriculture (e.g. farm worker). Construction and manufacturing workers may be involved in different (or multiple) trades such as carpentry, roofing, material handling, packaging, and clean-up whereas farm workers may be coded to 4-digits as long as the type of crop is known. When we further scrutinized exposure assessment agreement by individual occupations, we found that while agreement was reasonable (≥80%) for many frequently reported jobs, some jobs had poor agreement results. For instance, among the Simplified Chinese jobs, agreement of JEM-based aromatic solvent exposure was only 4.2% for 307 jobs coded by centre 1 as "61: market-oriented skilled agricultural and fishery workers" (Supplementary table 2). Further manual review showed that many descriptions for these jobs were insufficient for determining the precise type of farming (e.g. "farmer", "grower", "agriculture work"). Consequently, coding centre 1 defaulted to assigning a generic two-digit agriculture job code "61" whereas centre 2 defaulted to a three-digit job code "611: market gardeners and crop growers," leading to low job coding and exposure assessment agreement. Our data showed other similar difficulties for the coders, such as in distinguishing truck versus car drivers for "delivery drivers" as well as in separating electrical/electronics mechanics with building electricians for "electricians." Overall, jobs with the lowest exposure assessment reliability in our study tended to have more ambiguous or less detailed job descriptions. We are completing a thorough review of all discordant job codes from different coders in our study. This will help ensure similar decisions are made for jobs with ambiguous descriptions to minimize inter-coder differences in subsequent exposure assessment procedures.

Our observations show an inherent disconnect between the interview and job coding stages in almost all large-scale occupational epidemiology studies relying on subject interviews to obtain job histories, including ours. Because interviewers typically have little knowledge in job coding (and even less knowledge in the specific job coding schemes used in the study), they could not assess whether job descriptions collected are sufficient for proper job coding. In theory, either training of interviewers in job coding or having interviewers code reported jobs in real time during interviews might remedy this issue. However both options are unlikely to be feasible as they place significant additional burden on the interviewers and study subjects. Additional improvements to the coding may also be possible by providing the coders with the subjects' responses to industry/job-specific questions that obtain additional information from subjects, particularly those with jobs with potential exposure. However, reviewing this information would add substantial time to the already lengthy coding process and thus it is generally only used in the assessment of exposure rather than classification of jobs. One potential solution to this challenge in future studies may be the application of automatic job coding algorithms (e.g. De Matteis et al., 2017; Rémen et al., 2018; Russ et al., 2016) during job history interviews, where job coding may take place in real-time and the study subjects may immediately clarify any ambiguous job descriptions. In our current work, we have additionally employed automatically assigned job-specific question modules to directly obtain exposure information from study participants with potentially exposed jobs (Friesen et al., 2016). Responses from these job-specific questions will be used in conjunction with the assigned job codes to further improve exposure assessment within the AsiaLymph study.

To our knowledge, ours is the largest study to investigate agreements in job coding from independent coders and resultant JEM-assessed exposures. Rather than selecting a small subset of jobs, we included all reported jobs in our analysis. Therefore our results may better represent job coding and related exposure assessment agreement performances in large-scale epidemiological studies with tens of thousands of reported jobs, where use of multiple job coders is the norm. The large sample size also allowed our study to include stratified analyses to investigate the potential changes in job coding and exposure assessment agreement in different occupations, industries, and chronological phases of the study. Jobs in our study were coded in a classification system that is nearly identical to ISCO-88, which is the main classification system used in numerous countries and comparable with many other national classification systems (Mannetje and Kromhout, 2003). Because our study provided extensive coding instructions, training, and feedback to all study centres, agreement results we reported may represent the upper range of concordance values that are feasible in similar study settings.

There are also limitations in our study. Because of the lack of a true gold standard for job coding, higher reliability in coding may not necessarily equate to more accurate coding. For instance, increases in coding agreement after more training and experience may be observed if all coders became fatigued and coded more jobs to more generic, unspecified job categories rather than the appropriate specific jobs (e.g. "8290: other machine operators and assemblers" instead of "8272: dairy-products machine operators"). Further analyses on generic job codes did uncover slight increases in the proportions of generic codes (0.7-5.3%) for all centres in phase 2 that perhaps suggest a certain level of coding fatigue; however this effect is too small to be the main driver of our main results (Supplementary table 7). Additionally, the JEM we selected was not specifically designed for the study region, and therefore our assessed exposures may not fully reflect actual exposures for the included jobs. However, due to the lack of general population JEMs tailored for workers in Asia, our work represents a practical approach for characterizing the potential impact of different job codes on JEM-assessed exposures for this analysis; Asia-specific exposure assessments will be assigned in future efforts. In addition, comparisons of the magnitude of kappa across groups must be interpreted cautiously, especially if the underlying exposure distribution differs between groups. Another limitation involves the crosswalk between the original coding in SOCS-ROC and final coding in ISCO-88. Although the two coding systems were nearly identical, there were a few instances of one-to-many matches that were resolved manually. Because the same person resolved all one-to-many matches during the crosswalk, coding agreement in ISCO-88 may be overestimated. However, because this procedure only applied to a small number of jobs (~3% of all jobs), the magnitude of the bias, if present, should be very small.

In summary, we reported inter-rater agreement in job coding and JEM-assessed occupational exposures in a large hospital-based case-control study in East Asia. JEM-exposure assessment agreement tended to be higher than job coding agreement. Use of tools or strategies that extract more detailed job descriptions, especially from potentially exposed jobs, may further increase both job coding and exposure assessment reliability.

REFERENCES

De Matteis, S., Jarvis, D., Young, H., Young, A., Allen, N., Potts, J., Darnton, A., Rushton, L., Cullinan, P., 2017. Occupational self-coding and automatic recording (OSCAR): a novel web-based tool to collect and code lifetime job histories in large population-based studies. Scand. J. Work. Environ. Health 43, 181–186. https://doi.org/10.5271/sjweh.3613

DGBAS, 1992. STANDARD OCCUPATIONAL CLASSIFICATION SYSTEM OF THE REPUBLIC OF CHINA; Revision 5 [WWW Document]. URL https://www.dgbas.gov.tw/ct.asp?xItem=15810&ctNode=3112 (accessed 9.13.19).

Friesen, M.C., Lan, Q., Ge, C., Locke, S.J., Hosgood, D., Fritschi, L., Sadkowsky, T., Chen, Y.-C., Wei, H., Xu, J., Lam, T.H., Kwong, Y.L., Chen, K., Xu, C., Su, Y.-C., Chiu, B.C.H., Ip, K.M.D., Purdue, M.P., Bassig, B.A., Rothman, N., Vermeulen, R., 2016. Evaluation of Automatically Assigned Job-Specific Interview Modules. Ann. Occup. Hyg. 60, 885–899. https://doi.org/10.1093/annhyg/mew029

ILO, 2010. ISCO-International Standard Classification of Occupations: Brief History [WWW Document]. URL http://www.ilo.org/public/english/bureau/stat/isco/intro2.htm (accessed 7.20.18).

Kennedy, S.M., Moual, N.L., Choudat, D., Kauffmann, F., 2000. 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. 57, 635–641. https://doi.org/10.1136/oem.57.9.635

Koeman, T., Offermans, N.S.M., Christopher-de Vries, Y., Slottje, P., Van Den Brandt, P.A., Goldbohm, R.A., Kromhout, H., Vermeulen, R., 2013. JEMs and incompatible occupational coding systems: effect of manual and automatic recoding of job codes on exposure assignment. Ann. Occup. Hyg. 57, 107–114. https://doi.org/10.1093/annhyg/mes046

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

Mannetje, A. 't, Kromhout, H., 2003. The use of occupation and industry classifications in general population studies. Int. J. Epidemiol. 32, 419–428. https://doi.org/10.1093/ije/dyg080

Rémen, T., Richardson, L., Pilorget, C., Palmer, G., Siemiatycki, J., Lavoué, J., 2018. Development of a Coding and Crosswalk Tool for Occupations and Industries. Ann. Work Expo. Health 62, 796–807. https://doi.org/10.1093/annweh/wxy052

Russ, D.E., Ho, K.-Y., Colt, J.S., Armenti, K.R., Baris, D., Chow, W.-H., Davis, F., Johnson, A., Purdue, M.P., Karagas, M.R., Schwartz, K., Schwenn, M., Silverman, D.T., Johnson, C.A., Friesen, M.C., 2016. Computer-based coding of free-text job descriptions to efficiently identify occupations in epidemiological studies. Occup. Environ. Med. 73, 417–424. https://doi.org/10.1136/oemed-2015-103152

Skorge, T.D., Eagan, T.M., Eide, G.E., Gulsvik, A., Bakke, P.S., 2009. Occupational exposure and incidence of respiratory disorders in a general population. Scand. J. Work. Environ. Health 35, 454–461. https://doi.org/10.5271/sjweh.1352

United Nations, 2008. International Standard Industrial Classification of All Economic Activities (ISIC), Rev. 4.

Supplementary Table 1: Percent agreement for JEM-assessed exposure to metals for frequently reported jobs (N ≥50) in the Simplified Chinese region

ISCO-88 code[1]	Job description	N	% agree[2]
7111	Miners and quarry workers	88	44.3
9312	Construction and maintenance labourers: roads, dams and similar constructions	213	59.6
7133	Plasterers	112	62.5
7141	Painters and related workers	76	63.2
7241	Electrical mechanics and fitters	53	67.9
7221	Blacksmiths, hammer-smiths and forging-press workers	69	69.6
7233	Agricultural- or industrial-machinery mechanics and fitters	243	72.4
8283	Electronic-equipment assemblers	67	74.6
7242	Electronics fitters	69	75.4
7137	Building and related electricians	146	76
7136	Plumbers and pipe fitters	73	78.1
7231	Motor vehicle mechanics and fitters	161	78.3
9322	Hand packers and other manufacturing labourers	315	78.4
3115	Mechanical engineering technicians	63	79.4
8261	Fibre-preparing-, spinning and winding machine operators	50	80
8290	Other machine operators and assemblers	126	80.2
7123	Concrete placers, concrete finishers and related workers	58	81
3111	Chemical and physical science technicians	74	81.1
3112	Civil engineering technicians	77	83.1
8263	Sewing machine operators	50	84
8211	Machine-tool operators	346	84.1
9333	Freight handlers	166	84.3
8122	Metal melters, casters and rolling-mill operators	96	85.4
zzzz	Not coded	63	87.3
7222	Tool-makers and related workers	135	88.1
8162	Steam-engine and boiler operators	75	90.7
8232	Plastic-products machine operators	73	91.8
7120	Building frame and related trades workers	197	92.9
8262	Weaving- and knitting-machine operators	115	93
7130	Building finishers and related trades workers	59	93.2
8131	Glass and ceramics kiln and related machine operators	79	93.7
3152	Safety, health and quality inspectors	236	94.5
7436	Sewers, embroiderers and related workers	65	95.4
8331	Motorised farm and forestry plant operators	65	95.4
7212	Welders and flamecutters	181	95.6
5169	Protective services workers not elsewhere-classified	106	96.2
7122	Bricklayers and stonemasons	103	97.1
9152	Doorkeepers, watchpersons and related workers	213	97.2
3439	Administrative associate professionals not-elsewhere classified	474	97.7
4100	Office clerks	2003	97.8
4190	Other office clerks	94	97.9
9331	Hand or pedal vehicle drivers	51	98
3416	Buyers	162	98.1
4131	Stock clerks	317	98.1
2310	College, university and higher education-teaching professionals	113	98.2
7433	Tailors, dressmakers and hatters	115	98.3
7442	Shoe-makers and related workers	120	98.3
7124	Carpenters and joiners	159	98.7
3434	Statistical, mathematical and related associate-professionals	104	99
7422	Cabinet makers and related workers	151	99.3
8320	Motor-vehicle drivers	431	99.3
8322	Car, taxi and van drivers	146	99.3
0110	Armed forces	530	99.4
3433	Bookkeepers	481	99.4
2320	Secondary education teaching professionals	230	99.6
5220	Shop salespersons and demonstrators	1445	99.7
6100	Market-oriented skilled agricultural and fishery workers	307	99.7
1210	Directors and chief executives	476	99.8
2221	Medical doctors	229	100
2230	Nursing and midwifery professionals	102	100
2300	Teaching professionals	138	100
2331	Primary education teaching professionals	254	100
2332	Preprimary education teaching professionals	96	100
3415	Technical and commercial sales representatives	175	100
3449	Customs, tax and related government associate-professionals not elsewhere classified	126	100
4132	Production clerks	61	100
4211	Cashiers and ticket clerks	200	100
5122	Cooks	423	100
5123	Waiters, waitresses and bartenders	223	100
5141	Hairdressers, barbers, beauticians and related-workers	70	100
5162	Police officers	99	100

6110	Market gardeners and crop growers	2251	100
6111	Field crop and vegetable growers	716	100
6112	Tree and shrub crop growers	173	100
6121	Dairy and livestock producers	63	100
7411	Butchers, fishmongers and related food preparers	54	100
7412	Bakers, pastry-cooks and confectionery makers	61	100
8324	Heavy truck and lorry drivers	101	100
9131	Domestic helpers and cleaners	53	100
9132	Helpers and cleaners in offices, hotels and other establishments	155	100
9133	Hand-launderers and pressers	53	100

1. as coded by centre 1 (Chengdu)
2. % agree = percent agreement

Supplementary Table 2: Percent agreement for JEM-assessed exposure to aromatic solvents for frequently reported jobs (N ≥50) in the Simplified Chinese region

ISCO-88 code	Job description	N	% agree
6100	Market-oriented skilled agricultural and fishery workers	307	4.2
7221	Blacksmiths, hammer-smiths and forging-press workers	69	56.5
9312	Construction and maintenance labourers: roads, dams and similar constructions	213	61
7133	Plasterers	112	61.6
3111	Chemical and physical science technicians	74	64.9
8283	Electronic-equipment assemblers	67	67.2
3115	Mechanical engineering technicians	63	68.3
7137	Building and related electricians	146	76
7111	Miners and quarry workers	88	76.1
9322	Hand packers and other manufacturing labourers	315	79
7136	Plumbers and pipe fitters	73	79.5
9333	Freight handlers	166	81.3
8290	Other machine operators and assemblers	126	81.7
7242	Electronics fitters	69	84.1
3112	Civil engineering technicians	77	84.4
7222	Tool-makers and related workers	135	84.4
7130	Building finishers and related trades workers	59	84.7
8232	Plastic-products machine operators	73	84.9
7231	Motor vehicle mechanics and fitters	161	85.7
6112	Tree and shrub crop growers	173	86.7
7233	Agricultural- or industrial-machinery mechanics and fitters	243	87.7
7442	Shoe-makers and related workers	120	88.3
7123	Concrete placers, concrete finishers and related workers	58	89.7
8211	Machine-tool operators	346	90.2
8122	Metal melters, casters and rolling-mill operators	96	90.6
8131	Glass and ceramics kiln and related machine operators	79	91.1
7436	Sewers, embroiderers and related workers	65	92.3
7120	Building frame and related trades workers	197	92.4
zzzz	Not coded	63	93.7
3152	Safety, health and quality inspectors	236	94.1
7241	Electrical mechanics and fitters	53	94.3
7122	Bricklayers and stonemasons	103	95.1
6121	Dairy and livestock producers	63	95.2
7422	Cabinet makers and related workers	151	95.4
8261	Fibre-preparing-, spinning- and winding machine operators	50	96

5169	Protective services workers not elsewhere classified	106	96.2
7124	Carpenters and joiners	159	96.2
8331	Motorised farm and forestry plant operators	65	96.9
8162	Steam-engine and boiler operators	75	97.3
7141	Painters and related workers	76	97.4
8262	Weaving- and knitting-machine operators	115	97.4
9152	Doorkeepers, watchpersons and related workers	213	97.7
7212	Welders and flamecutters	181	97.8
3439	Administrative associate professionals not elsewhere classified	474	97.9
4190	Other office clerks	94	97.9
8263	Sewing machine operators	50	98
4100	Office clerks	2003	98.1
4131	Stock clerks	317	98.1
9133	Hand-launderers and pressers	53	98.1
2310	College, university and higher education teaching professionals	113	98.2
3415	Technical and commercial sales representatives	175	98.3
7412	Bakers, pastry-cooks and confectionery makers	61	98.4
8322	Car, taxi and van drivers	146	98.6
3434	Statistical, mathematical and related associate professionals	104	99
8324	Heavy truck and lorry drivers	101	99
6110	Market gardeners and crop growers	2251	99.1
7433	Tailors, dressmakers and hatters	115	99.1
0110	Armed forces	530	99.2
1210	Directors and chief executives	476	99.2
3433	Bookkeepers	481	99.2
3449	Customs, tax and related government associate professionals not-elsewhere classified	126	99.2
5220	Shop salespersons and demonstrators	1445	99.2
6111	Field crop and vegetable growers	716	99.3
3416	Buyers	162	99.4
8320	Motor-vehicle drivers	431	99.5
2320	Secondary education teaching professionals	230	99.6
2221	Medical doctors	229	100
2230	Nursing and midwifery professionals	102	100
2300	Teaching professionals	138	100
2331	Primary education teaching professionals	254	100
2332	Preprimary education teaching professionals	96	100
4132	Production clerks	61	100
4211	Cashiers and ticket clerks	200	100
5122	Cooks	423	100
5123	Waiters, waitresses and bartenders	223	100
5141	Hairdressers, barbers, beauticians and related workers	70	100
5162	Police officers	99	100
7411	Butchers, fishmongers and related food preparers	54	100
9131	Domestic helpers and cleaners	53	100
9132	Helpers and cleaners in offices, hotels and other establishments	155	100
9331	Hand or pedal vehicle drivers	51	100

1. as coded by centre 1 (Chengdu)
2. % agree = percent agreement

Supplementary Table 3: Percent agreement for JEM-assessed exposure to vapours, dust, gases, and fumes for frequently reported jobs (N ≥50) in the Simplified Chinese region

ISCO-88 code	Job description	N	% agree
7411	Butchers, fishmongers and related food preparers	54	50
7130	Building finishers and related trades workers	59	52.5
3112	Civil engineering technicians	77	57.1
7412	Bakers, pastry-cooks and confectionery makers	61	57.4
7221	Blacksmiths, hammer-smiths and forging-press workers	69	60.9
zzzz	Not coded	63	61.9
9131	Domestic helpers and cleaners	53	62.3
7241	Electrical mechanics and fitters	53	64.2
9132	Helpers and cleaners in offices, hotels and other establishments	155	69
5123	Waiters, waitresses and bartenders	223	69.5
8320	Motor-vehicle drivers	431	69.6
9322	Hand packers and other manufacturing labourers	315	70.8
7122	Bricklayers and stonemasons	103	70.9
8290	Other machine operators and assemblers	126	73.8
9133	Hand-launderers and pressers	53	75.5
5162	Police officers	99	75.8
3115	Mechanical engineering technicians	63	77.8
8261	Fibre-preparing-, spinning- and winding machine operators	50	78
7233	Agricultural- or industrial-machinery mechanics and fitters	243	78.2
7242	Electronics fitters	69	79.7
8131	Glass and ceramics kiln and related machine operators	79	79.7
7436	Sewers, embroiderers and related workers	65	80
7231	Motor vehicle mechanics and fitters	161	80.1
7120	Building frame and related trades workers	197	80.7
9333	Freight handlers	166	80.7
7136	Plumbers and pipe fitters	73	80.8
8262	Weaving- and knitting-machine operators	115	82.6
9312	Construction and maintenance labourers: roads, dams and similar constructions	213	82.6
8211	Machine-tool operators	346	84.4
5169	Protective services workers not elsewhere classified	106	84.9
8122	Metal melters, casters and rolling-mill operators	96	86.5
4132	Production clerks	61	86.9
8324	Heavy truck and lorry drivers	101	87.1
3152	Safety, health and quality inspectors	236	87.7
4100	Office clerks	2003	87.9
3416	Buyers	162	88.9
7442	Shoe-makers and related workers	120	90
3439	Administrative associate professionals not elsewhere classified	474	90.1
7222	Tool-makers and related workers	135	90.4
8232	Plastic-products machine operators	73	90.4
3111	Chemical and physical science technicians	74	91.9
5122	Cooks	423	92
8263	Sewing machine operators	50	92
7433	Tailors, dressmakers and hatters	115	92.2
9331	Hand or pedal vehicle drivers	51	92.2
8331	Motorised farm and forestry plant operators	65	92.3
1210	Directors and chief executives	476	92.9
3449	Customs, tax and related government associate professionals not-elsewhere classified	126	92.9
4131	Stock clerks	317	93.1
7111	Miners and quarry workers	88	93.2
8322	Car, taxi and van drivers	146	93.2
4190	Other office clerks	94	93.6
2310	College, university and higher education teaching professionals	113	93.8
8283	Electronic-equipment assemblers	67	94
4211	Cashiers and ticket clerks	200	94.5
9152	Doorkeepers, watchpersons and related workers	213	94.8
2300	Teaching professionals	138	94.9
3434	Statistical, mathematical and related associate professionals	104	95.2
6121	Dairy and livestock producers	63	95.2
3415	Technical and commercial sales representatives	175	95.4
7133	Plasterers	112	95.5
5220	Shop salespersons and demonstrators	1445	95.6
0110	Armed forces	530	95.8
7137	Building and related electricians	146	95.9
7422	Cabinet makers and related workers	151	96
8162	Steam-engine and boiler operators	75	96
7212	Welders and flamecutters	181	96.7
6112	Tree and shrub crop growers	173	97.7
2320	Secondary education teaching professionals	230	97.8
3433	Bookkeepers	481	98.3

7123	Concrete placers, concrete finishers and related workers	58	98.3
5141	Hairdressers, barbers, beauticians and related workers	70	98.6
2221	Medical doctors	229	98.7
7124	Carpenters and joiners	159	98.7
2331	Primary education teaching professionals	254	99.2
6100	Market-oriented skilled agricultural and fishery workers	307	99.3
6110	Market gardeners and crop growers	2251	99.5
6111	Field crop and vegetable growers	716	99.6
2230	Nursing and midwifery professionals	102	100
2332	Preprimary education teaching professionals	96	100
7141	Painters and related workers	76	100

1. as coded by centre 1 (Chengdu)
2. % agree = percent agreement

Supplementary Table 4: Percent agreement for JEM-assessed exposure to metals for frequently reported jobs (N ≥50) in the Traditional Chinese region

ISCO-88 code	Job description	N	% agree
3114	Electronics and telecommunications engineering technicians	67	23.9
8223	Metal finishing-, plating- and coating-machine operators	52	42.3
2144	Electronics and telecommunications engineers	53	45.3
7241	Electrical mechanics and fitters	227	55.9
8290	Other machine operators and assemblers	102	57.8
8284	Metal-, rubber- and plastic-products assemblers	88	59.1
9312	Construction and maintenance labourers: roads, dams and similar constructions	60	70
7212	Welders and flamecutters	58	70.7
7311	Precision-instrument makers and repairers	51	72.5
7136	Plumbers and pipe fitters	55	72.7
8283	Electronic-equipment assemblers	155	80
zzzz	Not coded	91	82.4
7231	Motor vehicle mechanics and fitters	72	83.3
8232	Plastic-products machine operators	93	83.9
7133	Plasterers	76	85.5
3471	Decorators and commercial designers	74	93.2
8262	Weaving- and knitting-machine operators	76	93.4
3152	Safety, health and quality inspectors	174	93.7
1222	Production and operations department managers in manufacturing	71	94.4
4133	Transport clerks	61	95.1
4131	Stock clerks	166	95.2
1239	Other department managers not elsewhere classified	121	96.7
6153	Deep-sea fishery workers	61	96.7
1223	Production and operations department managers in construction	63	96.8
9322	Hand packers and other manufacturing labourers	244	97.1
1233	Sales and marketing department managers	81	97.5
3439	Administrative associate professionals not elsewhere classified	132	97.7
7432	Weavers, knitters and related workers	90	97.8
3416	Buyers	99	98
4222	Receptionists and information clerks	116	98.3
8263	Sewing machine operators	289	98.3
3415	Technical and commercial sales representatives	122	98.4
4142	Mail carriers and sorting clerks	62	98.4
7433	Tailors, dressmakers and hatters	213	98.6
9333	Freight handlers	143	98.6

0112	Armed forces	177	98.9
5162	Police officers	90	98.9
4111	Stenographers and typists	333	99.1
4121	Accounting and bookkeeping clerks	106	99.1
5220	Shop salespersons and demonstrators	545	99.1
4115	Secretaries	164	99.4
6111	Field crop and vegetable growers	390	99.5
5169	Protective services workers not elsewhere classified	224	99.6
9132	Helpers and cleaners in offices, hotels and other establishments	320	99.7
1210	Directors and chief executives	405	99.8
1224	Production and operations department managers in wholesale and retail trade	71	100
1227	Production and operations department managers in business-services	68	100
2230	Nursing and midwifery professionals	146	100
2320	Secondary education teaching professionals	87	100
2331	Primary education teaching professionals	99	100
2411	Accountants	75	100
3340	Other teaching associate professionals	61	100
3433	Bookkeepers	194	100
4100	Office clerks	142	100
4112	Word-processor and related operators	66	100
4211	Cashiers and ticket clerks	131	100
5122	Cooks	344	100
5123	Waiters, waitresses and bartenders	441	100
5141	Hairdressers, barbers, beauticians and related workers	107	100
5230	Stall and market salespersons	128	100
6110	Market gardeners and crop growers	92	100
8322	Car, taxi and van drivers	246	100
8323	Bus and tram drivers	103	100
8324	Heavy truck and lorry drivers	195	100
9133	Hand-launderers and pressers	86	100
9152	Doorkeepers, watchpersons and related workers	62	100

1. as coded by centre 3 (Hong Kong)
2. % agree = percent agreement

Supplementary Table 5: Percent agreement for JEM-assessed exposure to aromatic solvents for frequently reported jobs (N ≥50) in the Traditional Chinese region

ISCO-88 code	Job description	N	% agree
8223	Metal finishing-, plating- and coating-machine operators	52	25
7241	Electrical mechanics and fitters	227	57.3
8284	Metal-, rubber- and plastic-products assemblers	88	59.1
8290	Other machine operators and assemblers	102	62.7
7311	Precision-instrument makers and repairers	51	64.7
8283	Electronic-equipment assemblers	155	69
9312	Construction and maintenance labourers: roads, dams and similar constructions	60	71.7
7433	Tailors, dressmakers and hatters	213	71.8
8232	Plastic-products machine operators	93	74.2
3471	Decorators and commercial designers	74	75.7
7136	Plumbers and pipe fitters	55	80
zzzz	Not coded	91	85.7
7133	Plasterers	76	86.8
3114	Electronics and telecommunications engineering technicians	67	89.6
1222	Production and operations department managers in manufacturing	71	90.1
2144	Electronics and telecommunications engineers	53	90.6
3152	Safety, health and quality inspectors	174	92
8262	Weaving- and knitting-machine operators	76	92.1
1223	Production and operations department managers in construction	63	93.7
8263	Sewing machine operators	289	94.1
7231	Motor vehicle mechanics and fitters	72	94.4
3416	Buyers	99	94.9
1239	Other department managers not elsewhere classified	121	95
4133	Transport clerks	61	95.1
6111	Field crop and vegetable growers	390	95.4
7432	Weavers, knitters and related workers	90	95.6
4131	Stock clerks	166	95.8
7212	Welders and flamecutters	58	96.6
6153	Deep-sea fishery workers	61	96.7
9322	Hand packers and other manufacturing labourers	244	97.1
1233	Sales and marketing department managers	81	97.5
3439	Administrative associate professionals not elsewhere classified	132	97.7
6110	Market gardeners and crop growers	92	97.8
0110	Armed forces	177	98.3
4142	Mail carriers and sorting clerks	62	98.4

1210	Directors and chief executives	405	98.5
4111	Stenographers and typists	333	98.5
5162	Police officers	90	98.9
4121	Accounting and bookkeeping clerks	106	99.1
4222	Receptionists and information clerks	116	99.1
3415	Technical and commercial sales representatives	122	99.2
2230	Nursing and midwifery professionals	146	99.3
9333	Freight handlers	143	99.3
5220	Shop salespersons and demonstrators	545	99.4
3433	Bookkeepers	194	99.5
5169	Protective services workers not elsewhere classified	224	99.6
5122	Cooks	344	99.7
9132	Helpers and cleaners in offices, hotels and other establishments	320	99.7
5123	Waiters, waitresses and bartenders	441	99.8
1224	Production and operations department managers in wholesale and retail trade	71	100
1227	Production and operations department managers in business services	68	100
2320	Secondary education teaching professionals	87	100
2331	Primary education teaching professionals	99	100
2411	Accountants	75	100
3340	Other teaching associate professionals	61	100
4100	Office clerks	142	100
4112	Word-processor and related operators	66	100
4115	Secretaries	164	100
4211	Cashiers and ticket clerks	131	100
5141	Hairdressers, barbers, beauticians and related workers	107	100
5230	Stall and market salespersons	128	100
8322	Car, taxi and van drivers	246	100
8323	Bus and tram drivers	103	100
8324	Heavy truck and lorry drivers	195	100
9133	Hand-launderers and pressers	86	100
9152	Doorkeepers, watchpersons and related workers	62	100

1. as coded by centre 3 (Hong Kong)
2. % agree = percent agreement

Supplementary Table 6: Percent agreement for JEM-assessed exposure to vapours, dusts, gases, and fumes for frequently reported jobs (N ≥50) in the Traditional Chinese region

ISCO-88 code	Job description	N	% agree
4133	Transport clerks	61	23
3114	Electronics and telecommunications engineering technicians	67	32.8
7432	Weavers, knitters and related workers	90	38.9
zzzz	Not coded	91	50.5
8324	Heavy truck and lorry drivers	195	51.3
8223	Metal finishing-, plating- and coating-machine operators	52	51.9
9333	Freight handlers	143	67.1
7241	Electrical mechanics and fitters	227	70
7136	Plumbers and pipe fitters	55	70.9
4131	Stock clerks	166	72.3
8262	Weaving- and knitting-machine operators	76	72.4
3340	Other teaching associate professionals	61	73.8
7212	Welders and flamecutters	58	74.1
4142	Mail carriers and sorting clerks	62	74.2
3471	Decorators and commercial designers	74	74.3
8232	Plastic-products machine operators	93	77.4
2144	Electronics and telecommunications engineers	53	79.2
1223	Production and operations department managers in construction	63	79.4
8290	Other machine operators and assemblers	102	79.4
7311	Precision-instrument makers and repairers	51	80.4
8284	Metal-, rubber- and plastic-products assemblers	88	80.7
5162	Police officers	90	81.1
8323	Bus and tram drivers	103	81.6
9132	Helpers and cleaners in offices, hotels and other establishments	320	81.6
7433	Tailors, dressmakers and hatters	213	81.7
3152	Safety, health and quality inspectors	174	83.3
9312	Construction and maintenance labourers: roads, dams and similar constructions	60	83.3
5230	Stall and market salespersons	128	83.6
3439	Administrative associate professionals not elsewhere classified	132	84.1
1222	Production and operations department managers in manufacturing	71	84.5
7231	Motor vehicle mechanics and fitters	72	86.1
3416	Buyers	99	86.9
1233	Sales and marketing department managers	81	87.7
8283	Electronic-equipment assemblers	155	87.7
8263	Sewing machine operators	289	87.9

5123	Waiters, waitresses and bartenders	441	88
7133	Plasterers	76	88.2
1239	Other department managers not elsewhere classified	121	88.4
5122	Cooks	344	88.7
9133	Hand-launderers and pressers	86	90.7
4222	Receptionists and information clerks	116	91.4
4100	Office clerks	142	92.3
1224	Production and operations department managers in wholesale and retail trade	71	93
9322	Hand packers and other manufacturing labourers	244	93
6153	Deep-sea fishery workers	61	93.4
1210	Directors and chief executives	405	93.6
5220	Shop salespersons and demonstrators	545	93.6
4111	Stenographers and typists	333	93.7
0111	Armed forces	177	93.8
2230	Nursing and midwifery professionals	146	94.5
4115	Secretaries	164	94.5
4211	Cashiers and ticket clerks	131	94.7
8322	Car, taxi and van drivers	246	95.5
3415	Technical and commercial sales representatives	122	95.9
2411	Accountants	75	96
5169	Protective services workers not elsewhere classified	224	96
2320	Secondary education teaching professionals	87	96.6
6111	Field crop and vegetable growers	390	96.7
9152	Doorkeepers, watchpersons and related workers	62	96.8
4112	Word-processor and related operators	66	97
1227	Production and operations department managers in business-services	68	97.1
3433	Bookkeepers	194	98.5
2331	Primary education teaching professionals	99	99
4121	Accounting and bookkeeping clerks	106	99.1
5141	Hairdressers, barbers, beauticians and related workers	107	99.1
6110	Market gardeners and crop growers	92	100

1. as coded by centre 3 (Hong Kong)
2. % agree = percent agreement

Supplementary Table 7: Proportions of generic job codes* in phase 1 versus 2 of job coding for different study centres

Centres	Phase 1	Phase 2
1 – Chengdu	29.4%	32.8%
2 – Tianjin	22.3%	27.6%
3 – Hong Kong	8.4%	14.9%
4 – Taiwan	9.0%	9.7%

*generic job codes are defined as any 3, 2, 1 digit ISCO-88 job codes plus the following 4-digit codes that describes “other, not elsewhere classified” occupations: 1239; 1319; 2139; 2149; 2229; 2359; 2419; 2429; 3119; 3139; 3229; 3340; 3419; 3429; 3439; 3449; 4190; 5139; 5149; 5169; 6129; 7129; 8139; 8159; 8229; 8269; 8290.

Supplementary Figures 1-4: Screen captures of the occupational history questions on the computer assisted personal interview (CAPI) device used during interview

1. First question in the occupational history section.

D. OCCUPATIONAL HISTORY

Now I'd like to ask you some questions about the kind of work you have done.
We are interested in every job, at home, or outside the home, part-time or full-time, paid or unpaid, including work on a farm, any self-employment, or work for companies or family businesses (excluding housewife), which you held for a total of 12 months or longer since you first began working.

1D.1	Are you currently employed, not employed, or retired?	○ EMPLOYED ○ NOT EMPLOYED ○ RETIRED

Provide Extra Comments

<< >>

2. Ever employed question.

1D.3	Did you ever have any jobs, held for a total of 12 months or longer, either outside the home or at home (?)...	○ Yes ○ No ○ Don't Know

Provide Extra Comments

<< >>

3. Example of completed job history grid

If you held more than one job at a company (or at home), or more than one job at the same time, we would like to talk about each job separately. Also, please include any seasonal work and any time while in the military. Let's begin by listing only the employer name, job title, and years worked at each of these jobs.

jobHistory Grid		EMPLOYER-1D.4 What was the name of the employer or workplace where you (first/next) worked for a total of 12 months or longer?	JOB TITLE-1D.5 What was the job title of the (first/next) job you held for 12 months or longer at (EMPLOYER-1D.4)?	START-1D.6 When did you start working as a (JOB TITLE-1D.5)? How old were you or what year was it?		STOP 1D.7 When did you stop working as a (JOB TITLE-1D.5) at (EMPLOYER-1D.4)? How old were you or what year was it?	
				Age, OR	Year	Age, OR	Year
	1	Happy Valley Farms	farmer	18		24	
	2	ABC Toy Company	plastic toy assembler	25		35	
	3	Protect Insurance Company	secretary	36		37	
	4	Hope Hospital	registered nurse	38		51	
	5						
	6						

Do you have any more jobs to add? If yes, please click here!!

Confirm	So just to confirm, the most recent job you held ended in [last Age, Year (ID.7) in grid]. Is that correct?	◉ Yes ○ No

Provide Extra Comments

<< >>

4. Additional questions asked of each job, with example answers

1D.8	When you worked at **Happy Valley Farms** from **18** to **24** what did they make, or what service did they provide?	grew vegetables
1D.9	How many months per year did you usually work on this job?	9 MONTHS PER YEAR
1D.10	On average, how many days per week did you work on this job?	6 DAYS PER WEEK
1D.11	On average, about many hours per day did you work on this job?	10 HOURS PER DAY
1D.12	In this job, on average, about how many hours did you spend outdoors on a normal working day on this job?	8 HOURS PER DAY OUTDOORS
1D.13	What were your main activities or duties as a **farmer** at **Happy Valley Farms**?	hoeing, planting, harvesting crops
1D.14	In this job, did you ever use paints, stains or varnishes or work in an area where they were used?	○ Yes ◉ No ○ Don't Know
1D.15	In this job, did you ever use solvents, glues, degreasing agents (to clean metal parts), gasoline or other fuels, or work in an area where they were used?	○ Yes ◉ No ○ Don't Know
1D.16	In this job, did you ever use particle board, plywood, or veneered woods or work in an area where they were used?	○ Yes ◉ No ○ Don't Know

Provide Extra Comments

<< >>

CHAPTER 5

Occupational Exposure to Benzene and Mortality Risk of Lymphohaematopoietic cancers in the Swiss National Cohort

C Ge[1]; A Spoerri[2]; M Egger[2,3]; N Rothman[4]; Q Lan[4]; A Huss[1,2]; R Vermeulen[1,5]; For the Swiss National Cohort

1. Institute for Risk Assessment Sciences, Utrecht University, The Netherlands;
2. Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland;
3. School of Social and Community Medicine, University of Bristol, UK;
4. Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD, USA;
5. Julius Centre for Public Health Sciences and Primary Care, University Medical Centre, Utrecht, The Netherlands

To be submitted

ABSTRACT

Objectives: Previous studies established a causal relationship between occupational benzene exposure and acute myeloid leukaemia (AML). However, mixed results have been reported for associations between benzene exposure and other myeloid and lymphoid malignancies. Our work aimed to examine whether occupational benzene exposure is associated with increased mortality from overall lymphohaematopoietic (LH) cancer and major subtypes.

Methods: Mortality records were linked to a Swiss census-based cohort from two national censuses in 1990 and 2000. Cases were defined as having any LH cancers registered in death certificates. We performed occupational exposure assessment by applying a quantitative benzene job-exposure matrix (BEN-JEM) to census-reported occupations. Exposure was calculated as the products of exposure proportions and levels (P x L). Cox proportional hazards models were used to calculate LH cancer death hazard ratios (HRs) and 95% confidence intervals (CI) associated with benzene exposure, both continuously and in ordinal categories.

Results: Our study included approximately 2.97 million persons and 13,415 LH cancer cases. We observed increased mortality risks per unit (P x L) increase in continuous benzene exposure for AML (HR 1.03; 95% CI 1.00, 1.06) and diffuse large B-cell lymphoma (DLBCL; HR 1.09; 95% CI 1.04, 1.14). When exposure was assessed categorically, increasing trends in risks were observed with increasing benzene exposure for AML (p=0.04), DLBCL (p=0.02), and follicular lymphoma (FL; p=0.05).

Conclusion: In a national cohort from Switzerland we found that occupational exposure to benzene is associated with elevated mortality risks for AML, DLBCL, and possibly FL.

INTRODUCTION

Benzene is a ubiquitous air pollutant. In the workplace, benzene exposure occurs in industries such as oil and gas extraction, refinement of petroleum products, and shoe production (Loomis et al., 2017). In Europe and Canada, studies have estimated that approximately 1-2% of the total working population were exposed to benzene in the workplace (Kauppinen et al., 2000; Peters et al., 2014). In China, around 500,000 workers in 2000 were employed in the shoemaking industry alone, where exposure levels were usually above the Chinese occupational benzene exposure limit of 40 mg/m^3 from 1979-2001 (Huang, 2000; L. Wang et al., 2006).

The International Agency for Research on Cancer (IARC) classifies benzene as a group 1 carcinogen in humans (IARC, 2018; Loomis et al., 2017). Association between occupational benzene exposure and acute myeloid leukaemia (AML) was observed in a number of studies, including recent industrial cohort studies with quantitative exposure estimates (Schnatter et al. 2012; Stenehjem et al. 2015; Linet et al. 2018). However, the relationship between benzene exposure and other lymphohaematopoietic (LH) cancers is less clear (IARC, 2018; Loomis et al., 2017).

For the present study, we assessed historical occupational benzene exposure for more than 2.97 million persons in the Swiss National Cohort (SNC) using a quantitative benzene job exposure matrix (JEM) and evaluated the associated mortality risks from overall lymphohaematopoietic cancer and major subtypes, including AML, chronic myeloid leukaemia (CML), non-Hodgkin lymphoma (NHL), acute lymphoid leukaemia (ALL), chronic lymphoid leukaemia (CLL), diffuse large B-cell lymphoma (DLBCL), small cell B-cell lymphoma (SCBCL), mantle cell lymphoma (MCL), T/NK-cell lymphoma (T/NK-NHL), and follicular lymphoma (FL), multiple myeloma (MM), and Hodgkin lymphoma (HL).

METHODS

Study population

The SNC is a longitudinal study based on two national censuses in Switzerland in 1990 and 2000. Records from the censuses were linked with mortality and emigration records using deterministic and probabilistic methods (Bopp et al., 2009; Spoerri et al., 2010). Key variables used in the linkages include sex, date of birth, marital status, nationality, religion, place of residence, spousal information, and family structure. We excluded persons below the age of 30 from our analysis because 1) census linkage for these subjects was less complete due to their higher mobility and likelihood of living alone (Bopp et al., 2009), and 2) mortality in these subjects were low.

Standardized registration of deaths and participation in the national census were both mandatory in Switzerland. For instance, for the year 2000, the census coverage was 98.6% (Renaud, 2004) and 94% of all deaths could be linked to census records. The SNC study obtained ethics approval from the Cantonal Ethics Committees of Bern. Additional information on the cohort and study is available online at www.swissnationalcohort.ch.

Exposure assessment and case definition

Census participants, by self-enumeration, reported approximately 18,000 distinct jobs in the two Swiss censuses. The SNC dataset already included standardized occupation coding using the 1988 version of the International Standard Classification of Occupations (ISCO-88), which was assigned by the Swiss Federal Office and reviewed by one of the authors (AH). Occupations reported by census participants when they entered the risk period (either in 1990 or 2000) were used for exposure assessment. We estimated occupational exposure to benzene using the BEN-JEM, a quantitative general population JEM developed by one of the authors (RV) (Spycher et al., 2017). For each ISCO-88 occupation, the BEN-JEM assesses the proportion of workers exposed (P) and the mean level of benzene exposure (L) in parts per million (ppm). The BEN-JEM also assigns different proportions and levels of exposure for different time periods, taking into account the decreasing trend of occupational benzene exposure in Europe and North America over time (Creely et al., 2007). Relevant BEN-JEM assessment periods for our study include 1985-94 and 1998-2000. Exposure for a particular job was calculated as the products of exposure proportions and levels (P x L). We also assessed exposure with ordinal exposure categories. Benzene exposures were categorized as low, medium, and high exposure when the P x L was <2, 2-10, and >10, respectively (Spycher et al., 2017). Jobs with zero exposure proportion or level were categorized as unexposed.

Cases were defined as deaths from LH neoplasms, as registered anywhere on death certificates. Versions 8 and 10 of the International Classification of Diseases and Related Health Problems (ICD8/10) were used to identify LH cancer categories and subtypes. Causes of death were recorded in Switzerland using ICD8 before 1995 and ICD10 since 1995. Disease categories and associated ICD codes used in our study are listed in Table 1.

Table 1: Lymphohaematopoietic cancer and subtypes of interest with corresponding ICD8/10 codes, as registered anywhere on death certificate

Disease category	ICD8	ICD10
Lymphohaematopoietic cancer	200-209	C81-C96
Myeloid leukaemia	205	C92
AML	205.0	C92.0
CML	205.1	C92.1
Lymphoid neoplasms	200-204	C81-C91
Non-Hodgkin Lymphoma	200, 202, 204	C82-C91
ALL	204.0	C91.0
CLL	204.1	C91.1
DLBCL		C83.3
SCBCL (including LPL, MZL, and WM)		C83.0 + C88.0
MCL		C83.1
T/NK-NHL		C84 + C86
FL	202.0	C82
Multiple myeloma	203	C90
Hodgkin lymphoma	201	C81

ICD8/10 = International Classification of Diseases and Related Health Problem Version8/10; AML = acute myeloid leukaemia; CML = chronic myeloid leukaemia; ALL = acute lymphoid leukaemia; CLL = chronic lymphoid leukaemia; DLBCL = diffuse large B-cell lymphoma; SCBCL = small cell B-cell lymphoma; LPL = lymphoplasmacytic lymphoma; MZL = marginal zone lymphoma; WM = Waldenström's macroglobulinemia; MCL = mantle cell leukaemia; T/NK-NHL = mature T/NK cell lymphoma; FL = follicular lymphoma.

Statistical analysis

We used Cox proportional hazard models to evaluate the associations between occupational benzene exposure and LH cancer-related deaths. Our models used participant age as the underlying time scale. For the main analysis, participants entered the risk set on December 4th 1990, the day of the 1990 census. Persons who were less than 30 years of age or who immigrated to Switzerland between December 4th 1990 and December 4th 2000 entered the risk set on December 4th 2000, the day of the 2000 census. Observation time ended on the earliest of the following: date of death, date of emigration, or December 31st 2016. For disease categories introduced in ICD10 (i.e. DLBCL, SCBCL, T/NK-NHL, and MCL), subjects entered the risk period on January 1st 1995, when ICD10 use began in Switzerland. We adjusted for sex in the minimally adjusted model and additionally adjusted for nationality, education, language region, and marital status in the full model. We stratified analyses to account for differences in baseline mortality risk over time ("strata" command in Stata).

For models with benzene exposure as a categorical variable, we assessed trends in the risk estimates by calculating the p-value in a likelihood ratio test after setting exposure categories to values 0, 1, 2, and 3. If the number of cases in the high exposure category was less than five, we combined cases in both the medium and high exposure categories into one "medium/high" exposure category. No risk estimates were calculated for health outcomes with less than five cases in any exposure categories. We tested models for proportionality assumption using statistical tests based on Schoenfeld residuals and the assumption was met for all exposure variables. All analyses were performed using Stata version 14 (StatCorp, College Station, Texas, USA).

RESULTS

The 1990 census included 3.055 million persons aged between 30 and retirement age (62 years for women and 65 years for men). We excluded 577,226 persons with missing or uncertain occupational information (i.e. ISCO-88 coded to major group only), 64,766 pensioners, and 255,251 people who were homemakers or unemployed, leaving 2.158 million persons for the analysis restricted to 1990–2000. The 2000 census had 3,470 million persons aged between 30 and retirement age. We excluded 836,270 persons with missing or uncertain occupational information, 100,947 persons on pension, and 498,414 people who were homemakers or unemployed, leaving 2.035 million persons for the analysis restricted to 2000–2016. For our final combined 1990–2016 analyses, a total of 2,976 million persons were included. Across our entire study population, 19,926 deaths from LH cancers occurred in the risk period, of which 98% could be linked to a census record. Of the 19,539 census-linked cases, we excluded 3,714 cases because they had uncertain occupational information, 1,106 cases because they were pensioners, and 1,304 cases because they were homemakers or unemployed. A total of 13,415 LH cases were included in our final analysis. Characteristics of our study population at baseline are shown in Table 2

Increased hazard ratios (HR) for AML were observed in both the minimally and fully adjusted continuous benzene exposure model (HR 1.03 per P x L unit exposure increase; 95% CI 1.00, 1.06 in both models) (Table 3). Hazard ratios for DLBCL also showed an associated with continuous benzene exposure (HR 1.08 per P x L unit exposure increase; 95% CI 1.04, 1.14 in both models).

Table 2. Study population characteristics at baseline (December 4th 1990)

	Occupational benzene exposure level[1]			
	None	**Low**	**Medium**	**High**
N	1,655,412	234,586	246,741	20,775
Mean entry age (SD)	45 (9)	45 (9)	45 (9)	44 (9)
Female (%)	43	34	29	2
Education level[2] (%)				
Low	18	36	23	19
Medium	53	45	63	66
High	29	19	14	15
Foreign nationals (%)	17	36	29	40
Language region (%)				
German	72	74	74	73
French	24	23	22	23
Italian	4	3	4	4
Marital status (%)				
Single	12	11	11	12
Married	72	76	76	75
Widowed	4	4	3	2
Divorced	12	10	10	11

[1] Exposure = sum of product of the proportions of workers exposed and level of exposure in ppm (P x L) across all occupations reported at censuses. Low, medium, and high exposure groups correspond to P x L values of <2, 2-10, and >10, respectively.
[2] Low = compulsory education or less; medium = upper secondary level education; high = tertiary level education.

Based on the fully adjusted models with categorical benzene exposure, increasing risk trends for AML (p=0.04), DLBCL (p=0.02), and FL (p=0.05) were observed with increasing benzene exposure (Table 4). Elevated risk point estimates were observed for AML for the medium (HR 1.11; 95% CI 0.97, 1.27) and high exposure group (HR 1.35; 95% CI 0.90, 2.03); however these increases were associated with larger uncertainties. HR for DLBCL was 1.22 (95% CI 0.94, 1.59) for the medium exposure group and 2.78 (95% CI 1.52, 5.09) for the high exposure group. We also observed clear increased risks for FL for the combined medium/high exposure group (HR 1.58; 95% CI 1.09, 2.29) and for HL for the low exposure group (HR 1.37; 95% CI 1.03, 1.82).

Elevated HR point estimates were observed for some other LH cancers in our fully adjusted models. HR for myeloid leukaemia was 1.09 (95% CI 0.96, 1.22) for the medium exposure group and 1.06 (95% CI 0.72, 1.57) for the high exposure group. For

CML, HR was 1.06 (95% CI 0.81, 1.39) for the medium exposure group and 1.08 (95% CI 0.44, 2.61) for the high exposure group.

Table 3. Hazard ratios for mortality from LH cancers and subtypes with benzene exposure[1] as a continuous variable for persons in the Swiss National Cohort

	Minimally adjusted[2] HR (95% CI)	Fully adjusted[2] HR (95% CI)
Any LH cancer	1.00 (0.99-1.02)	1.00 (0.99-1.01)
Myeloid leukaemia	1.02 (1.00-1.04)	1.01 (0.99-1.04)
AML	1.03 (1.00-1.06)	1.03 (1.00-1.06)
CML	1.02 (0.96-1.08)	1.01 (0.95-1.08)
Lymphoid neoplasms	1.00 (0.98-1.01)	1.00 (0.98-1.01)
NHL	1.00 (0.98-1.02)	1.00 (0.98-1.02)
ALL	0.96 (0.84-1.90)	0.96 (0.84-1. 09)
CLL	0.99 (0.95-1.03)	0.99 (0.95-1.03)
DLBCL	1.09 (1.04-1.14)	1.09 (1.04-1.14)
SCBCL (including LPL, MZL, and WM)	0. 90 (0.79-1.03)	0.89 (0.77-1.02)
MCL	0.83 (0.65-1.06)	0.84 (0.66-1.06)
T/NK-NHL	0.98 (0.91-1.06)	0.98 (0.91-1.06)
FL	1.06 (0.98-1.14)	1.05 (0.97-1.14)
MM	0.99 (0.96-1.02)	0.99 (0.96-1.02)
HL	1.04 (0.98-1.10)	1.04 (0.97-1.10)

[1] Exposure = sum of product of the proportions of workers exposed and level of exposure in ppm (P x L) across all occupations reported at censuses.
[2] Minimally adjusted model was adjusted for sex and fully adjusted model additionally adjusted for nationality, education, language region and marital status

LH = lymphohaematopoietic; AML = acute myeloid leukaemia; CML = chronic myeloid leukaemia; NHL = non-Hodgkin lymphoma; ALL = acute lymphoid leukaemia; CLL = chronic lymphoid leukaemia; DLBCL = diffuse large B-cell lymphoma; SCBCL = small cell B-cell lymphoma; LPL = lymphoplasmacytic lymphoma; MZL = marginal zone lymphoma; WM = Waldenström's macroglobulinemia; MCL = mantle cell leukaemia; T/NK-NHL = mature T/NK cell lymphoma; FL = follicular lymphoma; MM = multiple myeloma; and HL = Hodgkin lymphoma

Table 4. Hazard ratios for mortality from LH cancers and subtypes with benzene exposure[1] as a categorical variable for persons in the Swiss National Cohort

	N	Minimally adjusted[2] HR (95% CI)	Fully adjusted[2] HR (95% CI)	
Any LH cancer				
unexposed	10360	referent	referent	
low exposure	1333	1.00 (0.95-1.06)	1.00 (0.94-1.06)	
medium exposure	1602	1.03 (0.98-1.09)	1.01 (0.96-1.07)	
high exposure	120	1.00 (0.84-1.20)	0.99 (0.83-1.19)	
p-trend				0.32
Myeloid leukemias				
unexposed	2102	referent	referent	
low exposure	273	1.03 (0.91-1.17)	1.03 (0.90-1.17)	
medium exposure	337	1.10 (0.98-1.24)	1.09 (0.96-1.22)	
high exposure	26	1.09 (0.74-1.60)	1.06 (0.72-1.57)	
p-trend				0.11
AML				
unexposed	1554	referent	referent	
low exposure	200	1.03 (0.89-1.20)	1.04 (0.90-1.21)	
medium exposure	248	1.12 (0.98-1.28)	1.11 (0.97-1.27)	
high exposure	24	1.39 (0.93-2.08)	1.35 (0.90-2.03)	
p-trend				0.04
CML				
unexposed	385	referent	referent	
low exposure	52	1.03 (0.77-1.38)	1.01 (0.75-1.35)	
medium exposure	64	1.08 (0.83-1.41)	1.06 (0.81-1.39)	
high exposure	5	1.07 (0.44-2.58)	1.08 (0.44-2.61)	
p-trend				0.56
Lymphoid neoplasms				
unexposed	7853	referent	referent	
low exposure	1013	1.00 (0.94-1.07)	1.00 (0.94-1.07)	
medium exposure	1189	1.01 (0.95-1.07)	0.99 (0.93-1.05)	
high exposure	88	0.97 (0.78-1.19)	0.96 (0.78-1.18)	
p-trend				0.94
NHL				
unexposed	5206	referent	referent	
low exposure	666	0.99 (0.92-1.08)	0.99 (0.91-1.07)	
medium exposure	803	1.02 (0.94-1.10)	1.00 (0.92-1.08)	
high exposure	62	1.02 (0.79-1.31)	1.01 (0.78-1.29)	
p-trend				0.72

ALL				
unexposed	172	referent	referent	
low exposure	19	0.96 (0.59-1.54)	0.97 (0.60-1.58)	
medium/high exposure	22	0.93 (0.59-1.46)	0.92 (0.58-1.45)	
p-trend				0.88
CLL				
unexposed	1028	referent	referent	
low exposure	127	0.93 (0.77-1.11)	0.92 (0.76-1.11)	
medium exposure	181	1.07 (0.91-1.26)	1.06 (0.90-1.25)	
high exposure	12	0.93 (0.52-1.64)	0.93 (0.53-1.65)	
p-trend				0.70
DLBCL				
unexposed	389	referent	referent	
low exposure	47	0.98 (0.73-1.33)	0.99 (0.73-1.34)	
medium exposure	67	1.24 (0.96-1.62)	1.22 (0.94-1.59)	
high exposure	11	2.77 (1.51-5.06)	2.78 (1.52-5.09)	
p-trend				0.02
SCBCL (ncluding LPL, MZL, and WM)				
unexposed	237	referent	referent	
low exposure	32	1.02 (0.71-1.48)	1.00 (0.69-1.45)	
medium/high exposure	35	0.86 (0.60-1.24)	0.82 (0.57-1.17)	
p-trend				0.31
MCL				
unexposed	121	referent	referent	
low exposure	13	0.81 (0.45-1.43)	0.79 (0.44-1.41)	
Medium/high exposure	22	1.01 (0.64-1.59)	1.03 (0.65-1.64)	
p-trend				0.89
FL				
unexposed	159	referent	referent	
low exposure	16	0.83 (0.49-1.38)	0.83 (0.49-1.40)	
medium/high exposure	38	1.64 (1.14-2.36)	1.58 (1.09-2.29)	
p-trend				0.05
T/NK-NHL				
unexposed	321	referent	referent	
low exposure	49	1.23 (0.91-1.66)	1.24 (0.91-1.68)	
Medium/high exposure	48	0.94 (0.69-1.28)	0.94 (0.69-1.29)	
p-trend				0.92
MM				
unexposed	2350	referent	referent	
low exposure	293	0.97 (0.86-1.10)	0.97 (0.86-1.10)	
medium exposure	344	0.99 (0.88-1.11)	0.97 (0.87-1.09)	
high exposure	21	0.79 (0.51-1.21)	0.79 (0.51-1.22)	
p-trend				0.48
HL				
unexposed	328	referent	referent	
low exposure	57	1.37 (1.03-1.82)	1.37 (1.03-1.82)	
medium exposure	46	0.95 (0.70-1.30)	0.94 (0.69-1.29)	
high exposure	7	1.79 (0.84-3.79)	1.78 (0.84-3.78)	
p-trend				0.36

[1] Exposure = sum of product of the proportions of workers exposed and level of exposure in ppm (P x L) across all occupations reported at censuses. Low, medium, and high exposure groups correspond to P x L values of <2, 2-10, and >10, respectively.
[2] Minimally adjusted model was adjusted for sex and fully adjusted model additionally adjusted for nationality, education, language region and marital status

LH = lymphohaematopoietic; AML = acute myeloid leukaemia; CML = chronic myeloid leukaemia; NHL = non-Hodgkin lymphoma; ALL = acute lymphoid leukaemia; CLL = chronic lymphoid leukaemia; DLBCL = diffuse large B-cell lymphoma; SCBCL = small cell B-cell lymphoma; LPL = lymphoplasmacytic lymphoma; MZL = marginal zone lymphoma; WM = Waldenström's macroglobulinemia; MCL = mantle cell leukaemia; FL = follicular lymphoma; T/NK-NHL = mature T/NK cell lymphoma; MM = multiple myeloma; and HL = Hodgkin lymphoma

DISCUSSION

For myeloid leukaemia and its subtypes, the strongest association with benzene exposure was observed for AML. Increases in mortality risks were observed when benzene exposure was assessed both continuously and categorically. Our findings are consistent with the current consensus that benzene exposure is causally linked to AML (IARC, 2018; Loomis et al., 2017). For the other subtype of myeloid leukaemia, CML, we found elevated HR point estimates for persons in the medium and high exposure categories. Our results are similar to several studies on benzene-exposed workers in industrial cohorts that reported elevated risk point estimates for CML, but no clear results for an exposure-response relationship (Guénel et al., 2002; Linet et al., 2015, 2018; Rushton & Romaniuk, 1997). In a pooled analysis of benzene-exposed petroleum workers, clear increases in CML risks and an exposure-response trend were both reported (Schnatter et al. 2012).

Among lymphomas, increases in HR for DLBCL were associated with both continuous and categorical benzene exposure metrics. Our findings are consistent with results from a study with 40 exposed cases among a cohort of solvent-exposed female workers in the US, which reported increases in DLBCL risks with increases in benzene exposure intensity and probability (R. Wang et al., 2009). A study of similar design in Norwegian oil industry male workers, however, observed no association between benzene exposure and DLBCL based on 15 exposed cases (Stenehjem et al., 2015). Two population-based case-control studies also found no risk increases in DLBCL in relation to occupational benzene exposure. One of these studies was performed in a European cohort assembled from the Czech Republic, France, Germany, Ireland, Italy and Spain which included 28 exposed cases (Cocco et al., 2010); the other was from China with nine exposed cases (Wong et al., 2010). Our analyses based on categorical benzene exposure also suggest a possible relationship between benzene exposure and FL. However, we did not observe elevated FL risks in our continuous exposure models. In a hospital-based case-control study in Shanghai, Wong and colleagues (Wong et al., 2010) reported an odds ratio of 7.00 (95% CI 1.45, 33.70) based on seven exposed cases. Increases in FL risk point estimates were also reported by other studies performed in benzene exposed workers in industrial cohorts (Stenehjem et al., 2015; R. Wang et al., 2009) and in the general population (Cocco et al., 2010; Orsi et al., 2010).

Our results do not support an association between benzene exposure and overall NHL mortality risk. A key challenge in comparing results between different studies investigating risks of NHL is the heterogeneous classification of the disease category. Under the current 2008 World Health Organization (WHO) classification scheme, NHL includes subtypes such as ALL, CLL, and MM (Swerdlow et al., 2008). Historically, however, ALL and CLL were not classified as subtypes of NHL, which may lead to different findings across studies using different classification schemes. For instance, among three recent studies conducted in the Chinese population, two reported positive associations between benzene exposure and NHL (Bassig et al., 2015; Linet et al., 2015) and one reported no association (Wong et al., 2010). All three studies, however, used different definitions for NHL: Wong and colleagues included all lymphoid leukaemia (ALL; CLL) and MM cases; Bassig and colleagues included ALL and CLL but not MM cases; Linet and colleagues excluded all ALL, CLL, and MM cases (Bassig et al., 2015; Linet et al., 2015; Wong et al., 2010). Two other recent studies reported suggestive associations between benzene and combined B-cell NHL (Cocco et al., 2010; Stenehjem et al., 2015); however this disease category excludes T-cell and NK-cell lymphomas and thus may not be directly comparable to our results for NHL. Our findings suggest benzene exposure may be a subtype-specific risk factor for DLBCL and possibly for FL. Our study, along with several other studies which reported differential NHL subtype risks for benzene-exposed workers (Cocco et al., 2010; Orsi et al., 2010; Stenehjem et al., 2015; R. Wang et al., 2009; Wong et al., 2010), highlight the importance of assessing and reporting risk by disease subtype when possible, as there may be significant etiologic heterogeneity between different NHL subtypes. This is consistent with evidence supporting different known or suspected lifestyle risk factors for different types of NHL (Morton et al., 2014). In addition, there is abundant experimental and human evidence showing that benzene is lymphomagenic and can cause a variety of biological perturbations, including CD4+ T cell toxicity and an increase in lymphocyte chromosomal aberrations (IARC, 2018).

For HL, increased mortality risks were limited to the low benzene exposure group and not observed in our continuous exposure models. Earlier studies generally report no associated between benzene exposure and HL (Bloemen et al., 2004; Orsi et al., 2010). In a meta-analysis of lymphoma subtypes by Vlaanderen and colleagues, no association was observed between occupational benzene exposure and HL (Vlaanderen et al., 2011).

There are several limitations for our study. Occupations reported in a single census (either 1990 census occupation for subjects entering the risk period in 1990 or 2000 census occupation for subjects entering in 2000) were the only information available for the assessment of occupational benzene exposure. Therefore our study was unable to account for any changes in occupation prior to or after the census dates, leading to potential misclassification of exposure. This misclassification is, however, unlikely to be differential. Non-differential errors reduce the magnitude of the observed risk estimates. Errors in estimating individual exposure may have also occurred because exposure was assessed using the estimated exposure concentration and proportion of exposed workers in different standardized job categories in the JEM. Previous work investigating the biases created by aggregation of exposure level and probability in JEMs showed that the effects are dependent on a number of factors and difficult to predict (Burstyn et al., 2012). Finally, we were unable to account for lifestyle factors such as smoking and occupational exposures other than benzene. Other exposures, such as to chlorinated solvents, may confound our findings as they may be related to both benzene exposure and risk of LH cancers. However, the fact that the established association between benzene exposure and AML was observed in all our models indicates that the overall exposure assessment quality is sufficient and the limitations do not preclude our study from finding other exposure-disease associations (Vlaanderen et al., 2011).

Our study also has several strengths. The SNC is a large cohort that includes information on nearly three million eligible participants collected in two mandatory censuses. The size of the cohort allowed for analyses of 3,055 benzene-exposed LH cancer cases. For occupational studies investigating risks of rare LH cancer subtypes, our study is one of the largest in terms of exposed cases, including more than 100 exposed cases each of DLBCL and HL, and more than 40 exposed cases each of ALL, FL, and SCBCL. In addition, participation rates in the Swiss national censuses and successful linkage rates to death records were very high, thus our study population represents a nearly complete coverage of the Swiss population with minimal selection bias. Occupational exposure assessment was made with the application of an established quantitative JEM that has been applied in another study, where a relationship between maternal benzene exposure and childhood leukaemia was found (Spycher et al., 2017). The time axis is an important JEM feature for increasing the validity of exposure assessment, as patterns of industrial benzene use and exposure have changed in the last few decades, leading to lower exposure in most workplaces around the world (Creely et al., 2007; Koh et al., 2014).

Our work highlights the fact that registry-based occupation and disease information may be used, in conjunction with a generic JEM, to study occupational cancers in the general population. Our results provide evidence that occupational exposure to benzene is associated with elevated mortality risks for AML, DLBCL, and possibly FL in the Swiss population. Additional work is needed to replicate and extend these findings in other settings, particularly in studies with other populations and complete occupational histories for more detailed exposure assessment.

REFERENCES

Bassig, B. A., Friesen, M. C., Vermeulen, R., Shu, X.-O., Purdue, M. P., Stewart, P. A., Xiang, Y.-B., Chow, W.-H., Zheng, T., Ji, B.-T., Yang, G., Linet, M. S., Hu, W., Zhang, H., Zheng, W., Gao, Y.-T., Rothman, N., & Lan, Q. (2015). Occupational Exposure to Benzene and Non-Hodgkin Lymphoma in a Population-Based Cohort: The Shanghai Women's Health Study. *Environmental Health Perspectives*, *123*(10), 971–977. https://doi.org/10.1289/ehp.1408307

Bloemen, L., Youk, A., Bradley, T., Bodner, K., & Marsh, G. (2004). Lymphohaematopoietic cancer risk among chemical workers exposed to benzene. *Occupational and Environmental Medicine*, *61*(3), 270–274. https://doi.org/10.1136/oem.2003.007013

Bopp, M., Spoerri, A., Zwahlen, M., Gutzwiller, F., Paccaud, F., Braun-Fahrländer, C., Rougemont, A., & Egger, M. (2009). Cohort Profile: The Swiss National Cohort—a longitudinal study of 6.8 million people. *International Journal of Epidemiology*, *38*(2), 379–384. https://doi.org/10.1093/ije/dyn042

Burstyn, I., Lavoué, J., & Tongeren, M. V. (2012). Aggregation of Exposure Level and Probability into a Single Metric in Job-Exposure Matrices Creates Bias. *Annals of Occupational Hygiene*, *56*(9), 1038–1050. https://doi.org/10.1093/annhyg/mes031

Cocco, P., t'Mannetje, A., Fadda, D., Melis, M., Becker, N., Sanjosé, S. de, Foretova, L., Mareckova, J., Staines, A., Kleefeld, S., Maynadié, M., Nieters, A., Brennan, P., & Boffetta, P. (2010). Occupational exposure to solvents and risk of lymphoma subtypes: Results from the Epilymph case–control study. *Occupational and Environmental Medicine*, *67*(5), 341–347. https://doi.org/10.1136/oem.2009.046839

Creely, K. S., Cowie, H., Van Tongeren, M., Kromhout, H., Tickner, J., & Cherrie, J. W. (2007). Trends in inhalation exposure—A review of the data in the published scientific literature. *The Annals of Occupational Hygiene*, *51*(8), 665–678. https://doi.org/10.1093/annhyg/mem050

Guénel, P., Imbernon, E., Chevalier, A., Crinquand-Calastreng, A., & Goldberg, M. (2002). Leukemia in relation to occupational exposures to benzene and other agents: A case-control study nested in a cohort of gas and electric utility workers. *American Journal of Industrial Medicine*, *42*(2), 87–97. https://doi.org/10.1002/ajim.10090

Huang, P. (2000). Advances in adhesives used in the shoe industry (in Chinese). *Guangdong Chemical Engineering*, *1*, 13–16.

IARC. (2018). *Benzene; IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 120*. http://publications.iarc.fr/_publications/media/download/5074/69ba844af3f3018bc06e63dee-8a52c2e4494a771.pdf

Kauppinen, T., Toikkanen, J., Pedersen, D., Young, R., Ahrens, W., Boffetta, P., Hansen, J., Kromhout, H., Blasco, J. M., Mirabelli, D., Orden-Rivera, V. de la, Pannett, B., Plato, N., Savela, A., Vincent, R., & Kogevinas, M. (2000). Occupational exposure to carcinogens in the European Union. *Occupational and Environmental Medicine*, *57*(1), 10–18. https://doi.org/10.1136/oem.57.1.10

Koh, D.-H., Bhatti, P., Coble, J. B., Stewart, P. A., Lu, W., Shu, X.-O., Ji, B.-T., Xue, S., Locke, S. J., Portengen, L., Yang, G., Chow, W.-H., Gao, Y.-T., Rothman, N., Vermeulen, R., & Friesen, M. C. (2014). Calibrating a population-based job-exposure matrix using inspection measurements to estimate historical occupational exposure to lead for a population-based cohort in Shanghai, China. *Journal of Exposure Science and Environmental Epidemiology*, *24*(1), 9–16. https://doi.org/10.1038/jes.2012.86

Linet, M. S., Gilbert, E. S., Vermeulen, R., Dores, G. M., Yin, S.-N., Portengen, L., Hayes, R. B., Ji, B.-T., Lan, Q., Li, G.-L., Rothman, N., Chinese Center for Disease Control and Prevention–US National Cancer Institute Benzene Study Group, Ding, C.-Y., Dores, G. M., Gao, Y., Gilbert, E. S., Hayes, R. B., Ji, B.-T., Lan, Q., … Zhou, J.-S. (2018). Benzene Exposure Response and Risk of Myeloid Neoplasms in Chinese Workers: A Multicenter Case-Cohort Study. *Journal of the National Cancer Institute*. https://doi.org/10.1093/jnci/djy143

Linet, M. S., Yin, S.-N., Gilbert, E. S., Dores, G. M., Hayes, R. B., Vermeulen, R., Tian, H.-Y., Lan, Q., Portengen, L., Ji, B.-T., Li, G.-L., Rothman, N., & Chinese Center for Disease Control and Prevention-U.S. National Cancer Institute Benzene Study Group. (2015). A retrospective cohort study of cause-specific mortality and incidence of hematopoietic malignancies in Chinese benzene-exposed workers. *International Journal of Cancer. Journal International Du Cancer*, *137*(9), 2184–2197. https://doi.org/10.1002/ijc.29591

Loomis, D., Guyton, K. Z., Grosse, Y., Ghissassi, F. E., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Vilahur, N., Mattock, H., & Straif, K. (2017). Carcinogenicity of benzene. *The Lancet Oncology*, *18*(12), 1574–1575. https://doi.org/10.1016/S1470-2045(17)30832-X

Morton, L. M., Slager, S. L., Cerhan, J. R., Wang, S. S., Vajdic, C. M., Skibola, C. F., Bracci, P. M., de Sanjosé, S., Smedby, K. E., Chiu, B. C. H., Zhang, Y., Mbulaiteye, S. M., Monnereau, A., Turner, J. J., Clavel, J., Adami, H.-O., Chang, E. T., Glimelius, B., Hjalgrim, H., … Sampson, J. N. (2014). Etiologic heterogeneity among non-Hodgkin lymphoma subtypes: The InterLymph Non-Hodgkin Lymphoma Subtypes Project. *Journal of the National Cancer Institute. Monographs*, *2014*(48), 130–144. https://doi.org/10.1093/jncimonographs/lgu013

Orsi, L., Monnereau, A., Dananche, B., Berthou, C., Fenaux, P., Marit, G., Soubeyran, P., Huguet, F., Milpied, N., Leporrier, M., Hemon, D., Troussard, X., & Clavel, J. (2010). Occupational exposure to organic solvents and lymphoid neoplasms in men: Results of a French case-control study. *Occupational and Environmental Medicine*, *67*(10), 664–672. https://doi.org/10.1136/oem.2009.049460

Peters, C. E., Ge, C. B., Hall, A. L., Davies, H. W., & Demers, P. A. (2014). CAREX Canada: An enhanced model for assessing occupational carcinogen exposure. *Occupational and Environmental Medicine*, oemed-2014-102286. https://doi.org/10.1136/oemed-2014-102286

Renaud, A. (2004). *Methodology report: Coverage estimation for the Swiss population census 2000.* Swiss Federal Statistical Office.

Rushton, L., & Romaniuk, H. (1997). A case-control study to investigate the risk of leukaemia associated with exposure to benzene in petroleum marketing and distribution workers in the United Kingdom. *Occupational and Environmental Medicine*, *54*(3), 152–166.

Schnatter, A. R., Glass, D. C., Tang, G., Irons, R. D., & Rushton, L. (2012). Myelodysplastic Syndrome and Benzene Exposure Among Petroleum Workers: An International Pooled Analysis. *JNCI: Journal of the National Cancer Institute*, *104*(22), 1724–1737. https://doi.org/10.1093/jnci/djs411

Spoerri, A., Zwahlen, M., Egger, M., & Bopp, M. (2010). The Swiss National Cohort: A unique database for national and international researchers. *International Journal of Public Health*, *55*(4), 239–242. https://doi.org/10.1007/s00038-010-0160-5

Spycher, B. D., Lupatsch, J. E., Huss, A., Rischewski, J., Schindera, C., Spoerri, A., Vermeulen, R., & Kuehni, C. E. (2017). Parental occupational exposure to benzene and the risk of childhood cancer: A census-based cohort study. *Environment International*, *108*(Supplement C), 84–91. https://doi.org/10.1016/j.envint.2017.07.022

Stenehjem, J. S., Kjærheim, K., Bråtveit, M., Samuelsen, S. O., Barone-Adesi, F., Rothman, N., Lan, Q., Grimsrud, T. K., & UK Inflammatory Breast Cancer Working group. (2015). Benzene exposure and risk of lymphohaematopoietic cancers in 25 000 offshore oil industry workers. *British Journal of Cancer*, *112*(9), 1603–1612. https://doi.org/10.1038/bjc.2015.108

Swerdlow, S., Campo, E., Harris, N., Jaffe, E., Pileri, S., Stein, H., Thiele, J., & Vardiman, J. (2008). *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues.* International Agency for Research on Cancer.

Vlaanderen, J., Lan, Q., Kromhout, H., Rothman, N., & Vermeulen, R. (2011). Occupational Benzene Exposure and the Risk of Lymphoma Subtypes: A Meta-analysis of Cohort Studies Incorporating Three Study Quality Dimensions. *Environmental Health Perspectives*, *119*(2), 159–167. https://doi.org/10.1289/ehp.1002318

Wang, L., Zhou, Y., Liang, Y., Wong, O., Armstrong, T., Robert Schnatter, A., Wu, Q., Fang, J., Ye, X., Fu, H., & Irons, R. D. (2006). Benzene exposure in the shoemaking industry in China, a literature survey, 1978–2004. *Regulatory Toxicology and Pharmacology*, *46*(2), 149–156. https://doi.org/10.1016/j.yrtph.2006.06.009

Wang, R., Zhang, Y., Lan, Q., Holford, T. R., Leaderer, B., Zahm, S. H., Boyle, P., Dosemeci, M., Rothman, N., Zhu, Y., Qin, Q., & Zheng, T. (2009). Occupational exposure to solvents and risk of non-Hodgkin lymphoma in Connecticut women. *American Journal of Epidemiology*, *169*(2), 176–185. https://doi.org/10.1093/aje/kwn300

Wong, O., Harris, F., Armstrong, T. W., & Hua, F. (2010). A hospital-based case–control study of non-Hodgkin lymphoid neoplasms in Shanghai: Analysis of environmental and occupational risk factors by subtypes of the WHO classification. *Chemico-Biological Interactions*, *184*(1), 129–146. https://doi.org/10.1016/j.cbi.2009.10.016

CHAPTER 6

Respirable Crystalline Silica Exposure, Smoking, and Lung Cancer Subtype Risks: A Pooled Analysis of Case–Control Studies

C Ge[1], S Peters[1], A Olsson[2], L Portengen[1], J Schüz[2], J Almansa[1], T Behrens[3], B Pesch[3], B Kendzia[3], W Ahrens[4], B Bencko[5], S Benhamou[6], P Boffetta[7,8], B Bueno-de-Mesquita[9], N Caporaso[10], D Consonni[11], P Demers[12], E Fabiánová[13,14], G Fernández-Tardón[15], J Field[16], F Forastiere[17], L Foretova[18], P Guénel[19], P Gustavsson[20], V Ho[21], V Janout[22], KH Jöckel[23], S Karrasch[24,25,26], MT Landi[10], J Lissowska[27], D Luce[28], D Mates[29], J McLaughlin[30], F Merletti[31], D Mirabelli[31], N Plato[20], H Pohlabeln[4], L Richiardi[31], P Rudnai[32], J Siemiatycki[21], B Świątkowska[33], A Tardón[15], HE Wichmann[34,35], D Zaridze[36], T Brüning[3], K Straif[2], H Kromhout[1], and R Vermeulen[1]

Published: *American Journal of Respiratory and Critical Care Medicine*, 2020, Vol. 202, Iss. 3, 412–421.

doi: 10.1164/rccm.201910-1926OC

Affiliations

1. Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands;
2. International Agency for Research on Cancer (IARC), World Health Organization (WHO), Lyon, France;
3. Institute for Prevention and Occupational Medicine of the German Social Accident Insurance-Institute of the Ruhr University, Bochum, Germany;
4. Leibniz Institute for Prevention Research and Epidemiology-Bremen Institute for Prevention Research and Social Medicine (BIPS), Bremen, Germany;
5. Institute of Hygiene and Epidemiology, First Faculty of Medicine, Charles University, Prague, Czech Republic;
6. Inserm Unit 1018, Villejuif, France;
7. Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York;
8. Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy;
9. The National Institute for Public Health and Environmental Protection, Bilthoven, the Netherlands;
10. National Cancer Institute, Bethesda, Maryland;
11. Unitàdi epidemiologia, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy;
12. Occupational Cancer Research Centre, Cancer Care Ontario, Toronto, Ontario, Canada;
13. Regional Authority of Public Health, Banská Bystrica, Slovakia;
14. Faculty of Health, Catholic University, Ružomberok, Slovakia;
15. Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Institute of Health Research of the Principality of Asturias-Foundation for Biosanitary Research of Asturias (ISPA-FINBA), Faculty of Medicine, University of Oviedo, Oviedo, Spain;
16. Roy Castle Lung Cancer Research Programme, Cancer Research Centre, University of Liverpool, Liverpool, United Kingdom;
17. Consiglio Nazionale delle Ricerche-Istituto per la Ricerca e l'Innovazione Biomedica (CNR-Irib), Palermo, Italy;
18. Masaryk Memorial Cancer Institute, Brno, Czech Republic;
19. Center for Research in Epidemiology and Population Health (CESP), Team Exposome and Heredity, Inserm Unit 1018, University Paris-Saclay, Villejuif, France;
20. The Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden;
21. University of Montreal Hospital Research Centre, University of Montreal, Montreal, Quebec, Canada;
22. Faculty of Health Sciences, Palacky University, Olomouc, Czech Republic;
23. Institute for Medical Informatics, Biometry, and Epidemiology, University of Duisburg-Essen, Essen, Germany;
24. Institute and Outpatient Clinic for Occupational, Social, and Environmental Medicine, Inner City Clinic, University Hospital of Munich;
25. Institute of Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany;
26. Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research, Munich, Neuherberg, Germany;
27. The M. Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland;
28. Université de Rennes I, Inserm Unit 1085, École des hautes études en santé publique (EHESP), Institut de recherche en santé, environnement ettravail (Irset), Pointe-à-Pitre, France;
29. National Institute of Public Health, Bucharest, Romania;
30. Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada;
31. Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Torino, Italy;
32. National Public Health Center, Budapest, Hungary;
33. The Nofer Institute of Occupational Medicine, Lodz, Poland;
34. Institut für Medizinische Informatik Biometrie Epidemiologie, Ludwig-Maximilians-Universität, Munich, Germany;
35. Institut für Epidemiologie, Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg, Germany; and
36. Russian Cancer Research Centre, Moscow, Russia

ABSTRACT

Rationale: Millions of workers around the world are exposed to respirable crystalline silica. Although silica is a confirmed human lung carcinogen, little is known regarding the cancer risks associated with low levels of exposure and risks by cancer subtype.

Objectives: We aimed to address current knowledge gaps in lung cancer risks associated with low levels of occupational silica exposure and the joint effects of smoking and silica exposure on lung cancer risks.

Methods: Subjects from 14 case–control studies from Europe and Canada with detailed smoking and occupational histories were pooled. A quantitative job-exposure matrix was used to estimate silica exposure by occupation, time period, and geographical region. Logistic regression models were used to estimate exposure–disease associations and the joint effects of silica exposure and smoking on risk of lung cancer. Stratified analyses by smoking history and cancer subtypes were also performed.

Measurements and Main Results: Our study included 16,901 cases and 20,965 control subjects. Lung cancer odds ratios ranged from 1.15 (95% confidence interval, 1.04–1.27) to 1.45 (95% confidence interval, 1.31–1.60) for groups with the lowest and highest cumulative exposure, respectively. Increasing cumulative silica exposure was associated (P trend < 0.01) with increasing lung cancer risks in nonsilicotics and in current, former, and never-smokers. Increasing exposure was also associated (P trend ≤ 0.01) with increasing risks of lung adenocarcinoma, squamous cell carcinoma, and small cell carcinoma. Supermultiplicative interaction of silica exposure and smoking was observed on overall lung cancer risks; superadditive effects were observed in risks of lung cancer and all three included subtypes.

Conclusions: Silica exposure is associated with lung cancer at low exposure levels. An exposure–response relationship was robust and present regardless of smoking, silicosis status, and cancer subtype.

INTRODUCTION

Occupational exposure to respirable crystalline silica (silica hereafter) occurs in tens of millions of workers globally in a wide range of industries, including construction, mining, and quarrying, as well as manufacturing of bricks, ceramics, and metal products (Rushton, 2011; WHO, 2007). Silica is classified as a human lung carcinogen by the International Agency for Research on Cancer (IARC), the U.S. National Institute for Occupational Safety and Health, and the U.S. National Toxicology Program (IARC, 2012; NIOSH, 2002; NTP, 2016). A pooled analysis of 1,072 lung cancer cases from 10 industry-based studies showed that the risk of cancer increased monotonically with increases in cumulative silica exposure (Steenland et al., 2001). Additional evidence of an exposure–response relationship between silica and lung cancer was observed in different industrial cohorts (Liu et al., 2013; Sogl et al., 2012) as well as in case–control studies in different countries (Cassidy et al., 2007; De Matteis et al., 2012; Vida et al., 2010).

Despite the strong epidemiologic evidence of an exposure–response relationship between silica and lung cancer, questions still remain regarding certain aspects of the carcinogenicity of silica, including the role of cigarette smoking as a potential confounder and effect modifier (Steenland & Ward, 2014), whether an exposure threshold exists for silica-related lung cancer (Manno et al., 2018), whether silicosis is a prerequisite for developing silica-related lung cancer (Checkoway & Franzblau, 2000; Kurihara & Wada, 2004), the effect of silica exposure on risks of different histological subtypes of lung cancer (Cassidy et al., 2007; De Matteis et al., 2012), and the joint effect of exposure to silica and smoking on risk of lung cancer and its subtypes (Cassidy et al., 2007; Liu et al., 2013).

In the current study we present findings from the Pooled Analysis of Case-Control Studies on the Joint Effects of Occupational Carcinogens in the Development of Lung Cancer (SYNERGY) project, which is a pooled analysis of lung cancer case-control studies from Europe and Canada (Olsson et al., 2011). Occupational exposure to quartz silica was estimated via a quantitative general population job-exposure matrix (SYN-JEM) (Peters et al., 2016). The aims of our work were to assess *1*) the risks of lung cancer in relation to various indices of occupational silica exposure by cancer subtype, smoking status, and silicosis status; *2*) the interaction of silica exposure and smoking on the risk of lung cancer and its major subtypes on the additive and multiplicative scale; and *3*) the excess lifetime risks (ELRs) of lung cancer associated with different levels of occupational silica exposure.

METHODS

Study Population

The SYNERGY project is a pooled analysis of 14 population-and hospital-based lung cancer case–control studies in 13 European countries and Canada (*see* Supplementary Table 1). Detailed description of the population was presented elsewhere (Olsson et al., 2011). Lifetime occupational and smoking histories were available for all subjects. Self-reports of physician-diagnosed silicosis were collected in the AUT-Munich (Arbeit und Technik-Munich), EAGLE (Environment and Genetics in Lung Cancer Etiology), HdA (Humanisierung des Arbeitslebens), and INCO (International Agency for Research on Cancer Multicenter Case-Control Study of Occupation, Environment, and Lung Cancer in Central and Eastern Europe) studies by in-person or next-of-kin interview (full silicosis questions available in Supplementary Table 1). Ethical approvals for the SYNERGY project were obtained from all participating countries, as well as the IARC institutional review board. More information about the project is available at http://synergy.iarc.fr.

Exposure Assessment

The elaborated SYN-JEM and the underlying models for exposure to quartz silica have been described in detail elsewhere (Peters et al., 2013, 2016; Peters, Vermeulen, Portengen, et al., 2011). Briefly, 23,640 historical personal respirable quartz measurements were combined with exposure ratings from a general population JEM, the Domtoren-JEM (Peters, Vermeulen, Cassidy, et al., 2011). Quantitative quartz silica exposure estimates (in milligram per cubic meter) representing annual average exposure levels were derived for each job title, region, and year combination. Silica concentrations before 1960 were assumed to be the same as those in 1960. JEM linkage to the population was performed via the *International Standard Classification of Occupations* (version 1968, or ISCO-68) (ILO, 2010). Cumulative exposure (in milligram per cubic meter years) was calculated as the sum of the products of modeled exposure intensities and years of employment for all jobs over a subject's entire working history.

Statistical Analysis

The overall analysis protocol for silica is similar to those previously applied to characterize lung cancer risks for exposure to diesel engine exhaust and asbestos in the SYNERGY study (Olsson et al., 2011, 2017). Odds ratios (ORs) and 95% confidence intervals (CIs) for lung cancer associated with various categorical indices of occupational silica exposure were calculated using unconditional logistic regression models. Trend analysis *P* values were obtained by including the various indices of exposure as continuous variables in models for all subjects and for exposed subjects only. In our main categorical models, lung cancer risks were calculated for the following silica exposure metrics: ever/never exposure, duration of exposure (1–9, 10–19, 20–29, and >29 yr), time since last exposure (<5, 5–9, 10–19, 20–29, 30–39, and >39 yr), and cumulative exposure (quartiles of exposure distribution among control subjects: >0–0.39, 0.4–1.09, 1.1–2.39, and ≥2.4 mg/m^3-years). Adjustments were made for age group (<45, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and >74 yr), sex, study, smoking (log [cigarette pack-years + 1]), smoking cessation since interview/diagnosis (current smokers: >0–7, 8–15, 16–25, and >25 yr; never-smokers), and ever employment in "list A jobs." List A jobs are occupations with known occupational lung cancer risks (e.g., welders, long-distance truck drivers, or boiler operators), and their inclusion in the model served as an adjustment for exposures to other occupational lung carcinogens. The list was first published in 1982, then updated in 1995 and 2000 to include exposures reviewed by the IARC up to volume 75 of the *IARC Monographs on the Identification of Carcinogenic Hazards to Humans* (Ahrens & Merletti, 1998; Mirabelli et al., 2001). We defined smokers as subjects who smoked more than one cigarette per day for more than one year; pack-years was calculated as the sum of the products of smoking duration in years and average smoking of 20-cigarette packs per day.

Various silica cumulative exposure lag times (0, 5, 10, 15, and 20 yr) were applied in the main models, but only results with zero lag are presented because models with no lag had the best model fit according to minimized Akaike information criterion values. Stratified analyses for cancer risks associated with cumulative exposure categories were also calculated for subjects with different major cancer subtypes, without reported silicosis, and with different smoking habits.

For analyses of silica exposure as a continuous variable, both untransformed and natural log-transformed cumulative exposure were used. For the model with log-transformed exposure, nonexposed subjects were assigned two-thirds of the lowest cumulative exposure value among the exposed group (0.0036 mg/m^3-years). To further explore the shape of the exposure–response relationship, we performed thin-plate regression spline analyses as implemented in the R package mgcv (Wood, 2011), with relative maximum likelihood selected as the method for smoothing parameter estimation and total number of basis functions limited to three. The 95% CIs around the splines were based on simulations from posterior distributions of model coefficients with random draws from a multivariate normal distribution parameterized by the estimated mean vector and covariance matrix of the model coefficients. All splines were truncated at the 99th percentile to focus on results that were the most relevant and best supported by our exposure data.

Multiplicative interactions between silica exposure and smoking on risks of overall lung cancer and major cancer subtypes were assessed using *P* values from the cross-product interaction terms between silica exposure and smoking in the logistic models. For additive interactions, relative excess risks due to interaction (RERI) were calculated using ORs from the adjusted logistic models as defined by Rothman and Greenland (Rothman & Greenland, 1998) and implemented in the R package epiR (Stevenson et al., 2020).

The ELRs of lung cancer at age 80 years associated with 45 years of occupational silica exposure at various concentrations were calculated according to life table methods described by Vermeulen and colleagues (Vermeulen et al., 2014) with all-cause and lung cancer mortality rates from the European Union in 2008 as the referent (European Commission, 2008). Silica exposure levels for our ELR calculations were selected based on the current recommended 8-hour threshold limit value by the American Conference of Governmental Industrial Hygienists at 0.025 mg/m^3 (ACGIH, 2010), the recently updated 0.05 mg/m^3 permissible exposure limit from U.S. Occupational Safety and Health Administration (OSHA, 2017), and the exposure limit of 0.1 mg/m^3 in the latest European Union directive (2019/130) on the protection of workers from carcinogens(EU Parliament and Council, 2019).

Statistical analyses were conducted using SAS (version 9.3; SAS Institute) and R (version 3.5) (R Core Team, 2020).

RESULTS

After excluding participants with incomplete information on covariates (804 cases and 848 control subjects), 16,901 lung cancer cases (4,752 adenocarcinomas, 6,503 squamous cell carcinomas, 2,730 small cell carcinomas, 2,822 other/unspecified lung cancers, and 94 not available) and 20,965 control subjects remained for our main analyses (Table 1). Silicosis status was available in 50% of the study population (n = 18,931), among which 108 cases of silicosis were reported. Occupations with the highest modeled silica exposure concentrations in SYN-JEM were chimney bricklayers, stone cutters/carvers, and hand monument carvers; the most frequently reported exposed job titles among the control subjects in our population were farm helpers, general farmers, and construction bricklayers (more occupations in these categories are available in Supplementary Table 2).

Table 1. Selected Study Population Characteristics by Lung Cancer Status and Silica Exposure

	Ever Exposed to Silica				Never Exposed to Silica			
Characteristic	**Cases (*n*)**	**%**	**Control Subjects (*n*)**	**%**	**Cases (*n*)**	**%**	**Control Subjects (*n*)**	**%**
Sex								
M	4,649	94.4	4,140	92.2	8,956	74.8	12,311	74.7
F	274	5.6	348	7.8	3,022	25.2	4,166	25.3
Age group								
<45 yr	142	2.9	194	4.3	573	4.8	1,177	7.1
45–64 yr	2,503	50.8	2,055	45.8	6,260	52.3	8,299	50.4
>64 yr	2,278	46.3	2,239	49.9	5,145	43.0	7,001	42.5
Smoking status								
Never-smoker	248	5.0	1,253	27.9	1,121	9.4	5,900	35.8
Former smoker	1,736	35.3	2,010	44.8	3,696	30.9	6,210	37.7
Current smoker	2,939	59.7	1,225	27.3	7,161	59.8	4,367	26.5
Smoking pack-years								
Never-smoker	248	5.0	1,253	27.9	1,121	9.4	5,900	35.8
<10	227	4.6	683	15.2	582	4.9	2,386	14.5
10–19	475	9.6	598	13.3	1,127	9.4	2,264	13.7
>19	3,973	80.7	1,954	43.5	9,148	76.4	5,927	36.0
Years since quitting smoking								
Never-smoker	248	5.0	1,253	27.9	1,121	9.4	5,900	35.8
<0–7 yr	638	13.0	317	7.1	1,388	11.6	1,105	6.7
8–15 yr	494	10.0	461	10.3	1,037	8.7	1,437	8.7
16–25 yr	379	7.7	590	13.1	792	6.6	1,756	10.7
>25 yr	225	4.6	642	14.3	479	4.0	1,912	11.6
Current smoker	2,939	59.7	1,225	27.3	7,161	59.8	4,367	26.5
List A job								
Ever employment	829	16.8	597	13.3	958	8.0	767	4.7
Never employment	4,094	83.2	3,891	86.7	10,905	92.0	15,563	95.3
Silicosis								
Reported silicosis	57	1.2	33	0.7	13	0.1	5	0
No reported silicosis	2,882	58.5	2,311	51.5	6,091	50.9	7,539	45.8
Unknown	1,984	40.3	2,144	47.8	5,874	49.0	8,933	54.2
Lung cancer subtype								
Adenocarcinoma	1,069	21.7			3,683	30.7		
Small cell carcinoma	869	17.7			1,861	15.5		
Squamous cell carcinoma	2,251	45.7			4,252	35.5		
Other/unspecified	711	14.4			2,111	17.6		
Not available	23	0.5			71	0.6		

Increased overall lung cancer risks were observed for silica-exposed versus non-exposed subjects across three occupational exposure metrics, including ever exposure (OR, 1.30 [95% CI, 1.23–1.38]), longer exposure duration (longest category: >29 yr; OR, 1.48 [95% CI, 1.34–1.63]), and higher cumulative exposure (highest category: ≥2.4 mg/m^3-years; OR, 1.45 [95% CI, 1.31–1.60]) (Table 2). Elevated lung cancer risk increases were found for groups with the lowest exposure duration and cumulative exposure; ORs were 1.22 (95% CI, 1.12–1.31) and 1.15 (95% CI, 1.04–1.27) for subjects with exposure duration of one to nine years and cumulative exposure <0.4 mg/m^3-years, respectively. Increasing cancer risk trends were also associated (P trends < 0.01) with both increasing exposure duration and increasing cumulative exposure. Lung cancer risks for those who were more recently exposed also tended to be higher than for those who were last exposed a longer time ago, but confidence in this risk trend is lower (P trend = 0.10 among exposed subjects). Results for analyses restricted to subjects who did not report silicosis were similar to those from the main analyses (Table 3).

Increasing risks of all three included lung cancer subtypes were observed with increasing silica cumulative exposure (Table 4). We observed elevated risks of squamous cell carcinoma for all cumulative exposure groups, including the lowest (OR, 1.22 [95% CI, 1.06–1.39]). Clear increased risks of small cell carcinoma were found for groups with cumulative exposures >0.4 mg/m^3-years, with an OR of 1.70 (95% CI, 1.43–2.02) for the highest exposed group. Adenocarcinoma risks were generally lower than those observed in small cell and squamous cell carcinomas; adenocarcinoma OR for the highest exposed group was 1.17 (95% CI, 1.00–1.37).

The continuous model with untransformed exposure showed that every 1 mg/m^3-year increase in cumulative silica exposure increased lung cancer risk by a factor of 1.06 (95% CI, 1.04–1.08). In the model with log-transformed exposure, lung cancer risk increased by a factor of 1.05 (95% CI, 1.04–1.06) for every unit increase in log-cumulative exposure. Nonparametric spline analysis showed monotonic increases in risks of overall lung cancer and its subtypes for both untransformed and log-transformed silica cumulative exposure (Figure 1). Individual splines for overall lung cancer and all subtypes with corresponding 95% CI are available in Supplementary Figures 1 and 2.

Table 2. Lung Cancer Odds Ratios Associated with Various Indices of Occupational Silica Exposure

Occupational Silica Exposure	Cases (*n*)	%	Control Subjects (*n*)	%	OR*	95% CI
Never	11,978	70.9	16,477	78.6	1.0	Referent
Ever exposure	4,923	29.1	4,488	21.4	1.30	1.23–1.38
Duration, yr						
1–9	2,035	12.0	1,936	9.2	1.22	1.12–1.31
10–19	926	5.5	905	4.3	1.20	1.08–1.34
20–29	635	3.8	519	2.5	1.45	1.26–1.66
>29	1,327	7.9	1,128	5.5	1.48	1.34–1.63
Test for trend, P value					<0.01	
P value excluding never exposed					<0.01	
Cumulative exposure, mg/m^3-years						
>0–0.39	1,113	6.6	1,128	5.4	1.15	1.04–1.27
0.4–1.09	1,221	7.2	1,120	5.3	1.33	1.21–1.47
1.1–2.39	1,231	7.3	1,122	5.4	1.29	1.17–1.42
≥2.4	1,358	8.0	1,118	5.3	1.45	1.31–1.60
Test for trend, P value					<0.01	
P value excluding never exposed					<0.01	
Time since last exposure†, yr						
<5	934	5.5	815	3.9	1.43	1.18–1.73
5–9	462	2.7	351	1.7	1.43	1.15–1.77
10–19	679	4.0	569	2.7	1.36	1.13–1.63
20–29	617	3.7	536	2.6	1.26	1.08–1.47
30–39	931	5.5	812	3.9	1.30	1.15–1.47
>39	1,300	7.7	1,405	6.7	1.09	0.99–1.20
Test for trend, P value						
P value excluding never exposed					0.10	

Definition of abbreviations: CI = confidence interval; OR = odds ratio.
*OR adjusted for study, age group, sex, smoking (pack-years, time since quitting smoking), and list A jobs.
†OR in "time since last exposure" is additionally adjusted for duration (continuous) of silica exposure. Trend test limited to exposed subjects.

Table 3. Lung cancer odds ratios (OR) associated with cumulative occupational silica exposure in subjects without silicosis

Cumulative silica exposure *(mg/m^3-years)*	Cases (n)	OR*	95% CI
Never	6091	1.0	Referent
>0–0.39	665	1.22	1.07–1.40
0.4–1.09	720	1.50	1.31–1.71
1.1–2.39	757	1.48	1.30–1.69
≥2.4	740	1.42	1.25–1.63
Test for trend, p-value		<0.01	
Excl. never exposed		<0.01	

Definition of abbreviations: CI = confidence interval; OR = odds ratio.
*OR adjusted for study, age group, sex, smoking (pack-years, time-since-quitting smoking), and List A jobs

Table 4. Lung Cancer Major Subtype Risks Associated with Cumulative Occupational Silica Exposure

	Adenocarcinoma			Squamous Cell Carcinoma			Small Cell Carcinoma		
Cumulative Exposure (*mg/m³-years*)	**Cases (*n*)**	**OR***	**95% CI**	**Cases (*n*)**	**OR***	**95% CI**	**Cases (*n*)**	**OR***	**95% CI**
Never	3,683	1.0	Referent	4,252	1.0	Referent	1,861	1.0	Referent
>0–0.39	283	1.14	0.98–1.33	455	1.22	1.06–1.39	194	1.07	0.89–1.28
0.4–1.09	282	1.18	1.02–1.37	557	1.51	1.33–1.71	204	1.41	1.17–1.68
1.1–2.39	240	1.03	0.88–1.20	593	1.46	1.29–1.65	229	1.48	1.25–1.76
≥2.4	264	1.17	1.00–1.37	646	1.55	1.37–1.76	242	1.70	1.43–2.02
Test for trend, P value		0.01			<0.01			<0.01	
P value excluding never exposed		0.02			<0.01			<0.01	

Definition of abbreviations: CI = confidence interval; OR = odds ratio.
*OR adjusted for study, age group, sex, smoking (pack-years, time since quitting smoking), and list A jobs.

Stratified analyses showed that, regardless of smoking status, increasing cumulative silica exposure was associated (*P* trends for all subjects < 0.01) with increasing lung cancer risks (Table 5). Risks of lung cancer for different silica exposure groups were similar for former and current smokers, with ORs of 1.47 (95% CI, 1.27–1.70) and 1.39 (95% CI, 1.20–1.62) for the highest exposed group, respectively. For never-smokers, the OR point estimates for all silica cumulative exposure categories were above 1, with the highest exposed group having an OR of 1.40 (95% CI, 1.03–1.86).

Table 5. Lung Cancer Risks Associated with Cumulative Occupational Silica Exposure by Smoking Status

	Never-Smokers			Former Smokers			Current Smokers		
Cumulative exposure (*mg/m³-years*)	**Cases (*n*)**	**OR***	**95% CI**	**Cases (*n*)**	**OR†**	**95% CI**	**Cases (*n*)**	**OR‡**	**95% CI**
Never	1,121	1.0	Referent	3,696	1.0	Referent	7,161	1.0	Referent
>0–0.39	60	1.17	0.85–1.57	366	1.07	0.92–1.25	687	1.19	1.03–1.39
0.4–1.09	59	1.07	0.78–1.43	433	1.37	1.18–1.59	729	1.33	1.15–1.55
1.1–2.39	60	1.02	0.75–1.36	441	1.35	1.16–1.57	730	1.29	1.11–1.50
≥2.4	69	1.40	1.03–1.86	496	1.47	1.27–1.70	793	1.39	1.20–1.62
Test for trend, P value		<0.01			<0.01			<0.01	
P value excluding never exposed		0.02			<0.01			0.07	

Definition of abbreviations: CI = confidence interval; OR = odds ratio.
*OR adjusted for sex, study, age group, and list A jobs.
†OR adjusted for sex, study, age group, list A jobs, pack-years, and time since quitting smoking.
‡OR adjusted for sex, study, age group, list A jobs, and pack-years.

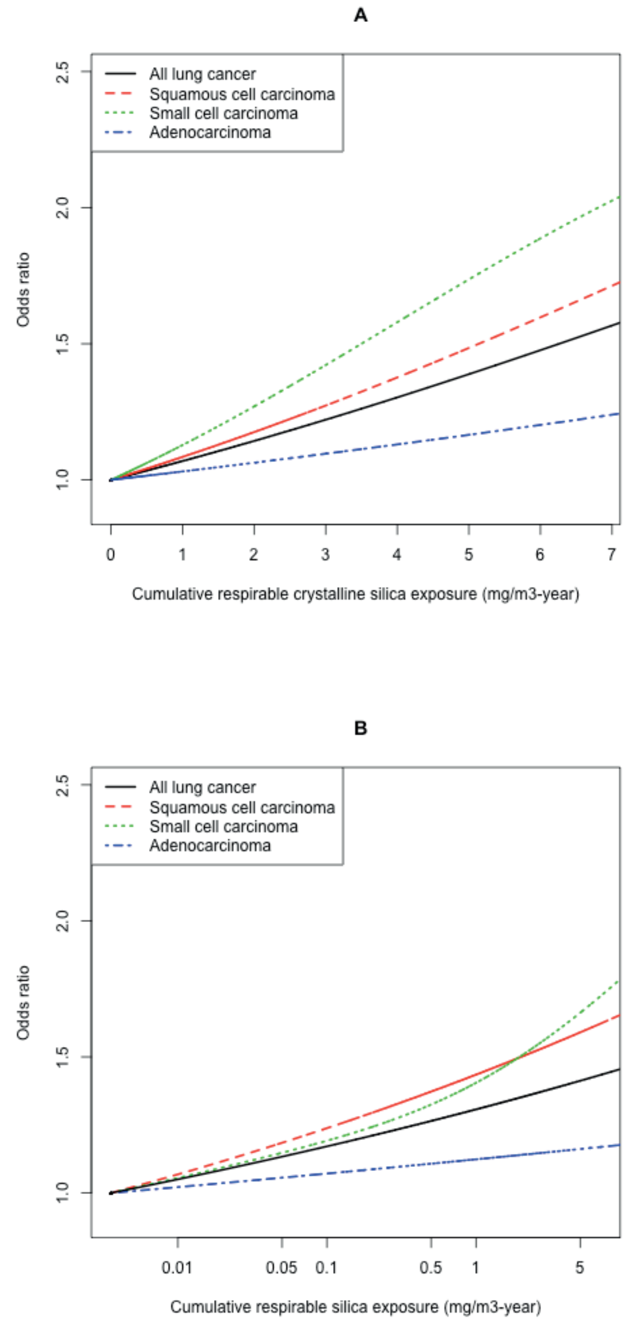


Figure 1. Spline analyses results on exposure–response relationships between lung cancer with (*A*) cumulative exposure and (*B*) natural log-transformed cumulative exposure.

Interactions beyond the additive model between smoking and occupational silica exposure were observed for overall lung cancer (RERI = 2.34 [95% CI, 1.85–2.83]), adenocarcinoma (RERI = 0.70 [95% CI, 0.26–1.15]), squamous cell carcinoma (RERI = 4.86 [95% CI, 3.63–6.09]), and small cell carcinoma (RERI = 5.13 [95% CI, 3.03–7.23]) (Tables 6 and 7). Supermultiplicative joint effect of smoking and silica exposure was observed on overall lung cancer risk ($P < 0.01$). OR point estimates also suggest supermultiplicative interactions for risks of adenocarcinoma and squamous cell carcinoma, though these effect estimates were associated with higher uncertainties ($P = 0.17$ and $P = 0.23$, respectively) because of smaller sample sizes.

Lung cancer ELRs were 0.22%, 0.45%, and 0.96% for workers exposed to 0.025, 0.05, and 0.1 mg/m^3 of silica, respectively.

Table 6. Interactions between Occupational Silica Exposure and Smoking for All Lung Cancers

	All Lung Cancers			
Exposure Status	**Control Sujects (*n*)**	**Cases (*n*)**	**OR***	**95% CI**
Never-smoker and never silica	5,900	1,121	1.0	Referent
Never-smoker and ever silica	1,253	248	1.02	0.87–1.19
Ever-smoker and never silica	10,577	10,857	6.37	5.91–6.87
Ever-smoker and ever silica	3,235	4,675	8.72	8.0–9.52
P value multiplicative interaction			<0.01	
RERI			2.34	1.85–2.83

Definition of abbreviations: CI = confidence interval; OR = odds ratio; RERI = relative excess risks due to interaction.
*OR adjusted for sex, study, age group, and list A jobs.

Table 7. Interactions between Occupational Silica Exposure and Smoking for Major Lung Cancer Subtypes

	Adenocarcinoma			Squamous Cell Carcinoma			Small Cell Carcinoma		
Exposure Status	**Cases (*n*)**	**OR***	**95% CI**	**Cases (*n*)**	**OR***	**95% CI**	**Cases (*n*)**	**OR***	**95% CI**
Never-smoker and never silica	589	1.0	Referent	195	1.0	Referent	82	1.0	Referent
Never-smoker and ever silica	111	1.01	0.81–1.24	62	1.22	0.90–1.62	29	1.49	0.96–2.27
Ever-smoker and never silica	3,094	3.90	3.52–4.32	4,057	11.0	9.47–12.8	1,779	13.6	10.9–17.3
Ever-smoker and ever silica	958	4.61	4.06–5.23	2,189	16.1	13.7–18.9	840	19.2	15.3–24.7
P value multiplicative interaction		0.17			0.23			0.80	
RERI		0.70	0.26–1.15		4.86	3.63–6.09		5.13	3.03–7.23

For definition of abbreviations, *see* Table 6.
*OR adjusted for sex, study, age group, and list A jobs.

DISCUSSION

In a large, international, pooled case–control study with more than 16,000 lung cancer cases, we found increases in lung cancer risks associated with continuous silica cumulative exposure as well as different categorical exposure metrics, including ever exposure, longer exposure duration, and higher cumulative exposure.

Positive associations between occupational silica exposure and lung cancer have been reported mainly in industrial cohorts. In a pooled analysis of 10 silica-exposed industrial cohorts, Steenland and colleagues reported a lung cancer risk increase of 1.07 for every unit increase in log-transformed cumulative silica exposure in milligram per cubic meter years with zero lag (Steenland et al., 2001). The corresponding risk increase reported by Liu and colleagues in a cohort of 34,018 workers in China was 1.06 (Liu et al., 2013). These estimates were very similar to the result from our analysis with log-cumulative exposure (OR, 1.05). Results from our corresponding spline analyses were consistent with the exposure–response relationships observed in the linear cumulative exposure logistic models; monotonic risk increases were observed for lung cancer and its subtypes.

Our results showed that silica is associated with lung cancer at very low cumulative exposures with no apparent threshold at concentrations investigated. ORs were 1.15 and 1.33 for our two lowest exposed groups, which had median cumulative exposures of 0.22 and 0.73 mg/m^3-years, respectively. Few other studies quantified lung cancer risks at levels near or below 1 mg/m^3-year. A meta-analysis with data from 19 studies calculated a pooled risk estimate of 1.19 (95% CI, 1.01–1.39) for workers with a median cumulative exposure of 0.42 mg/m^3-years (Poinen-Rughooputh et al., 2016). Liu and colleagues reported an OR point estimate of 1.12 (1.26 with 25 yr lag) for Chinese workers in the lowest exposed group with median exposure of 0.56 mg/m^3-years (Liu et al., 2013). However, results by Sogl and coworkers, who assessed silica exposure in German uranium mines using a measurement-based JEM, observed no lung cancer effects below cumulative silica exposures of 10 mg/m^3-years (Sogl et al., 2012).

For some carcinogens and related cancers, there is good evidence that disease relative risks after cessation of exposure are below unity when compared with groups with continued exposure (e.g., cigarette smoking) (IARC, 2007; Vlaanderen et al., 2014). We tested whether such a pattern was present in our population using time-since-exposure categories. We observed results suggesting that higher lung cancer risks were associated with more recent silica exposure. To our knowledge, this is the only study that included this metric for silica exposure and more evidence is needed to support this finding.

Whether silicosis is a prerequisite for silica-related lung cancer had been a topic of debate, primarily because results from earlier studies failed to support a consistent association between silica and lung cancer after excluding subjects with silicosis (Checkoway & Franzblau, 2000; Kurihara & Wada, 2004). A number of more recent studies set up analyses specifically to address this issue and reported evidence of a positive relationship between silica exposure and lung cancer without clinical silicosis (Checkoway et al., 1999; De Matteis et al., 2012; Liu et al., 2013; Poinen-Rughooputh et al., 2016; Sogl et al., 2012; Taeger et al., 2008). Results from our restricted analysis of subjects without silicosis similarly support a direct association between silica and lung cancer. Although underreporting of silicosis owing to self-reports by the index subject or proxy was possible in our study, the effects observed were unlikely to be caused solely by the misclassification of silicosis owing to the rarity of the condition in the general population.

Our findings suggest that lung squamous cell and small cell carcinomas are more strongly associated with silica exposure than lung adenocarcinoma. Research on lung cancer subtypes related specifically to silica exposure is rare. Two other large case–control studies in Europe and Canada similarly reported increased risks for all three major subtypes in relation to silica exposure, with the strongest association observed in squamous cell carcinoma (Cassidy et al., 2007; Vida et al., 2010). A large case–control study in Italy found elevated risk only for squamous cell and small cell carcinomas but not for adenocarcinoma (De Matteis et al., 2012). Most subjects in the three aforementioned studies, however, were also included in the current study and represented approximately 35% of our total participants.

Increases in overall lung cancer risk with increasing cumulative exposure were found regardless of smoking status. Our findings are in accordance with those from Liu and colleagues, in which never-smokers with cumulative silica exposure >1.12 mg/m^3-years had a lung cancer hazard ratio of 1.60 (95% CI, 1.01–2.55) (Liu et al., 2013). Ours is the first study to report an exposure–response between cumulative exposure to silica and lung cancer among never-smokers. Superadditive interactions of silica exposure and cigarette smoking were observed for overall lung cancer as well as all three major subtypes. Supermultiplicative interaction was also observed for all lung cancers combined. One other study reported a superadditive joint effect of silica exposure and smoking on lung cancer (Liu et al., 2013), and one reported no evidence for a joint effect beyond the multiplicative model (Cassidy et al., 2007).

Our study population comprised a large number of cases exposed to silica (n = 4,923) and allowed for stratification and interaction analyses for different cancer subtypes and risk factors. Despite having a large study population, our power to investigate silica exposure-related cancer risks in women were limited. This is because the number of exposed cases in women (n = 274) was much smaller than those in men (n = 4,649). Analyses restricted to females showed imprecise results with OR point estimates that were generally >1 (*see* Supplementary Table 3a). Male-specific results are also available in Supplementary Table 3b.

We performed quantitative exposure assessment specific for exposure to quartz silica, which allowed for quantification of exposure–disease risks and exploration of the shape of the exposure–response cuxrves in a population-based case–control setting. However, our estimates of silica exposure may be affected by exposure misclassification and less accurate than some industrial cohort–based studies, particularly those with detailed work history and extensive historical silica measurements. This misclassification was likely to be nondifferential with respect to case status and would result in a bias of risk estimates toward the null. Owing to sparse measurement data for years before 1960 in our JEM, we assumed in backward extrapolation that silica exposure did not further increase in years before 1960. In a previous publication we have explored different time-trend assumptions in our exposure model (Peters et al., 2013). Naturally, the assigned silica exposures in the population (and the slope of the exposure–response) would vary if different time-trend assumptions were made, but these changes have little effect on the exposure status and ranking of cumulative exposure among our study population. When we restricted our categorical exposure model to include only subjects who started work after 1960 (*see* Supplementary Table 4.2), the silica lung cancer exposure–response in general and, more specifically, elevated lung cancer risks associated with lower categories of cumulative silica exposure were still observed.

Our study included more complete information on individual covariates than most industry-based studies, which allowed for the control of important potential confounders such as smoking and exposures to other occupational lung carcinogens in our models. As an alternative to adjusting for coexposures to other lung carcinogens with ever employment in list A jobs, we performed a sensitivity analysis controlling for Domtoren-JEM–assessed ever exposure to diesel engine exhaust, hexavalent chromium, asbestos, and polycyclic aromatic hydrocarbons in our categorical exposure model. Results of this analysis (*see* Supplementary Table 4.4) were very similar compared with our main results. The associations we observed between silica and lung cancer were also robust in other sensitivity analyses with different subgroups (*see* Supplement Methods and Results, and Tables 4.1–4.5).

Current definitions of "tolerable" ELR owing to occupational exposure to carcinogens vary by jurisdiction, ranging from the 0.4% in the Netherlands and Germany to 0.1% generally accepted by the U.S. Occupational Safety and Health Administration (AGS, 2019; Health Council of the Netherlands, 2019; Rodricks et al., 1987). According to our calculations, lifetime occupational silica exposure at 0.05 and 0.1 mg/m^3 would result in respective lung cancer ELRs of 0.45% and 0.96%, which clearly exceed this range of tolerable risks. Lifetime silica exposure at 0.025 mg/m^3 would result in approximately two lung cancers in 1,000 workers, which falls below the Dutch/German limit of 0.4% but above the U.S. limit of 0.1%. Other studies have estimated similar lung cancer ELRs at low levels of silica exposure, with one study estimating an ELR of 0.23% to 0.48% for workers exposed to 0.07 mg/m^3 of silica and another estimating an ELR of 0.2% to 0.3% for workers with an exposure level of 0.01 mg/m^3 (IARC, 2012; Liu et al., 2013; Steenland et al., 2001). The ELR findings from other studies and ours suggest that lower occupational silica exposure limits may be considered to protect exposed workers from excess lung cancer risks. Lastly, because our exposure assessment was specific for quartz silica and did not include other forms of silica, our ELR may not reflect risks from exposures to other forms of crystalline silica such as cristobalite and tridymite. However, because quartz is by far the most common form of crystalline silica, exposure prevalence and disease burden associated with other crystalline silica polymorphs are likely to be much smaller than those associated with quartz exposure (IARC, 1997).

CONCLUSIONS

In a large pooled analysis of lung cancer case–control studies, we observed a positive association and exposure–response relationship between occupational silica exposure and lung cancer. The exposure–disease association was consistent regardless of tobacco smoking history or silicosis status. Silica-exposed workers had higher risks for all investigated lung cancer subtypes; risks were higher for squamous cell and small cell carcinomas than for adenocarcinoma. Our findings support efforts to further reduce occupational exposure to silica for the protection of exposed workers against risks of developing lung cancer.

REFERENCES

ACGIH. (2010). *Silica, Crystalline: α-Quartz and Cristobalite: TLV® Chemical Substances 7th Edition Documentation*. ACGIH.

AGS. (2019, March 29). *TRGS 910 Risikobezogenes Maßnahmenkonzept für Tätigkeiten mit krebserzeugenden Gefahrstoffen (Technical Rules for Hazardous Substances 910: Risk-based action plan for activities with carcinogenic hazardous substances—In German)*. https://www.baua.de/DE/Angebote/Rechtstexte-und-Technische-Regeln/Regelwerk/TRGS/pdf/TRGS-910.pdf?__blob=publicationFile&v=4

Ahrens, W., & Merletti, F. (1998). A Standard Tool for the Analysis of Occupational Lung Cancer in Epidemiologic Studies. *International Journal of Occupational and Environmental Health*, 4(4), 236–240. https://doi.org/10.1179/oeh.1998.4.4.236

Cassidy, A., Mannetje, A., van Tongeren, M., Field, J. K., Zaridze, D., Szeszenia-Dabrowska, N., Rudnai, P., Lissowska, J., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Fevotte, J., Fletcher, T., Brennan, P., & Boffetta, P. (2007). Occupational Exposure to Crystalline Silica and Risk of Lung Cancer: A Multicenter Case-Control Study in Europe. *Epidemiology*, *18*(1), 36–43.

Checkoway, H., & Franzblau, A. (2000). Is silicosis required for silica-associated lung cancer? *American Journal of Industrial Medicine*, *37*(3), 252–259.

Checkoway, Harvey, Hughes, J. M., Weill, H., Seixas, N. S., & Demers, P. A. (1999). Crystalline silica exposure, radiological silicosis, and lung cancer mortality in diatomaceous earth industry workers. *Thorax*, *54*(1), 56–59. https://doi.org/10.1136/thx.54.1.56

De Matteis, S., Consonni, D., Lubin, J. H., Tucker, M., Peters, S., Vermeulen, R. C., Kromhout, H., Bertazzi, P. A., Caporaso, N. E., Pesatori, A. C., Wacholder, S., & Landi, M. T. (2012). Impact of occupational carcinogens on lung cancer risk in a general population. *International Journal of Epidemiology*, *41*(3), 711–721. https://doi.org/10.1093/ije/dys042

EU Parliament and Council. (2019). DIRECTIVE (EU) 2019/130 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 January 2019 amending Directive 2004/37/EC on the protection of workers from the risks related to exposure to carcinogens or mutagens at work. *Official Journal of the European Union*, *L30*(112). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0130&from=EN

European Commission. (2008). *Eurostat 2008 dataset on all causes and lung cancer mortality in European Union countries*. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_cd_aro&lang=en.

Health Council of the Netherlands. (2019, March 13). *Diesel Engine Exhaust: Health-based recommended occupational exposure limit*. https://www.gezondheidsraad.nl/binaries/gezondheidsraad/documenten/adviezen/2019/03/13/dieselmotoremissie/Diesel+Engine+Exhaust.pdf

IARC. (1997). *Silica, Some Silicates, Coal Dust and para-Aramid Fibrils* (Volume 68; IARC Monographs on the Evaluation of Carcinogenic Risks to Humans). https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Silica-Some-Silicates-Coal-Dust-And-Em-Para-Em--Aramid-Fibrils-1997

IARC. (2007). *IARC Handbooks of Cancer Prevention: Tobacco Control* (Volume 11). https://publications.iarc.fr/Book-And-Report-Series/Iarc-Handbooks-Of-Cancer-Prevention/Tobacco-Control-Reversal-Of-Risk-After-Quitting-Smoking-2007

IARC. (2012). *Silica dust, Crystalline, in the form for Quartz or Cristobalite* (No. 100C; IARC Monographs on the Evaluation of Carcinogenic Risks to Humans). IARC. https://monographs.iarc.fr/wp-content/uploads/2018/06/mono100C-14.pdf

ILO. (2010, June 10). *ISCO-International Standard Classification of Occupations: Brief History*. http://www.ilo.org/public/english/bureau/stat/isco/intro2.htm

Kurihara, N., & Wada, O. (2004). Silicosis and smoking strongly increase lung cancer risk in silica-exposed workers. *Industrial Health*, *42*(3), 303–314.

Liu, Y., Steenland, K., Rong, Y., Hnizdo, E., Huang, X., Zhang, H., Shi, T., Sun, Y., Wu, T., & Chen, W. (2013). Exposure-Response Analysis and Risk Assessment for Lung Cancer in Relationship to Silica Exposure: A 44-Year Cohort Study of 34,018 Workers. *American Journal of Epidemiology*, *178*(9), 1424–1433. https://doi.org/10.1093/aje/kwt139

Manno, M., Levy, L., Johanson, G., & Cocco, P. (2018). Silica, silicosis and lung cancer: What level of exposure is acceptable? *La Medicina Del Lavoro*, *109*(6), 478–480. https://doi.org/10.23749/mdl.v109i6.7928

Mirabelli, D., Chiusolo, M., Calisti, R., Massacesi, S., Richiardi, L., Nesti, M., & Merletti, F. (2001). [Database of occupations and industrial activities that involve the risk of pulmonary tumors]. *Epidemiologia E Prevenzione*, *25*(4–5), 215–221.

NIOSH. (2002). *Health Effects of Occupational Exposure to Respirable Crystalline Silica* (No. 2002–129; NIOSH Hazard Review).

NTP. (2016). *Silica, Crystalline (Respirable Size)* (Report on Carcinogens, Fourteenth Edition).

Olsson, A. C., Gustavsson, P., Kromhout, H., Peters, S., Vermeulen, R., Brüske, I., Pesch, B., Siemiatycki, J., Pintos, J., Brüning, T., Cassidy, A., Wichmann, H.-E., Consonni, D., Landi, M. T., Caporaso, N., Plato, N., Merletti, F., Mirabelli, D., Richiardi, L., … Straif, K. (2011). Exposure to diesel motor exhaust and lung cancer risk in a pooled analysis from case-control studies in Europe and Canada. *American Journal of Respiratory and Critical Care Medicine*, *183*(7), 941–948. https://doi.org/10.1164/rccm.201006-0940OC

Olsson, A. C., Vermeulen, R., Schüz, J., Kromhout, H., Pesch, B., Peters, S., Behrens, T., Portengen, L., Mirabelli, D., Gustavsson, P., Kendzia, B., Almansa, J., Luzon, V., Vlaanderen, J., Stücker, I., Guida, F., Consonni, D., Caporaso, N., Landi, M. T., … Straif, K. (2017). Exposure–Response Analyses of Asbestos and Lung Cancer Subtypes in a Pooled Analysis of Case–Control Studies. *Epidemiology (Cambridge, Mass.)*, *28*(2), 288–299. https://doi.org/10.1097/EDE.0000000000000604

OSHA. (2017). *Small Entity Compliance Guide for the Respirable Crystalline Silica Standard for General Industry and Maritime*. https://www.osha.gov/Publications/OSHA3911.pdf

Peters, S., Kromhout, H., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Vermeulen, R. (2013). Sensitivity Analyses of Exposure Estimates from a Quantitative Job-exposure Matrix (SYN-JEM) for Use in Community-based Studies. *Annals of Occupational Hygiene*, *57*(1), 98–106. https://doi.org/10.1093/annhyg/mes045

Peters, S., Vermeulen, R., Cassidy, A., Mannetje, A. 't, Tongeren, M. van, Boffetta, P., Straif, K., & Kromhout, H. (2011). Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case–control study. *Occupational and Environmental Medicine*, *68*(2), 148–153. https://doi.org/10.1136/oem.2010.055608

Peters, S., Vermeulen, R., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Kromhout, H. (2011). *Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies*. *13*(11), 3262–3268. https://doi.org/10.1039/C1EM10628G

Peters, S., Vermeulen, R., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Kromhout, H. (2016). SYN-JEM: A Quantitative Job-Exposure Matrix for Five Lung Carcinogens. *Annals of Occupational Hygiene*, *60*(7), 795–811. https://doi.org/10.1093/annhyg/mew034

Poinen-Rughooputh, S., Rughooputh, M. S., Guo, Y., Rong, Y., & Chen, W. (2016). Occupational exposure to silica dust and risk of lung cancer: An updated meta-analysis of epidemiological studies. *BMC Public Health*, *16*(1), 1137. https://doi.org/10.1186/s12889-016-3791-5

R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Rodricks, J. V., Brett, S. M., & Wrenn, G. C. (1987). Significant risk decisions in federal regulatory agencies. *Regulatory Toxicology and Pharmacology*, *7*(3), 307–320. https://doi.org/10.1016/0273-2300(87)90038-9

Rothman, K., & Greenland, S. (1998). *Modern Epidemiology*. Lippincott - Raven.

Rushton, L. (2011). Chronic Obstructive Pulmonary Disease and Occupational Exposure to Silica. *Reviews on Environmental Health*, *22*(4), 255–272. https://doi.org/10.1515/REVEH.2007.22.4.255

Sogl, M., Taeger, D., Pallapies, D., Brüning, T., Dufey, F., Schnelzer, M., Straif, K., Walsh, L., & Kreuzer, M. (2012). Quantitative relationship between silica exposure and lung cancer mortality in German uranium miners, 1946–2003. *British Journal of Cancer*, *107*(7), 1188–1194. https://doi.org/10.1038/bjc.2012.374

Steenland, K., Mannetje, A., Boffetta, P., Stayner, L., Attfield, M., Chen, J., Dosemeci, M., DeKlerk, N., Hnizdo, E., Koskela, R., & Checkoway, H. (2001). Pooled exposure–response analyses and risk assessment for lung cancer in 10 cohorts of silica-exposed workers: An IARC multicentre study. *Cancer Causes & Control*, *12*(9), 773–784. https://doi.org/10.1023/A:1012214102061

Steenland, K., & Ward, E. (2014). Silica: A lung carcinogen. *CA: A Cancer Journal for Clinicians*, *64*(1), 63–69. https://doi.org/10.3322/caac.21214

Stevenson, M., Sergeant, E., Nunes, T., Heuer, C., Marshall, J., Sanchez, J., Thornton, R., Reiczigel, J., Robison-Cox, J., Sebastiani, P., Solymos, P., Yoshida, K., Jones, G., Pirikahu, S., Firestone, S., Kyle, R., Popp, J., & Reynard, M. J. and C. (2020). *epiR: Tools for the Analysis of Epidemiological Data* (2.0.17) [Computer software]. https://CRAN.R-project.org/package=epiR

Taeger, D., Krahn, U., Wiethege, T., Ickstadt, K., Johnen, G., Eisenmenger, A., Wesch, H., Pesch, B., & Bruning, T. (2008). A study on lung cancer mortality related to radon, quartz, and arsenic exposures in German uranium miners. *Journal of Toxicology and Environmental Health. Part A*, *71*(13–14), 859–865. https://doi.org/10.1080/15287390801987972

Vermeulen, R., Silverman, D. T., Garshick, E., Vlaanderen, J., Portengen, L., & Steenland, K. (2014). Exposure-response estimates for diesel engine exhaust and lung cancer mortality based on data from three occupational cohorts. *Environmental Health Perspectives*, *122*(2), 172–177. https://doi.org/10.1289/ehp.1306880

Vida, S., Pintos, J., Parent, M.-E., Lavoué, J., & Siemiatycki, J. (2010). Occupational exposure to silica and lung cancer: Pooled analysis of two case-control studies in Montreal, Canada. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *19*(6), 1602–1611. https://doi.org/10.1158/1055-9965.EPI-10-0015

Vlaanderen, J., Portengen, L., Schüz, J., Olsson, A., Pesch, B., Kendzia, B., Stücker, I., Guida, F., Brüske, I., Wichmann, H.-E., Consonni, D., Landi, M. T., Caporaso, N., Siemiatycki, J., Merletti, F., Mirabelli, D., Richiardi, L., Gustavsson, P., Plato, N., … Vermeulen, R. (2014). Effect Modification of the Association of Cumulative Exposure and Cancer Risk by Intensity of Exposure and Time Since Exposure Cessation: A Flexible Method Applied to Cigarette Smoking and Lung Cancer in the SYNERGY Study. *American Journal of Epidemiology*, *179*(3), 290–298. https://doi.org/10.1093/aje/kwt273

WHO. (2007). *The Global Occupational Health Network newsletter: Elimination of silicosis*. www.who.int/occupational_health/publications/newsletter/gohnet12e.pdf

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B*, *73*(1), 3–36.

Supplementary Methods

Sensitivity analysis

To assess if risks associated with categorical silica cumulative exposure differed between studies that used population versus hospital controls, stratified analyses by type of control were performed. Restricted risk analyses were also performed on workers who held blue-collar jobs and who started work after 1960 to assess if risks differed for workers with lower socio-economic status and for workers who worked in more recent periods when exposure estimates are more reliable, respectively. Because workers in mining and agriculture may have different silica exposure patterns compared to other occupations, additional restricted analyses were performed on the study population without subjects who were ever employed in mining or agriculture. As an alternative control for potential confounding exposures to other occupational lung carcinogens, we constructed our main categorical cumulative exposure model without adjustment for List A jobs and with additional adjustment for ever exposure to asbestos, diesel exhaust, hexavalent chromium, and polycyclic aromatic hydrocarbons as assessed by the DOM-JEM. Additionally, study-specific ORs with adjustments identical to the main models were calculated to assess heterogeneity between studies. Extent of heterogeneity between OR estimates from different studies was assessed using the p-value of the Cochran's Q statistic and as a percentage in I^2 (Higgins et al., 2003).

Supplementary Results

Sensitivity analysis

Associations between cumulative silica exposure and lung cancer were found regardless of the type of controls used (Supplementary Table 4.1), though ORs from studies with hospital controls (4,310 cases; 4,868 controls) were lower and more imprecise compared to those with population controls (12,263 cases; 16,287 controls). Similarly, exposure-response associations were observed when we restricted analyses to blue-collar workers, to workers who started work after 1960, to subjects never employed in mining, and to subjects never employed in agriculture (Supplementary Tables 4.2-4.3). Compared to our main categorical model, slight attenuation of risk estimates was generally observed when we restricted our analyses to blue-collar workers (12,444 cases; 13,111 controls) and to subjects never employed in mining (16,186 cases; 20,508 controls). In contrast, risk estimates were generally higher when we limited our analyses to subjects who started work after 1960 (4,471 cases; 6,478 controls) and who never worked in agriculture (14,462 cases; 18,218 controls). When we adjusted for potential confounding occupational exposures with individual exposures rather than List A jobs, we continue to observe increasing lung cancer risk trend with increasing cumulative exposure, albeit with slightly attenuated ORs across cumulative exposure categories compared to risks observed in our main model (Supplementary Table 4.4).

A moderate amount of heterogeneity (I^2 = 40.8%; Q=31.9, p=0.03) was observed between included studies. Heterogeneity reduced significantly after removal of one study (AUT) from our study population (I^2=5.9%; Q=15.3, p=0.64). The silica-lung cancer exposure-response pattern in the more homogenous subgroup of studies was similar to results of the main analyses, albeit with attenuated ORs for the two highest exposed groups (13,721 cases; 17,716 controls; Supplementary Table 4.5).

Supplementary References

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal*, *327*(7414), 557–560.

Peters, S., Vermeulen, R., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Kromhout, H. (2011). *Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies*. *13*(11), 3262–3268. https://doi.org/10.1039/C1EM10628G

Supplementary Table 1. Description of the studies included in these analyses in the SYNERGY project

			Cases		Controls					
Study	Country	Data collection	N	Response rate (%)	N	Response rate (%)	Quartz exposure	Silicosis Self-report	Control Source **	Interview††
AUT-Munich	Germany	1990–1995	3180	77	3249	41	1931-1995	Yes†	P	S
CAPUA	Spain	2000–2010	559	91	512	96	1926-2010	No	H	S
EAGLE	Italy	2002–2005	1908	87	2065	72	1932-2005	Yes†	P	S
HdA	Germany	1988–1993	1004	69	1002	68	1926-1993	Yes§	P	S
ICARE	France	2001–2007	2739	63	3449	77	1937-2007	No	P	S & NOK
INCO	Czech Republic	1999–2002	304	94	452	80	1937-2002	Yesll	H	S
INCO	Hungary	1998–2001	391	90	305	100	1931-1999	Yesll	H	S
INCO	Poland	1998–2002	793	88	835	88	1933-2001	Yesll	P & H	S
INCO	Romania	1998–2002	179	90	225	99	1943-2001	Yesll	H	S
INCO	Russia	1998–2001	599	96	580	90	1938-2000	Yesll	H	S
INCO	Slovakia	1998–2002	345	90	285	84	1937-2002	Yesll	H	S
INCO/LLP	United Kingdom	1998–2005	441	78	916	84	1934-2004	Yesll	P	S
LUCA	France	1989–1992	280	98	282	98	1927-1992	No	H	S
LUCAS	Sweden	1985–1990	1014	87	2307	85	1923-1990	No	P	S & NOK
MONTREAL	Canada	1996–2002	1176	85	1505	69	1936-1999	No	P	S & NOK
MORGEN*	Netherlands	1993–1997	43	N/A	115	N/A	1945-1994	No	P	S
PARIS	France	1988–1992	169	95	227	95	1929-1992	No	H	S
ROME	Italy	1993–1996	326	74	321	63	1926-1995	No	H	S
TORONTO	Canada	1997–2002	365	62	844	71	1929-2002	No	P & H	S
TURIN/VENETO	Italy	1990–1994	1086	79	1489	80	1925-1994	No	P	S
Overall	14 countries	1985–2010	16 901	78%	20 965	69%	1923-2010		H=21%	NOK=7.3%

*Nested case-control study: 45% of invited participants to the original cohort completed the baseline questionnaire.
†: Interview question: Up until two years ago has a doctor ever told you that you have or have had silicosis?
‡: Interview question: More than a year ago did a doctor ever tell you that you had silicosis?
§: Interview question: Up until two years ago has a doctor ever told you that you have or have had silicosis?
ll: Interview question: Have you ever had silicosis (in list with 8 other medical conditions); how old?
**: P = population controls; H = hospital controls
††: S = subject; NOK = Next-of-kin

Supplementary Table 2. The ten jobs with highest modelled silica exposure in SYN-JEM and ten most prevalent exposed jobs among the controls in the SYNERGY population*

ISCO-68†	Job title	GM (mg/m³)‡
Highest exposed jobs in SYN-JEM		
9–51.25	Bricklayer (chimney)	0.11
8–20.90	Stone cutters and carvers	0.10
8–20.80	Monument carver (hand)	0.10
7–11.70	Sampler (mine)	0.09
9–59.45	Demolition worker	0.09
7–12.20	Stone splitter	0.08
8–20.70	Stone carver (hand)	0.06
8–99.40	Clay slip maker	0.05
7–11.05	Miner (general)	0.05
8–99.30	Clay mixer	0.05
Most prevalent exposed jobs among SYNERGY controls		
6–21.10	Farm helper (general)	0.02
6–11.10	General farmer	0.02
9–51.20	Bricklayer (construction)	0.03
6–21.05	Farm worker (general)	0.01
9–59.10	House builder (general)	0.04
7–11.05	Miner (general)	0.05
6–28.20	Motorised farm equipment operator	0.02
9–59.90	Other construction workers	0.02
6–27.40	Gardener	0.02
9–52.10	Reinforced concreter (general)	0.02

*Table adapted from Table 3 in SYN-JEM silica exposure modelling manuscript by Peters and colleagues (Peters et al., 2011).
†ISCO-68: International Standard Classification of Occupations, version 1986
‡GM: geometric mean of quartz silica exposure, modelled by SYN-JEM

Supplementary Table 3a. Lung cancer odds ratios (OR) associated with categorical indices of occupational silica exposure in women

Occupational silica exposure	Exposure category	Cases	%	Controls	%	OR*	95% CI
Reference group	Non-exposed	3022	91.7	4166	92.3	1.0	Referent
Ever exposure	Ever	274	8.3	348	7.7	1.11	0.91–1.34
Duration (years)	1–9	170	5.2	183	4.1	1.08	0.84–1.39
	10–19	49	1.5	73	1.6	1.13	0.75–1.69
	20–29	21	0.6	25	0.6	1.40	0.74–2.63
	>29	34	1.0	67	1.5	1.05	0.67–1.63
	Test for trend, p-value					0.27	
	Excl. never RCS exposed					0.25	
Cumulative exposure	>0–0. 39	108	3.3	102	2.3	1.07	0.77–1.48
(mg/m^3-years)	0.4–1.09	80	2.4	92	2.0	1.24	0.88–1.74
	1.1–2. 39	51	1.5	93	2.1	1.02	0.69–1.47
	≥2.4	35	1.1	61	1.4	1.10	0.69–1.74
	Test for trend, p-value					0.97	
	Excl. never RCS exposed					0.50	
Time since last exposure (years)†	<5	28	0.8	33	3.9	1.62	0.74–3.59
	5–9	19	0.6	15	1.7	2.00	0.79–5.19
	10–19	22	0.7	42	2.7	0.95	0.46–1.92
	20–29	28	0.8	42	2.6	1.12	0.61–2.05
	30–39	55	1.7	55	3.9	1.46	0.91–2.32
	>39	122	3.7	161	6.7	0.97	0.72–1.30
	Test for trend, p-value						
	*Excl. never RCS exposed**					0.77	

*OR adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

† OR in "time since last exposure" is additionally adjusted for duration (continuous) of silica exposure. Trend test limited to exposed subjects.

Supplementary Table 3b. Lung cancer odds ratios (OR) associated with categorical indices of occupational silica exposure in men

Occupational silica exposure	Exposure category	Cases	%	Controls	%	OR*	95% CI
Reference group	Non-exposed	8956	65.8	12311	74.8	1.0	Referent
Ever exposure	Ever	4649	34.2	4140	25.2	1.31	1.24–1.39
Duration (years)	1–9	1865	13.7	1753	10.7	1.22	1.13–1.33
	10–19	877	6.4	832	5.1	1.20	1.07–1.34
	20–29	614	4.5	494	3.0	1.44	1.25–1.65
	>29	1293	9.5	1061	6.4	1.51	1.37–1.67
	Test for trend, p-value					<0.01	
	Excl. never RCS exposed					<0.01	
Cumulative exposure	>0–0.39	1005	7.4	1026	6.2	1.15	1.04–1.28
(mg/m^3-years)	0.4–1.09	1141	8.4	1028	6.2	1.34	1.21–1.48
	1.1–2.39	1180	8.7	1029	6.3	1.31	1.18–1.45
	≥2.4	1323	9.7	1057	6.4	1.45	1.31–1.61
	Test for trend, p-value					<0.01	
	Excl. never RCS exposed					<0.01	
Time since last exposure (years)†	<5	906	6.7	782	4.8	1.38	1.13–1.69
	5–9	443	3.3	336	2.0	1.37	1.09–1.72
	10–19	657	4.8	527	3.2	1.34	1.11–1.62
	20–29	589	4.3	494	3.0	1.25	1.06–1.47
	30–39	876	6.4	757	4.6	1.28	1.13–1.46
	>39	1178	8.7	1244	7.6	1.10	0.99–1.22
	Test for trend, p-value						
	*Excl. never RCS exposed**					0.12	

*OR adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

† OR in "time since last exposure" is additionally adjusted for duration (continuous) of silica exposure. Trend test limited to exposed subjects.

Supplementary Table 4. Sensitivity analyses for the association between silica cumulative exposure categories and lung cancer

Supplementary Table 4.1 Analyses stratified by type of controls

Cumulative exposure mg/m³-years	Studies with population controls* (12663 cases/16287 controls)			Studies with hospital controls* (4310 cases/4868 controls)		
	Cases/controls	OR†	95%CI	Cases/controls	OR†	95%CI
Unexposed	9136/13099	1.0	Referent	3127/3865	1.0	Referent
>0–0.39	969/964	1.19	1.07–1.33	116/113	1.10	0.82–1.47
0.4–1.09	913/828	1.37	1.23–1.54	222/195	1.24	0.99–1.54
1.1–2.39	899/784	1.42	1.26–1.59	287/258	1.07	0.89–1.31
≥2.4	746/612	1.54	1.36–1.76	558/437	1.29	1.11–1.51
Test for trend, p-value§		<0.01			<0.01	
Excl. never exposed		<0.01			0.29	

*Subjects from the INCO Poland and Toronto studies were included in both analyses, since both types of controls were used

† OR is adjusted for sex, study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

Supplementary Table 4.2 Analyses restricted to blue-collar workers and workers who started work after 1960

Cumulative exposure mg/m³-years	Restricting the study base to blue-collar workers (12444 cases/13111 controls)			Restricted to workers started after 1960 (4471 cases/6478 controls)		
	Cases/controls	OR*	95%CI	Cases/controls	OR*	95%CI
Unexposed	7953/9264	1.0	Referent	3642/5667	1.0	Referent
>0–0.39	1030/992	1.05	0.94 – 1.17	327/352	1.19	0.99–1.44
0.4–1.09	1133/967	1.27	1.15–1.41	235/234	1.34	1.07–1.66
1.1–2.39	1087/895	1.20	1.08–1.34	166/166	1.63	1.24–2.14
≥2.4	1241/993	1.34	1.21–1.49	95/101	1.26	0.90–1.77
Test for trend, p-value§		<0.01			0.01	
Excl. never exposed		<0.01			0.54	

*OR is adjusted for sex, study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

Supplementary Table 4.3 Analyses excluding subjects ever-employed in agriculture and mining industries

Cumulative exposure mg/m³-years	Subjects never-employed in agriculture (14462 cases/18218 controls)			Subjects never-employed in mining (16186 cases/20508 controls)		
	Cases/controls	OR*	95%CI	Cases/controls	OR*	95%CI
Unexposed	11769/16163	1.0	Referent	11973/16467	1.0	Referent
>0–0.39	610/552	1.21	1.06–1.39	1065/1094	1.13	1.02–1.24
0.4–1.09	667/470	1.58	1.38–1.81	1070/1018	1.30	1.18–1.44
1.1–2.39	613/430	1.48	1.28–1.71	1055/1010	1.26	1.14–1.40
≥2.4	803/603	1.52	1.34–1.72	1023/910	1.40	1.26–1.56
Test for trend, p-value§		<0.01			<0.01	
Excl. never exposed		0.02			<0.01	

*OR is adjusted for sex, study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

Supplementary Table 4.4 Analyses with alternative adjustment for potential confounding occupational exposures

Cumulative exposure mg/m³-years	All subjects		
	Cases/controls	OR*	95%CI
Unexposed	11978/16467	1.0	Referent
>0–0.39	1113/1128	1.12	1.01–1.24
0.4–1.09	1221/1120	1.30	1.18–1.44
1.1–2.39	1231/1122	1.25	1.13–1.38
≥2.4	1358/1118	1.39	1.26–1.54
Test for trend, p-value§		<0.01	
Excl. never exposed		<0.01	

*OR is adjusted for sex, study, age group, smoking (pack-years, time-since-quitting smoking), and ever exposure to asbestos, diesel exhaust, hexavalent chromium, and polycyclic aromatic hydrocarbons

Supplementary Table 4.5 Analyses by more homogenous group of studies (excluding AUT)

Cumulative exposure mg/m³-years	All studies except AUT (13721 cases/17716 controls)		
	Cases/controls	OR*	95%CI
Unexposed	10224/14279	1.0	Referent
>0–0.39	709/734	1.13	1.00–1.28
0.4–1.09	859/864	1.26	1.13–1.42
1.1–2.39	822/869	1.12	1.00–1.26
≥2.4	1107/970	1.37	1.23–1.52
Test for trend, p-value§		<0.01	
Excl. never exposed		<0.01	

*OR is adjusted for sex, study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

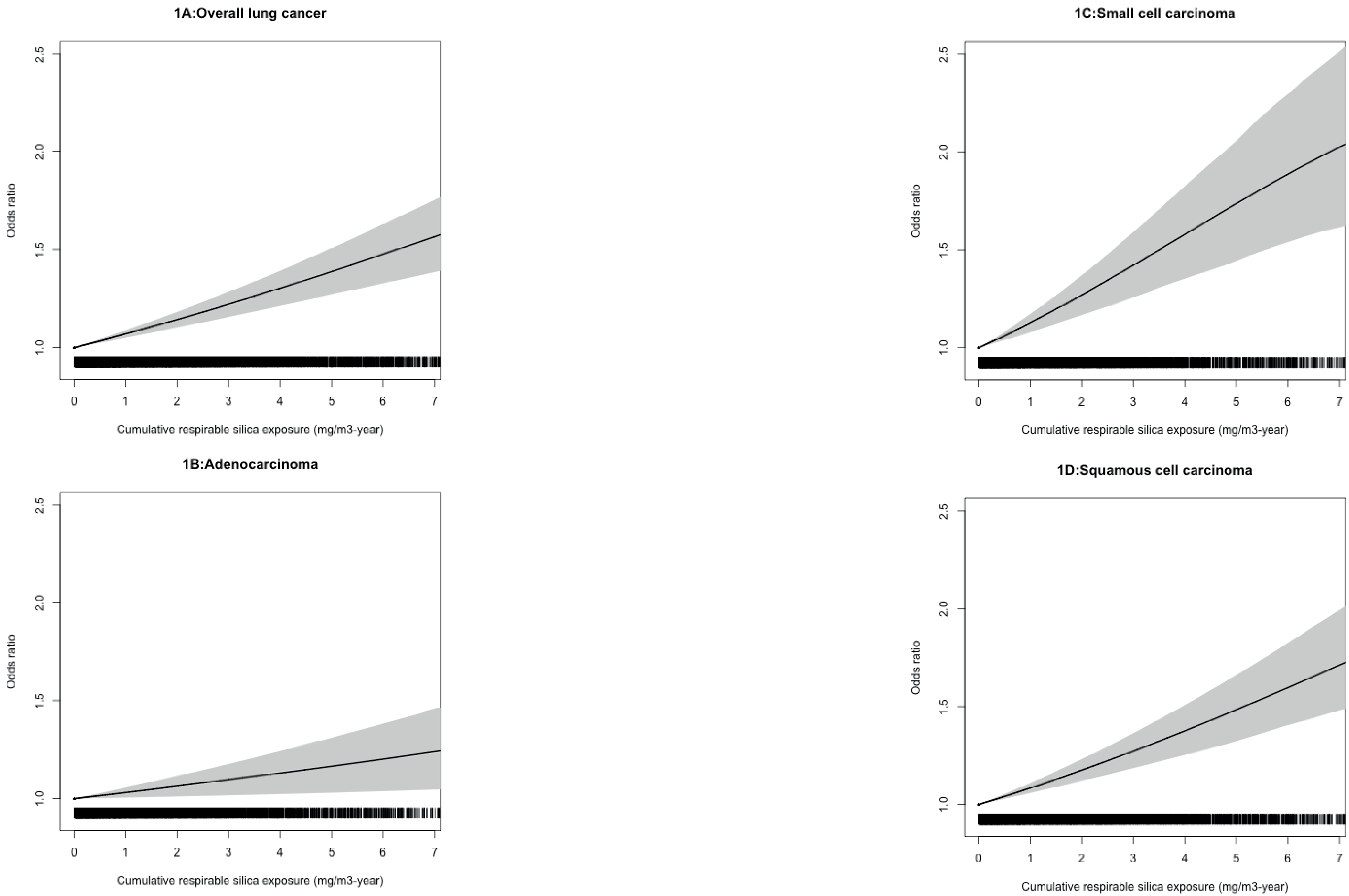


Figure 1: Splines and associated 95% confidence intervals for cumulative silica exposure on risks of 1A) overall lung cancer; 1B) adenocarcinoma; 1C) small cell carcinoma; 1D) squamous cell carcinoma. Bottom bars show frequency of subjects at associated exposure levels.
Abbreviation: mg/m^3-years = milligram per cubic metre years

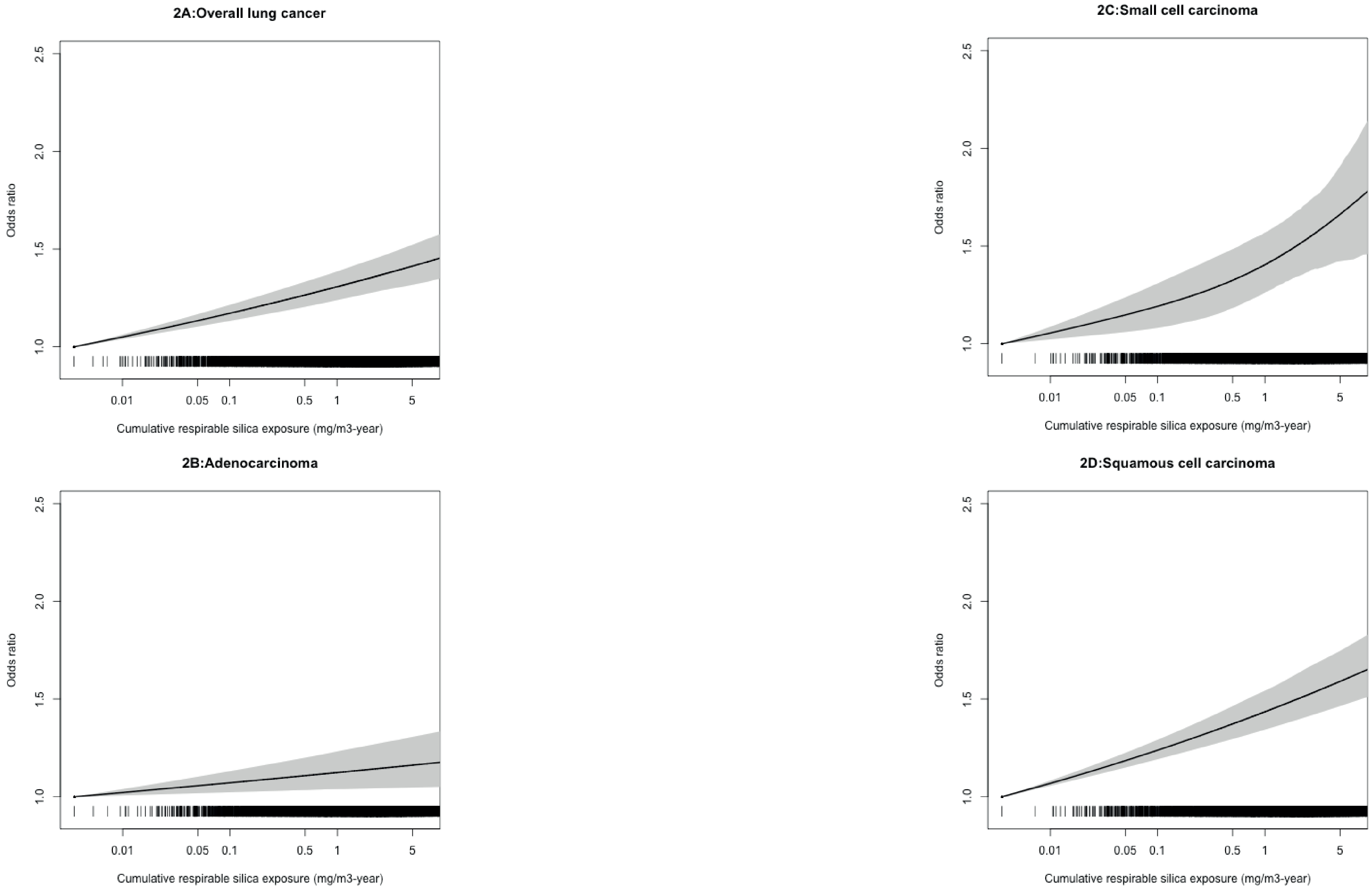


Figure 2: Splines and associated 95% confidence intervals for log-transformed cumulative silica exposure on risks of 1A) overall lung cancer; 1B) adenocarcinoma; 1C) small cell carcinoma; 1D) squamous cell carcinoma. Bottom bars show frequency of subjects at associated exposure levels. Abbreviation: mg/m³-years = milligram per cubic metre years

CHAPTER 7

Diesel Engine Exhaust Exposure, Smoking, and Lung Cancer Subtype Risks: A Pooled Exposure–Response Analysis of 14 Case–Control Studies

C Ge[1], S Peters[1], A Olsson[2], L Portengen[1], J Schüz[2], J Almansa[1], W Ahrens[3], V Bencko[4], S Benhamou[5], P Boffetta[6,7], B Bueno-de-Mesquita[8], N Caporaso[9], D Consonni[10], P Demers[11], E Fabiánová [12,13], G Fernández-Tardón[14], J Field[15], F Forastiere[16], L Foretova[17], P Guénel[18], P Gustavsson[19], V Janout[20], KH Jöckel[21], S Karrasch[22,23,24], MT Landi[9], J Lissowska[25], D Luce[26], D Mates[27], J McLaughlin[28], F Merletti[29], D Mirabelli[29], T Pándics[30], MÉ Parent[31], N Plato[19], H Pohlabeln[3], L Richiardi[29], J Siemiatycki[32], B Świątkowska[33], A Tardón[14], HE Wichmann[34,35], D Zaridze[36], K Straif[2], H Kromhout[1], and R Vermeulen[1]

Published: *American Journal of Respiratory and Critical Care Medicine*, 2020, Vol. 202, Iss. 3, 402–411.

doi: 10.1164/rccm.201911-2101OC

Affiliations

1. Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands;
2. International Agency for Research on Cancer (IARC/WHO), Lyon, France;
3. Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany;
4. Institute of Hygiene and Epidemiology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic;
5. INSERM U 1018, Villejuif, France;
6. Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York;
7. Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy;
8. The National Institute for Public Health and Environmental Protection, Bilthoven, the Netherlands;
9. National Cancer Institute, Bethesda, Maryland;
10. Epidemiology Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy;
11. Occupational Cancer Research Centre, Cancer Care Ontario, Toronto, Ontario, Canada;
12. Regional Authority of Public Health, Banská Bystrica, Slovakia;
13. Faculty of Health, Catholic University, Ružomberok, Slovakia;
14. Fundación para la Investigación e Innovación Biomédica en el Principado de Asturias – Instituto de Investigación Sanitaria del Principado (FINBA-ISPA), Faculty of Medicine, University of Oviedo and Centro de Investigación Biomédica en Red Epidemiología y Salud Pública (CIBERESP), Oviedo, Spain;
15. Roy Castle Lung Cancer Research Programme, Cancer Research Centre, University of Liverpool, Liverpool, United Kingdom;
16. Consiglio Nazionale delle Ricerche-Istituto per la Ricerca e l'Innovazione Biomedica (CNR-Irib), Palermo, Italy;
17. Masaryk Memorial Cancer Institute, Brno, Czech Republic;
18. Center for research in Epidemiology and Population Health (CESP), Cancer and Environment team, Inserm U1018, University Paris-Sud, University Paris-Saclay, Villejuif, France;
19. The Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden;
20. Faculty of Health Sciences, Palacky University, Olomouc, Czech Republic;
21. Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany;
22. Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital of Munich, Ludwig-Maximilians-Universität, Munich, Germany;
23. Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany;
24. Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research, Munich, Neuherberg, Germany;
25. The M. Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland;
26. Univ Rennes, Inserm, Ecole des hautes études en santé publique (EHESP), Irset (Institut de recherche en santé, environnement et travail) - UMR_S 1085, Pointe-à-Pitre, France;
27. National Institute of Public Health, Bucharest, Romania;
28. Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada;
29. Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and ll Centro di Riferimento per l'Epidemiologia e la Prevenzione Oncologica in Piemonte (CPO-Piemonte), Torino, Italy; 30. National Public Health Center, Budapest, Hungary;
31. Institut national de la recherche scientifique, University of Quebec, Laval, Quebec, Canada;
32. University of Montreal Hospital Research Centre, University of Montreal, Montreal, Quebec, Canada;
33. The Nofer Institute of Occupational Medicine, Lodz, Poland;
34. Institut für Medizinische Informatik Biometrie Epidemiologie, Ludwig Maximilians University, Munich, Germany;
35. Institut für Epidemiologie, Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg, Germany; and;
36. Russian Cancer Research Centre, Moscow, Russia

ABSTRACT

Rationale: Although the carcinogenicity of diesel engine exhaust has been demonstrated in multiple studies, little is known regarding exposure–response relationships associated with different exposure subgroups and different lung cancer subtypes.

Objectives: We expanded on a previous pooled case–control analysis on diesel engine exhaust and lung cancer by including three additional studies and quantitative exposure assessment to evaluate lung cancer and subtype risks associated with occupational exposure to diesel exhaust characterized by elemental carbon (EC) concentrations.

Methods: We used a quantitative EC job-exposure matrix for exposure assessment. Unconditional logistic regression models were used to calculate lung cancer odds ratios and 95% confidence intervals (CIs) associated with various metrics of EC exposure. Lung cancer excess lifetime risks (ELR) were calculated using life tables accounting for all-cause mortality. Additional stratified analyses by smoking history and lung cancer subtypes were performed in men.

Measurements and Main Results: Our study included 16,901 lung cancer cases and 20,965 control subjects. In men, exposure response between EC and lung cancer was observed: odds ratios ranged from 1.09 (95% CI, 1.00–1.18) to 1.41 (95% CI, 1.30–1.52) for the lowest and highest cumulative exposure groups, respectively. EC-exposed men had elevated risks in all lung cancer subtypes investigated; associations were strongest for squamous and small cell carcinomas and weaker for adenocarcinoma. EC lung cancer exposure response was observed in men regardless of smoking history, including in never-smokers. ELR associated with 45 years of EC exposure at 50, 20, and 1 μg/m^3 were 3.0%, 0.99%, and 0.04%, respectively, for both sexes combined.

Conclusions: We observed a consistent exposure–response relationship between EC exposure and lung cancer in men. Reduction of workplace EC levels to background environmental levels will further reduce lung cancer ELR in exposed workers.

INTRODUCTION

The International Agency for Research on Cancer (IARC) classifies diesel engine exhaust (hereafter diesel exhaust) as a group 1 human carcinogen (Benbrahim-Tallaa et al., 2012). Previous studies have provided consistent epidemiological evidence that lung cancer is associated with occupational exposure to diesel exhaust (Ilar et al., 2017; Olsson et al., 2011; Silverman et al., 2012; Steenland et al., 1998). Positive exposure–response relationships of diesel exhaust exposure and lung cancer were also reported by studies with quantitative exposure assessment for elemental carbon (EC), which is a measure of diesel exhaust exposure (Garshick et al., 2012; Silverman et al., 2012; Steenland et al., 1998; Vermeulen et al., 2014).

However, few studies have explored the risk of lung cancer associated with low exposure levels, and none have observed a positive association at lifetime cumulative EC exposure levels below 50 μg/m^3-years. Questions also remain regarding the role of cigarette smoking as a potential confounder or effect modifier in the relationship between EC exposure and lung cancer. For instance, although a handful of studies have shown suggestive elevated lung cancer risks in diesel exhaust–exposed workers who were never-smokers (Ilar et al., 2017; Parent et al., 2007; Villeneuve et al., 2011), only one study reported a significant effect (Silverman et al., 2012). The same study also reported attenuated lung cancer risk in subjects who were heavy smokers and highly exposed to diesel exhaust (i.e., a negative interaction). Finally, results reported by studies on risks of major lung cancer subtypes associated with diesel exhaust exposure have been inconsistent. Some studies reported the strongest association in large cell carcinoma compared with other major lung cancer subtypes (Ilar et al., 2017; Villeneuve et al., 2011), whereas others observed higher risks in squamous cell carcinoma (Parent et al., 2007; Pintos et al., 2012).

Previously, we published a study with pooled subjects from 11 lung cancer case–control studies from Europe and Canada (Olsson et al., 2011). In the current study, we increased the study population by including three additional studies (3,663 cases; 4,805 controls). Occupational exposure assessment was also enhanced with the use of a new job-exposure matrix (JEM), in which EC exposure was estimated quantitatively based on subject occupations. The purposes of our work were to evaluate *1*) the lung cancer risks associated with various indices of occupational diesel exhaust exposure by sex; *2*) the associations between diesel exhaust exposure and lung cancer by smoking status and cancer subtype in men; *3*) the joint effects of diesel exhaust exposure and smoking on the risk of lung cancer and its major subtypes on the additive and multiplicative scale in men; and *4*) the excess lifetime lung cancer risks associated with various levels of occupational diesel exhaust exposure in both sexes combined.

Methods

Study Population

Subjects from 14 hospital- and population-based lung cancer case–control studies in 13 European countries and Canada were pooled. A detailed description of the original study population is available elsewhere (Olsson et al., 2011). The current study updated the population with 3,663 cases and 4,805 controls from the TORONTO, CAPUA (Cáncer de Pulmón en Asturias), and ICARE (Investigation of Occupational and Environmental Causes of Respiratory Cancers) studies in Canada, Spain, and France, respectively (Supplementary Table 1). The project received ethical approvals from all participating countries and from the IARC institutional review board. More information about the SYNERGY project is available online at http://synergy.iarc.fr.

Job-Exposure Matrix and Exposure Assessment

A quantitative diesel engine exhaust JEM (DEE-JEM) was developed by C.G. and R.V. The DEE-JEM consists of EC exposure (in $\mu g/m^3$) assigned to all 1,506 five-digit International Standard Classification of Occupations (ISCO) (version 1968 or ISCO-68) (ILO, 2010) and was constructed based on 4,417 occupational EC measurements (data sources available in Supplementary Methods and Table 6). For occupations represented in the EC exposure measurements, the mean exposure concentrations were directly assigned. For occupations without measurement data, exposure concentrations from similar occupations with measurement data were assigned using expert decisions. An exposure probability factor was also assigned by expert decision to each exposed job (details on probability factors available in Supplementary Methods). The DEE-JEM was linked to study participant job histories by ISCO-68 occupations. Probability-weighted cumulative EC exposure (hereafter cumulative EC, expressed in $\mu g/m^3$-years) was calculated as the sum of the product of exposure levels, probabilities, and duration (in years) across all reported job periods for each subject. The DEE-JEM is available upon request from the corresponding author.

Main Statistical Analysis

Separately for men and women, unconditional logistic regression models were used to calculate the odds ratios (ORs) and 95% confidence intervals (CIs) of lung cancer associated with various categorical EC exposure metrics, including ever/never exposure, duration of exposure (<10, 10–19, 20–29, and >29 yr), and cumulative exposure (quartiles of exposure distribution among controls: >0–22, 23–70, 71–178, >178 $\mu g/m^3$-years). Trends were assessed using *P* values from the respective indices of EC exposure as continuous variables for all subjects and for exposed subjects only. Adjustments for the main analyses were determined *a priori* within the SYNERGY consortium and were identical with our previous occupational exposure publications (Olsson et al., 2011, 2017); these adjustments included study, age group (<45, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and >74 yr), smoking [log(cigarette pack-years + 1)], smoking cessation prior to interview/diagnosis (current smokers; >0–7, 8–15, 16–25, and >25 yr; and never-smokers), and having been ever employed in occupations with known lung cancer risks (list A jobs ever/never; full list in Supplementary Table 7). First published in 1982, list A jobs include occupations with definite lung cancer risks according to the IARC Monographs; the list was updated in 1995 and 2000 to cover all IARC-reviewed agents up to volume 75 of the Monographs (Ahrens & Merletti, 1998; Mirabelli et al., 2001). Smokers were defined as smoking more than one cigarette per day for more than 1 year. Smoking pack-years were calculated by summing the products of average daily smoking amount in 20-cigarette packs and smoking duration in years. Association between lung cancer and cumulative EC exposure as a continuous metric was assessed with a logistic linear regression model for men, women, and all subjects with identical adjustments as the categorical models. Models with various cumulative EC exposure lag times (i.e., omitting exposure in the last 5, 10, 15, or 20 years, or no omission at all) were constructed. According to minimized Akaike information criterion value, model fit was the best when lag time was 10 years—therefore, only results from models with a 10-year lag are presented.

Using the lung cancer risk from our linear continuous exposure model with all subjects, we calculated lung cancer excess lifetime risks (ELRs) at age 80 associated with 45 years of occupational EC exposure at 50, 20, and 1 $\mu g/m^3$ using life-table methods accounting for all-cause mortality outlined by Vermeulen and colleagues (Vermeulen et al., 2014). The selected exposure levels at 50, 20, and 1 $\mu g/m^3$ represented recommended limit values from the following: *1*) the German Committee for Hazardous Substances in 2017 based on a study on lung irritation after controlled human exposure (AGS, 2017); *2*) the U.S. National Institute of Occupational Safety and Health in 2003 that was later withdrawn (NIOSH, 2003); and *3*) the Health Council of the Netherlands in 2019 based on exposure–response estimates from Vermeulen and colleagues (Health Council of the Netherlands, 2019; Vermeulen et al., 2014), respectively. 2008 European data on mortality from all causes and lung cancer were used in our calculations (European Commission, 2008).

Extended Analysis for Male Subjects

To further investigate the exposure–response relationship between EC exposure and lung cancer in men, stratified analyses were performed to calculate lung cancer ORs associated with cumulative EC exposure categories with different major lung cancer subtypes and smoking histories. In addition, nonparametric thin-plate regression splines were created, as implemented in the R package mgcv, to visualize the shape of the exposure–response relationships between EC exposure and lung cancer subtypes in men. The number of basis functions was limited to three ($k = 3$), and the smoothing parameter was estimated using the relative maximum likelihood method. Spline model results were truncated at the 99th percentile of EC exposure to emphasize results with greater data support.

Additive interactions of cigarette smoking and EC exposure on lung cancer and subtype risks in men were assessed by calculating the excess risks due to interaction (RERI) using ORs from our logistic models as defined by Rothman and Greenland (Rothman & Greenland, 1998) and as implemented in the epi.interaction package in R. RERI values measure departure from additivity, with 0 representing no interaction on the additive scale (Knol et al., 2011). Interactions in men on the multiplicative scale were assessed using *P* values obtained from the cross products of smoking and EC exposure in the adjusted logistic models.

Statistical analyses were conducted using SAS (version 9.3; SAS Institute) and R (version 3.6).

RESULTS

A total of 37,866 subjects (16,901 cases; 20,965 controls) were included in our final analyses (Table 1). Among the lung cancer cases there were 4,752 adenocarcinomas, 810 large cell carcinomas, 2,730 small cell carcinomas, 6,503 squamous cell carcinomas, 2,012 other lung cancers, and 94 cases without subtype information.

In men, we observed elevated ORs for subjects with ever occupational exposure to EC (OR, 1.22; 95% CI, 1.15–1.29; Table 2). Increasing trends in lung cancer risks in men were associated with increases in both exposure duration and cumulative exposure (*P* trends < 0.01). Elevated male lung cancer ORs were also observed in the lowest categories of exposure duration (1–9 yr; OR, 1.07; 95% CI, 1.00–1.16) and cumulative exposure (>0–22 μg/m³-years; OR, 1.09; 95% CI, 1.00–1.19). In our female population, we observed no associations between lung cancer and different EC exposure metrics.

Table 1. Selected Study Population Characteristics by Lung Cancer Status and EC Exposure

	Ever Exposed to EC				Never Exposed to EC			
Characteristics	**Cases**	**%**	**Control Subjects**	**%**	**Cases**	**%**	**Control Subjects**	**%**
Sex								
M	8,045	95.4	8,181	94.1	5,560	65.6	8,270	67.4
F	386	4.6	512	5.9	2,910	34.4	4,002	32.6
Age group								
<45 yr	267	3.2	359	4.1	448	5.3	1,012	8.2
45–64 yr	4,195	49.8	4,120	47.4	4,568	53.9	6,234	50.8
>64 yr	3,969	47.1	4,214	48.5	3,454	40.8	5,026	41.0
Smoking status								
Never-smoker	379	4.5	2,287	26.3	990	11.7	4,866	39.7
Former smoker	2,966	35.2	3,880	44.6	2,466	29.1	4,340	35.4
Current smoker	5,086	60.3	2,526	29.1	5,014	59.2	3,066	25.0
Smoking pack-years								
Never-smoker	379	4.5	2,287	26.3	990	11.7	4,866	39.7
<10	381	4.5	1,287	14.8	428	5.1	1,782	14.5
10–19	765	9.1	1,206	13.9	837	9.9	1,656	13.5
>19	6,906	81.9	3,913	45.0	6,215	73.4	3,968	32.3
Years since quitting smoking								
Never-smoker	379	4.5	2,287	26.3	990	11.7	4,866	39.7
>0–7 yr	1,085	12.9	644	7.4	941	11.1	778	6.3
8–15 yr	836	9.9	883	10.2	695	8.2	1,015	8.3
16–25 yr	637	7.6	1,088	12.5	534	6.3	1,258	10.3
>25 yr	408	4.8	1,265	14.6	296	3.5	1,289	10.5
Current smoker	5,086	60.3	2,526	29.1	5,014	59.2	3,066	25.0
List A job								
Ever employment	1,143	13.6	866	10.0	644	7.6	498	4.1
Never employment	7,288	86.4	7,827	90.0	7,712	92.4	11,629	95.9
Lung cancer subtype								
Adenocarcinoma	1,953	23.2			2,799	33.0		
Large cell carcinoma	390	4.6			420	5.0		
Small cell carcinoma	1,427	16.9			1,303	15.4		
Squamous cell carcinoma	3,704	43.9			2,799	33.0		
Other/unspecified	914	10.8			1,098	13.0		
Not available	43	0.5			51	0.6		

Definition of abbreviation: EC = elemental carbon.

Table 2. Lung Cancer ORs Associated with Categorical Indices of Occupational EC Exposure

Occupational EC Exposure	Cases, *n* (%)	Controls, *n* (%)	OR*	95% CI
Men				
Never	5,560 (40.9)	8,270 (50.3)	1.0	Referent
Ever exposure	8,045 (59.1)	8,181 (49.7)	1.22	1.15–1.29
Duration, yr				
1–9	2,346 (17.2)	2,750 (16.7)	1.07	1.00–1.16
10–19	1,774 (13.0)	1,774 (10.8)	1.23	1.13–1.34
20–29	1,578 (11.6)	1,471 (8.9)	1.23	1.12–1.35
>29	2,347 (17.3)	2,186 (13.3)	1.39	1.28–1.51
Test for trend, P value†			<0.01	
P value excluding never exposed‡			<0.01	
Cumulative exposure, μg/m³-years				
>0–22	1,684 (12.4)	2,002 (12.2)	1.09	1.00–1.19
23–70	1,858 (13.7)	2,005 (12.2)	1.10	1.02–1.20
71–178	2,113 (15.5)	2,074 (12.6)	1.24	1.15–1.35
>178	2,390 (17.6)	2,100 (12.8)	1.43	1.32–1.54
Test for trend, P value†			<0.01	
P value excluding never exposed‡			<0.01	
Women				
Never	2,910 (88.3)	4,002 (88.7)	1.0	Referent
Ever exposure	386 (11.7)	512 (11.3)	1.00	0.85–1.18
Duration, yr				
1–9	235 (7.1)	273 (6.0)	1.02	0.83–1.26
10–19	86 (2.6)	112 (2.5)	1.07	0.77–1.47
20–29	25 (0.8)	49 (1.1)	0.69	0.39–1.17
>29	40 (1.2)	78 (1.7)	1.05	0.69–1.58
Test for trend, P value†			0.85	
P value excluding never exposed‡			0.74	
Cumulative exposure, μg/m3-years				
>0–22	165 (5.0)	179 (4.0)	1.03	0.80–1.33
23–70	118 (3.6)	162 (3.6)	1.03	0.78–1.36
71–178	64 (1.9)	99 (2.2)	0.92	0.64–1.31
>178	39 (1.2)	72 (1.6)	0.97	0.62–1.48
Test for trend, P value†			0.99	
P value excluding never exposed‡			0.82	

Definition of abbreviations: CI = confidence interval; EC = elemental carbon; OR = odds ratio.
*ORs are adjusted for study, age group, smoking pack-years [log(cigarette pack-years + 1)], time since quitting smoking, and list A jobs.
†Test for trend, *P* value obtained using the corresponding continuous exposure variable.
‡Test for trend, *P* value obtained using the corresponding continuous exposure variable among exposed subjects only.

Our continuous EC exposure models show that a 1 μg/m³-year increase in cumulative exposure was associated with an increase in lung cancer OR by a factor of 1.00001 (95% CI, 0.9987–1.00131) for women. The corresponding results for men and for all subjects were identical: lung cancer OR increased by a factor of 1.00034 (95% CI, 1.00021–1.00048) per μg/m³-years increase in cumulative EC exposure. Lung cancer ELRs associated with lifetime occupational EC exposure at 50, 20, and 1 μg/m³ were 3.0%, 0.99%, and 0.04%, respectively, for both sexes combined.

By lung cancer subtype, increasing cumulative EC exposure was associated with increasing ORs of squamous cell (*P* trend < 0.01) and small cell carcinomas (*P* trend = 0.02) in men (Table 3). For squamous cell carcinoma, all categories of cumulative EC exposure were associated with elevated ORs in men, including the lowest exposure (OR, 1.13; 95% CI, 1.01–1.26). The highest risks for both adenocarcinoma (OR, 1.23; 95% CI, 1.09–1.39) and large cell carcinoma (OR, 1.31; 95% CI, 1.02–1.67) were also observed in men in the group with the highest exposure.

Results from the nonparametric spline analyses for male subjects show monotonic increases in cancer risks for overall lung cancer and all four of the included subtypes (Figure 1). Among the lung cancer subtypes, squamous cell and small cell carcinomas show the strongest association with cumulative EC exposure, followed by large cell carcinoma and adenocarcinoma.

In our analyses stratified by smoking status, exposure–response associations between cumulative EC exposure and lung cancer were observed in men regardless of smoking history (Table 4). Lung cancer risks were similar for men in the highest EC exposure group who were never-smokers (OR, 1.41; 95% CI, 1.04–1.88), former smokers (OR, 1.47; 95% CI, 1.31–1.65), and current smokers (OR, 1.40; 95% CI, 1.24–1.57). Superadditive joint effects of smoking and EC exposure were observed in men for overall lung cancer and for all four cancer subtypes (Table 5). Suggestive super-multiplicative joint effects of smoking and EC exposure were observed for large cell carcinoma in men (*P* = 0.05).

Table 3. Lung cancer major subtype risks (OR) associated with cumulative occupational EC exposure in men

Cumulative EC Exposure, (µg/m³-years) by Lung Cancer Subtypes	Cases	OR*	95% CI
Adenocarcinoma			
Never	1513	1.0	Referent
>0–22	414	1.09	0.95–1.24
23–70	415	1.00	0.88–1.14
71–178	452	1.07	0.94–1.21
>178	531	1.23	1.09–1.39
Test for trend, p-value†		0.14	
Excl. never exposed‡		0.49	
Large cell carcinoma			
Never	257	1.0	Referent
>0–22	84	1.04	0.79–1.36
23–70	76	0.90	0.68–1.18
71–178	93	1.14	0.88–1.47
>178	109	1.31	1.02–1.67
Test for trend, p-value†		0.11	
Excl. never exposed‡		0.14	
Squamous cell carcinoma			
Never	2216	1.0	Referent
>0–22	742	1.13	1.01–1.26
23–70	819	1.14	1.03–1.27
71–178	982	1.37	1.24–1.52
>178	1069	1.54	1.39–1.70
Test for trend, p-value†		<0.01	
Excl. never exposed‡		0.01	
Small cell carcinoma			
Never	850	1.0	Referent
>0–22	249	0.99	0.84–1.16
23–70	334	1.20	1.03–1.39
71–178	360	1.31	1.14–1.53
>178	407	1.53	1.32–1.76
Test for trend, p-value†		0.02	
Excl. never exposed‡		0.39	

For definition of abbreviations, see Table 2.
*ORs adjusted for study, age group, smoking pack-years [log(cigarette pack-years + 1)], time since quitting smoking, and list A jobs.
†Test for trend, P value obtained using the corresponding continuous exposure variable.
‡Test for trend, P value obtained using the corresponding continuous exposure variable among exposed subjects only.

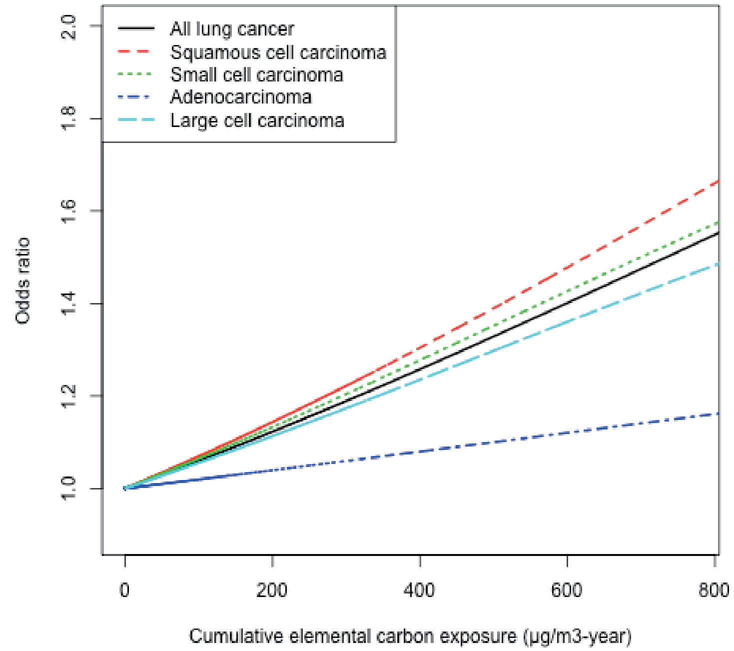


Figure 1. Spline analyses showing exposure–response relationships in men between cumulative elemental carbon exposure and risks of overall lung cancer plus subtypes.

Table 4. Lung Cancer Risks Associated with Cumulative Occupational EC Exposure by Smoking Status in Men

	Never-Smokers			Former Smokers			Current Smokers		
Cumulative EC exposure (µg/m³-years)	**Cases**	**OR***	**95% CI**	**Cases**	**OR†**	**95% CI**	**Cases**	**OR‡**	**95% CI**
Never	256	1.0	Referent	1,868	1.0	Referent	3,436	1.0	Referent
>0–22	66	1.40	1.03–1.88	624	1.11	0.98–1.26	994	1.04	0.92–1.18
23–70	41	0.94	0.65–1.33	656	1.23	1.09–1.40	1,161	1.01	0.90–1.14
71–178	55	1.17	0.85–1.60	764	1.33	1.18–1.50	1,294	1.15	1.03–1.29
>178	72	1.41	1.04–1.88	875	1.47	1.31–1.65	1,443	1.40	1.24–1.57
Test for trend, P value[x]		0.03			<0.01			<0.01	
P value excluding never exposed[k]		0.11			0.08			0.05	

For definition of abbreviations, *see* Table 2.
*ORs adjusted for study, age group, and list A jobs.
†ORs adjusted for study, age group, list A jobs, smoking pack-years [log(cigarette pack-years + 1)], and time since quitting smoking.
‡ORs adjusted for study, age group, list A jobs, and smoking pack-years [log(cigarette pack-years + 1)].
[x]Test for trend, *P* value obtained using the corresponding continuous exposure variable.
[k]Test for trend, *P* value obtained using the corresponding continuous exposure variable among exposed subjects only.

Table 5: Interactions between occupational elemental carbon (EC) exposure and smoking for overall lung cancer and major subtypes in men

Exposure Status by Lung Cancer Subtype	Controls	Cases	OR*	95%CI
All lung cancers				
Never Smoker & Never EC	2525	256	1.0	Referent
Never Smoker & Ever EC	1912	234	1.14	0.95–1.38
Ever Smoker & Never EC	5745	5304	8.71	7.62–10.0
Ever Smoker & Ever EC	6269	7811	11.4	9.93–13.0
p-value multiplicative†			0.18	
RERI‡			2.49	1.92–3.07
Adenocarcinoma				
Never Smoker & Never EC	2525	100	1.0	Referent
Never Smoker & Ever EC	1912	79	1.05	0.77–1.42
Ever Smoker & Never EC	5745	1413	6.14	4.99–7.63
Ever Smoker & Ever EC	6269	1733	7.22	5.87–8.98
p-value multiplicative†			0.47	
RERI‡			1.03	0.43–1.63
Large cell carcinoma				
Never Smoker & Never EC	2525	14	1.0	Referent
Never Smoker & Ever EC	1912	5	0.43	0.14–1.14
Ever Smoker & Never EC	5745	243	7.57	4.57–13.7
Ever Smoker & Ever EC	6269	357	9.35	5.66–16.8
p-value multiplicative†			0.05	
RERI‡			2.34	0.67–4.02
Squamous cell carcinoma				
Never Smoker & Never EC	2525	64	1.0	Referent
Never Smoker & Ever EC	1912	77	1.38	0.98–1.94
Ever Smoker & Never EC	5745	2152	13.4	10.5–17.1
Ever Smoker & Ever EC	6269	3535	18.1	14.4–24.0
p-value multiplicative†			0.99	
RERI‡			4.66	3.23–6.09
Small cell carcinoma				
Never Smoker & Never EC	2525	26	1.0	Referent
Never Smoker & Ever EC	1912	30	1.38	0.81–2.36
Ever Smoker & Never EC	5745	824	13.5	9.32–20.6
Ever Smoker & Ever EC	6269	1320	18.5	12.8–28.1
p-value multiplicative†			0.96	
RERI‡			4.56	2.42–6.69

Definition of abbreviations: CI = confidence interval; EC = elemental carbon; OR= odds ratio; RERI = excess risks due to interaction.
*ORs adjusted for study, age group, and list A jobs.
†P value for the EC and smoking interaction cross product term coefficient in fully adjusted logistic models; interaction on the multiplicative scale is present when $P < 0.05$.
‡Interaction on the additive scale is present when RERI deviates from 0.

DISCUSSION

In a large pooled case–control population, we observed positive associations between lung cancer and different occupational EC exposure metrics, including ever EC exposure, exposure duration, and cumulative exposure, in men. Increasing exposure duration and cumulative exposure were associated with increases in lung cancer risks in men, exhibiting monotonic exposure–response relationships. Our results are in accordance with, and further expand on, results from our earlier analysis within the SYNERGY study with 11 studies and semiquantitative exposure assessment, in which we reported a consistent exposure–response relationship between lung cancer and EC exposure (Olsson et al., 2011). Additional evidence of the exposure–response relationship between diesel exhaust exposure and lung cancer is provided by studies on workers in highly exposed industries such as mining (Silverman et al., 2012; Attfield et al., 2012; Neumeyer-Gromen et al., 2009; Säverin et al., 1999) and trucking (Garshick et al., 2012; Steenland et al., 1998).

In a metaregression analysis of the exposure–response relationship of lung cancer and diesel exhaust exposure based on data from three occupational cohort studies, Vermeulen and colleagues estimated that each $\mu g/m^3$-year increase in cumulative EC exposure results in a lung cancer relative risk (RR) of 1.00098 (Vermeulen et al., 2014). A subsequent sensitivity analysis reported a range of lung cancer RR of 1.0006 to 1.0012 per $\mu g/m^3$-years increase in cumulative EC exposure from several alternative models (Vermeulen & Portengen, 2016). These exposure–response slope estimates are approximately 2–3 times higher than our present linear model estimate of 1.00034 for all subjects. This difference may be due to factors such as occupational cohorts having higher cumulative EC exposures and more accurate exposure assessment in specific industries. Despite the differences on the exact risk magnitude, a consistent exposure–response trend between occupational diesel exhaust exposure and lung cancer was reported by studies with different designs among different populations.

We did not observe an exposure threshold for diesel exhaust–related lung cancer in men within the cumulative EC exposure ranges we investigated; increased lung cancer risk in men was observed in the lowest cumulative EC exposure group, with a median exposure of 11 $\mu g/m^3$-years. An additional sensitivity analysis with 10 cumulative exposure groups suggested (naturally, with less precision) an increased risk among the lowest exposure group with a median EC exposure of 3.3 $\mu g/m^3$-years (Supplementary Table 2). Few other studies investigated lung cancer risks in similar cumulative EC exposure ranges quantitatively. In occupational cohorts with higher EC exposures, one study reported a lung cancer OR of 1.31 (95% CI, 1.01–1.71) in U.S. trucking workers with a cumulative exposure of approximately 51 µg/m-years (Garshick et al., 2012), whereas another study reported a lung cancer OR of 0.74 (95% CI, 0.40–1.38) for U.S. miners with a cumulative EC exposure around 37 µg/m^3-years (Silverman et al., 2012).

We found that diesel exhaust exposure was associated with all four major lung in men, although differential risks were observed by subtype. Both our logistic regression and spline models showed that the associations were the strongest for squamous cell and small cell carcinomas, moderate for large cell carcinoma, and weakest for adenocarcinoma. Similar findings supportive of a stronger link between diesel exhaust exposure and lung squamous cell carcinoma were reported in populations in Canada (Parent et al., 2007; Pintos et al., 2012; Villeneuve et al., 2011), Finland (Guo et al., 2004), and Sweden (Boffetta et al., 2001; Ilar et al., 2017). This is the first report of a positive exposure–response relationship for diesel exhaust exposure and lung small cell carcinoma in men. Guo and colleagues observed a small cell carcinoma OR of 2.31 (95% CI, 1.02–5.25) for female Finnish workers in the low diesel exhaust exposure category based on six exposed cases (25). Elevated point estimates of small cell carcinoma risks were also observed in population-based studies from different countries (Guo et al., 2004; Ilar et al., 2017; Pintos et al., 2012). For adenocarcinoma, in accordance with our current observations, previous studies were consistent in reporting ORs that were lower than overall lung cancer risks (Boffetta et al., 2001; Guo et al., 2004; Ilar et al., 2017; Parent et al., 2007; Pintos et al., 2012; Villeneuve et al., 2011). Information on the risk of large cell carcinoma related to diesel exhaust exposure is limited; only two previous studies included large cell carcinoma in sub-type analyses (Ilar et al., 2017; Villeneuve et al., 2011). These studies reported exposure–response relationships for duration, intensity, and lifetime cumulative exposure to diesel exhaust and large cell carcinoma. In our male population, we observed a clear increased large cell carcinoma risk only in the group with the highest cumulative EC exposure (>178 µg/m^3-years), with a suggestive elevated OR estimate for the second highest exposed group.

We observed a lung cancer exposure–response risk trend in never-smoking men who were exposed to EC. Similarly, Silverman and colleagues reported a significant lung cancer OR of 7.30 (95% CI, 1.46–36.57) among highly exposed U.S. miners who never smoked (Silverman et al., 2012). The very high risk observed in the U.S. miners may be attributable to higher cumulative EC exposure in mining occupations or the fact that the estimate was based on only seven exposed cases.

The observed superadditive joint effects between EC exposure and smoking for overall lung cancer and its subtypes in men indicate that the absolute risk of cancer for men exposed to both EC and smoking was higher than the sum of the absolute risks of cancer from EC exposure and smoking alone (VanderWeele & Knol, 2014). Only one other study in Swedish dock workers investigated EC and smoking interaction on the additive scale and similarly reported a superadditive effect (Emmelin et al., 1993). Interaction in other studies was assessed on the multiplicative scale, in which supermultiplicative interaction represents a scenario in which the risk ratios (e.g., OR) of cancer for those exposed to both EC and smoking was higher than the product of the cancer risk ratios from EC exposure and smoking alone (VanderWeele & Knol, 2014). In two nonoverlapping Canadian population- based case–control studies, no significant multiplicative interaction was observed (Pintos et al., 2012; Villeneuve et al., 2011). Lastly, in the U.S. Miners Study, Silverman and colleagues reported a suggestive submultiplicative interaction, in which high exposure to both EC and cigarette smoke resulted in an attenuation of lung cancer risk increase (Silverman et al., 2012). In additional analyses wherein we explored cancer risks in four groups of male smokers (<10, 10–19, 20–39, and >39 pack-years, respectively) with cumulative EC exposures similar to those in Silverman and colleagues, we did not observe submultiplicative interactive effects and found consistent risk increases across all EC exposure categories for subjects with increasing pack-years of smoking (Supplementary Table 3).

Strengths of our study include a large pooled population with detailed smoking and occupational histories. Our sample size allowed for stratified analyses to explore the exposure–response relationship in different subgroups, whereas high-quality smoking and occupational histories allowed for the control of important potential confounders such as smoking and exposure to other occupational carcinogens. Exposure assessment was performed with a quantitative JEM developed using a combination of exposure measurements and expert assessment. The current DEE-JEM was developed independently from the Domtoren-JEM (DOM-JEM), an expert judgment JEM we used in an earlier analysis (Olsson et al., 2011). Despite this difference, results of both analyses showed consistent exposure–response relationships between occupational exposure to diesel exhaust and lung cancer. Reliability studies on occupational exposure assessment also suggested that incorporating measurements in the exposure assessment process may improve expert judgment (Ge et al., 2018; Teschke et al., 2002). Finally, the exposure–response relationship between EC exposure and lung cancer in our male population was robust and present in various sensitivity analyses, including when we limited analyses to a more homogeneous group of studies, when we limited our analyses to blue-collar workers only, and when we assessed EC exposure with alternative JEM configurations (Supplementary Tables 4.1-4.9).

There are also limitations in our work. Our DEE-JEM did not account for changes in exposure at different time periods and therefore may underestimate exposure for earlier periods when exposure was likely higher (Plato et al., 2019). The EC measurements used in our JEM were collected from 1985 to 2016 (median, 2002), whereas our subjects were assessed as exposed from 1923 to 2020 (median, 1968). However, the association between EC exposure and lung cancer was still present when we restricted our analyses to subjects exposed after 1960 (Supplementary Table 4.2). Because list A jobs included some jobs with potential diesel exhaust exposure, adjustment for ever-employment in any list A jobs in our main model may represent overadjustment for coexposures to other lung carcinogens. Removing all jobs with EC exposure from list A, however, may lead to underadjustment because many EC-exposed jobs have concurrent exposures to other lung carcinogens. We explored the coexposure adjustments using two additional sensitivity models: one with no adjustment and another adjusting for ever exposure to crystalline silica, asbestos, polycyclic aromatic hydrocarbons, and hexavalent chromium as assessed by the DOM-JEM (Supplementary Table 4.4). All three categorical EC models (i.e., the main model and the two sensitivity models) showed the EC-exposure–lung cancer response among men, suggesting that the association is unlikely to be fully explained by confounding due to exposures to other occupational lung carcinogens. Furthermore, because our JEM assigned EC exposures based on job titles, individual exposures may be misclassified in occupations with large exposure variability. This misclassification, however, was not likely to be differential by case status and introduced Berkson-like error that likely affected the precision, but not magnitude, of our risk estimates (Armstrong, 1990; Heid et al., 2004). Exposure misclassification of jobs within the DEE-JEM may also have occurred because ofthe fact that our EC exposure data were limited and did not represent all jobs in all study regions. If present, this would introduce classical error in our work and bias the observed effect toward the null, meaning that the true effect of diesel exhaust exposure on lung cancer may be stronger than our observed results. However, the aforementioned shortcomings related to retrospective exposure assessment are almost inevitable because of our study design and size. We have provided details on all data sources, assessment procedures, and various sensitivity analyses in an effort to maximize transparency.

Another notable limitation of our study is the lower statistical power to assess risk in female workers (390 exposed cases) compared with males (7,843 exposed cases). Our results on female cancer risks may also have been affected by more exposure misclassification of women compared with men because the supporting EC exposure data were collected almost exclusively among male workers. Adenocarcinoma, for which we observed the weakest association with diesel exhaust exposure among the lung cancer subtypes, was also more common in women than in men. However, our results should not be interpreted as diesel exhaust having no effect on lung cancer risks in women. A sensitivity analysis among women with lung cancer subtypes other than adenocarcinoma showed increased OR point estimates for cancer for all cumulative EC exposure groups, albeit with larger uncertainties (Supplementary Table 4.9).

In risk assessment for occupational carcinogen exposure, definitions for tolerable ELR range from 4 in 1,000 (0.4%) in the Netherlands and Germany to 1 in 1,000 (0.1%) in the United States (AGS, 2019; Health Council of the Netherlands, 2019; Rodricks et al., 1987). Of our three ELR estimates derived from different exposure limits, only the scenario with 1 $\mu g/m^3$ EC exposure and 0.04% ELR is below these levels. Another study using data from the U.S. trucking industry estimated that male workers exposed to 5 $\mu g/m^3$ EC would have a lung cancer ELR of 1–2% (Steenland et al., 1998). A separate study calculated a lung cancer ELR of 0.17% for workers exposed to 1 $\mu g/m^3$ EC using data from three U.S. mining and trucking industry cohorts (Vermeulen et al., 2014). Despite variations in the exact risk magnitude, estimates from different studies suggest that workplace EC levels should be at or near environmental background levels to reduce the lung cancer ELR for workers with lifetime exposure to diesel exhaust to tolerable levels, as defined by various national risk assessment agencies. Although multiple diesel engine emission control standards have been introduced in Europe since 2006 (Health Council of the Netherlands, 2019), these standards alone cannot be expected to reduce workplace EC exposure to environmental levels in the near future because they do not apply to the large number of existing diesel equipment that still is and will probably remain in use for many more years.

In summary, we observed a consistent exposure–response relationship between occupational diesel exhaust exposure and lung cancer in men in a large pooled analysis of case–control studies. Increased lung cancer risks were found in EC-exposed men who were never-smokers and smokers. Increased risks in men were also observed for all lung cancer subtypes included, with the strongest associations for squamous cell and small cell carcinomas and weaker for adenocarcinoma. The joint effects of EC exposure and smoking were superadditive on risks of overall lung cancer and all included subtypes. Our findings support efforts to further reduce workplace diesel exhaust exposure to protect workers against risks of lung cancer.

REFERENCES

AGS. (2017). Dieselmotoremissionen (DME). Begründung zu für Dieselmotoremissionen (DME) in TRGS 900 (Justification for diesel engine emissions (DME) in TRGS 900—In German). https://www.baua.de/DE/Angebote/Rechtstexte-und-Technische-Regeln/Regelwerk/TRGS/pdf/900/900-dieselmotorenemissionen-dme-russpartikel-als-ec.pdf?__blob=publicationFile&v=5

AGS. (2019, March 29). TRGS 910 Risikobezogenes Maßnahmenkonzept für Tätigkeiten mit krebserzeugenden Gefahrstoffen (Technical Rules for Hazardous Substances 910: Risk-based action plan for activities with carcinogenic hazardous substances—In German). https://www.baua.de/DE/Angebote/Rechtstexte-und-Technische-Regeln/Regelwerk/TRGS/pdf/TRGS-910.pdf?__blob=publicationFile&v=4

Ahrens, W., & Merletti, F. (1998). A Standard Tool for the Analysis of Occupational Lung Cancer in Epidemiologic Studies. International Journal of Occupational and Environmental Health, 4(4), 236–240. https://doi.org/10.1179/oeh.1998.4.4.236

Armstrong, B. G. (1990). THE EFFECTS OF MEASUREMENT ERRORS ON RELATWE RISK REGRESSIONS. American Journal of Epidemiology, 132(6), 1176–1184. https://doi.org/10.1093/oxfordjournals.aje.a115761

Attfield, M. D., Schleiff, P. L., Lubin, J. H., Blair, A., Stewart, P. A., Vermeulen, R., Coble, J. B., & Silverman, D. T. (2012). The Diesel Exhaust in Miners study: A cohort mortality study with emphasis on lung cancer. Journal of the National Cancer Institute, 104(11), 869–883. https://doi.org/10.1093/jnci/djs035

Benbrahim-Tallaa, L., Baan, R. A., Grosse, Y., Lauby-Secretan, B., Ghissassi, F. E., Bouvard, V., Guha, N., Loomis, D., & Straif, K. (2012). Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. The Lancet Oncology, 13(7), 663–664. https://doi.org/10.1016/S1470-2045(12)70280-2

Boffetta, P., Dosemeci, M., Gridley, G., Bath, H., Moradi, T., & Silverman, D. (2001). Occupational exposure to diesel engine emissions and risk of cancer in Swedish men and women. Cancer Causes & Control: CCC, 12(4), 365–374.

Emmelin, A., Nyström, L., & Wall, S. (1993). Diesel exhaust exposure and smoking: A case-referent study of lung cancer among Swedish dock workers. Epidemiology (Cambridge, Mass.), 4(3), 237–244.

European Commission. (2008). Eurostat 2008 dataset on all causes and lung cancer mortality in European Union countries. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_cd_aro&lang=en.

Garshick, E., Laden, F., Hart, J. E., Davis, M. E., Eisen, E. A., & Smith, T. J. (2012). Lung cancer and elemental carbon exposure in trucking industry workers. Environmental Health Perspectives, 120(9), 1301–1306. https://doi.org/10.1289/ehp.1204989

Ge, C. B., Friesen, M. C., Kromhout, H., Peters, S., Rothman, N., Lan, Q., & Vermeulen, R. (2018). Use and Reliability of Exposure Assessment Methods in Occupational Case–Control Studies in the General Population: Past, Present, and Future. Annals of Work Exposures and Health, 62(9), 1047–1063. https://doi.org/10.1093/annweh/wxy080

Guo, J., Kauppinen, T., Kyyrönen, P., Lindbohm, M.-L., Heikkilä, P., & Pukkala, E. (2004). Occupational exposure to diesel and gasoline engine exhausts and risk of lung cancer among Finnish workers. American Journal of Industrial Medicine, 45(6), 483–490. https://doi.org/10.1002/ajim.20013

Health Council of the Netherlands. (2019, March 13). Diesel Engine Exhaust: Health-based recommended occupational exposure limit. https://www.gezondheidsraad.nl/binaries/gezondheidsraad/documenten/adviezen/2019/03/13/dieselmotoremissie/Diesel+Engine+Exhaust.pdf

Heid, I. M., Küchenhoff, H., Miles, J., Kreienbrock, L., & Wichmann, H. E. (2004). Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. Journal of Exposure Science and Environmental Epidemiology, 14(5), 365–377. https://doi.org/10.1038/sj.jea.7500332

Ilar, A., Plato, N., Lewné, M., Pershagen, G., & Gustavsson, P. (2017). Occupational exposure to diesel motor exhaust and risk of lung cancer by histological subtype: A population-based case-control study in Swedish men. European Journal of Epidemiology, 32(8), 711–719. https://doi.org/10.1007/s10654-017-0268-5

ILO. (2010, June 10). ISCO-International Standard Classification of Occupations: Brief History. http://www.ilo.org/public/english/bureau/stat/isco/intro2.htm

Knol, M. J., VanderWeele, T. J., Groenwold, R. H. H., Klungel, O. H., Rovers, M. M., & Grobbee, D. E. (2011). Estimating measures of interaction on an additive scale for preventive exposures. European Journal of Epidemiology, 26(6), 433–438. https://doi.org/10.1007/s10654-011-9554-9

Mirabelli, D., Chiusolo, M., Calisti, R., Massacesi, S., Richiardi, L., Nesti, M., & Merletti, F. (2001). [Database of occupations and industrial activities that involve the risk of pulmonary tumors]. Epidemiologia E Prevenzione, 25(4–5), 215–221.

Neumeyer-Gromen, A., Razum, O., Kersten, N., Seidler, A., & Zeeb, H. (2009). Diesel motor emissions and lung cancer mortality—Results of the second follow-up of a cohort study in potash miners. International Journal of Cancer, 124(8), 1900–1906. https://doi.org/10.1002/ijc.24127

NIOSH. (2003). NIOSH manual of analytical methods (NMAM) fourth edition. Third supplement. (P. Schlecht & P. O'Connor, Eds.). https://www.cdc.gov/niosh/nioshtic-2/20024429.html

Olsson, A. C., Gustavsson, P., Kromhout, H., Peters, S., Vermeulen, R., Brüske, I., Pesch, B., Siemiatycki, J., Pintos, J., Brüning, T., Cassidy, A., Wichmann, H.-E., Consonni, D., Landi, M. T., Caporaso, N., Plato, N., Merletti, F., Mirabelli, D., Richiardi, L., … Straif, K. (2011). Exposure to diesel motor exhaust and lung cancer risk in a pooled analysis from case-control studies in Europe and Canada. American Journal of Respiratory and Critical Care Medicine, 183(7), 941–948. https://doi.org/10.1164/rccm.201006-0940OC

Olsson, A. C., Vermeulen, R., Schüz, J., Kromhout, H., Pesch, B., Peters, S., Behrens, T., Portengen, L., Mirabelli, D., Gustavsson, P., Kendzia, B., Almansa, J., Luzon, V., Vlaanderen, J., Stücker, I., Guida, F., Consonni, D., Caporaso, N., Landi, M. T., … Straif, K. (2017). Exposure–Response Analyses of Asbestos and Lung Cancer Subtypes in a Pooled Analysis of Case–Control Studies. Epidemiology (Cambridge, Mass.), 28(2), 288–299. https://doi.org/10.1097/EDE.0000000000000604

Parent, M.-E., Rousseau, M.-C., Boffetta, P., Cohen, A., & Siemiatycki, J. (2007). Exposure to diesel and gasoline engine emissions and the risk of lung cancer. American Journal of Epidemiology, 165(1), 53–62. https://doi.org/10.1093/aje/kwj343

Pintos, J., Parent, M.-E., Richardson, L., & Siemiatycki, J. (2012). Occupational exposure to diesel engine emissions and risk of lung cancer: Evidence from two case-control studies in Montreal, Canada. Occupational and Environmental Medicine, 69(11), 787–792. https://doi.org/10.1136/oemed-2012-100964

Plato, N., Lewné, M., & Gustavsson, P. (2019). A historical job-exposure matrix for occupational exposure to diesel exhaust using elemental carbon as an indicator of exposure. Archives of Environmental & Occupational Health, 0(0), 1–12. https://doi.org/10.1080/19338244.2019.1644277

Rodricks, J. V., Brett, S. M., & Wrenn, G. C. (1987). Significant risk decisions in federal regulatory agencies. Regulatory Toxicology and Pharmacology, 7(3), 307–320. https://doi.org/10.1016/0273-2300(87)90038-9

Rothman, K., & Greenland, S. (1998). Modern Epidemiology. Lippincott - Raven.

Säverin, R., Bräunlich, A., Dahmann, D., Enderlein, G., & Heuchert, G. (1999). Diesel exhaust and lung cancer mortality in potash mining. American Journal of Industrial Medicine, 36(4), 415–422. https://doi.org/10.1002/(SICI)1097-0274(199910)36:4<415::AID-AJIM2>3.0.CO;2-Q

Silverman, D. T., Samanic, C. M., Lubin, J. H., Blair, A. E., Stewart, P. A., Vermeulen, R., Coble, J. B., Rothman, N., Schleiff, P. L., Travis, W. D., Ziegler, R. G., Wacholder, S., & Attfield, M. D. (2012). The Diesel Exhaust in Miners study: A nested case-control study of lung cancer and diesel exhaust. Journal of the National Cancer Institute, 104(11), 855–868. https://doi.org/10.1093/jnci/djs034

Steenland, K., Deddens, J., & Stayner, L. (1998). Diesel exhaust and lung cancer in the trucking industry: Exposure–response analyses and risk assessment. American Journal of Industrial Medicine, 34(3), 220–228. https://doi.org/10.1002/(SICI)1097-0274(199809)34:3<220::AID-AJIM3>3.0.CO;2-Z

Teschke, K., Olshan, A. F., Daniels, J. L., Roos, A. J. D., Parks, C. G., Schulz, M., & Vaughan, T. L. (2002). Occupational exposure assessment in case–control studies: Opportunities for improvement. Occupational and Environmental Medicine, 59(9), 575–594. https://doi.org/10.1136/oem.59.9.575

VanderWeele, T. J., & Knol, M. J. (2014). A Tutorial on Interaction. Epidemiologic Methods, 3(1), 33–72. https://doi.org/10.1515/em-2013-0005

Vermeulen, R., & Portengen, L. (2016). Is diesel equipment in the workplace safe or not? Occupational and Environmental Medicine, 73(12), 846–848. https://doi.org/10.1136/oemed-2016-103977

Vermeulen, R., Silverman, D. T., Garshick, E., Vlaanderen, J., Portengen, L., & Steenland, K. (2014). Exposure-response estimates for diesel engine exhaust and lung cancer mortality based on data from three occupational cohorts. Environmental Health Perspectives, 122(2), 172–177. https://doi.org/10.1289/ehp.1306880

Villeneuve, P. J., Parent, M.-É., Sahni, V., Johnson, K. C., & Canadian Cancer Registries Epidemiology Research Group. (2011). Occupational exposure to diesel and gasoline emissions and lung cancer in Canadian men. Environmental Research, 111(5), 727–735. https://doi.org/10.1016/j.envres.2011.04.003

SUPPLEMENTARY METHODS

Elemental carbon (EC) data sources and additional description for the DEE-JEM

We chose EC as an exposure proxy for diesel engine exhaust because of its high specificity to diesel engine emissions and general acceptance as the best marker for diesel engine exhaust (Health Council of the Netherlands, 2019). The occupational EC exposure measurements for the JEM were obtained from three sources. Studies published from 1957 to 2007 that were included in an earlier review of EC occupational exposure by Pronk and colleagues (Pronk et al., 2009). An additional literature review was performed in the MEDLINE database for studies with EC measurements published between January 1st 2008 and May 31st 2017. Specifically, Medical Subject Headings (MeSH) terms "vehicle emissions" and "occupational exposures" were used in conjunction with all fields keywords "elemental carbon" and "diesel" to search for studies containing EC measurements. The search resulted in 34 matches and 9 publications contained relevant EC measurements for extraction (Bakke et al., 2014; Debia et al., 2016; Elihn et al., 2008; Galea et al., 2016; Hewett & Bullock, 2014; Lan et al., 2015; Lee et al., 2015; Sheesley et al., 2008; Shih et al., 2008). Two additional reports on EC exposures in firefighters were also added (Bott, 2010; Shih et al., 2008). Finally, occupational EC measurements from the UK Health and Safety Executive (HSE) National Exposure Database (NEDB) were also screened for extraction (HSE, 2019). For inclusion in our JEM, EC measurements had to be: 1) personal measurements or area measurements representative of personal exposure (e.g. inside a vehicle cabin); 2) sampled with duration longer than 1 hour; 3) representative of typical exposures experienced by workers (i.e. not worst-case or complaint-driven sampling); and 4) taken in actual workplaces rather than other simulated controlled settings. In total, 3,528 EC measurements were extracted from studies covered by the review by Pronk and colleagues, 700 were extracted from the additional literature review, and 189 were extracted from the NEDB. The EC measurements included 2,066 in the respirable fraction, 1,333 in the submicron fraction, 665 in the inhalable fraction, and 353 with no size fraction information. Measurements of all size fractions were treated equally as studies suggest the submicron size fraction captures approximately 75% of EC particulates whereas respirable and larger size fractions captures nearly all EC (Verma et al., 2003; Vermeulen et al., 2010). Sampling year for EC measurements used to construct the JEM ranged from 1985 to 2016 (median: 2002). Additional information on all EC measurements used for the DEE-JEM, including occupation, country, and sampling year, is available in Supplementary Table 6.

Assigned probabilities in the DEE-JEM consisted of one of three values in 0.1, 0.25, 0.5 and were given based on expert decision by two experts (CG, RV) consecutively. Probabilities were only assigned to occupations where the experts were confident that EC exposure does not occur for all workers with the same job title. A few ISCO-68 occupations at the 2- or 3-digit level received probabilities of 0.4 (n=3) and 0.6 (n=4) as median values of probabilities assigned to their respective 5-digit daughter occupations. In total, the DEE-JEM assigned EC exposure to 248 of 1,506 ISCO-68 jobs. Probability factors for these jobs were: 0.1 for 12 jobs, 0.25 for 84 jobs, 0.4 for 3 jobs, 0.5 for 46 jobs, 0.6 for 4 jobs and 1.0 for 100 jobs.

Sensitivity analyses

Stratified models were used to assess if cancer risks associated with cumulative EC exposure categories differed between population- versus hospital-based case control studies in men. Restricted models were created for male blue-collar workers and workers employed after 1960 to investigate whether cancer risks differed for workers with lower socioeconomic status and for workers whose exposures were more recent when diesel equipment became more common in the workplace, respectively. Because miners and farmers may account for large proportions of the exposed population and may have different exposure patterns than other occupations, restricted analyses were performed on the male study population without those ever-employed in mining and agriculture industries to see if risks differed compared to our main analyses. As an alternative to List-A job adjustment for exposures to other lung carcinogens, we controlled for ever exposure to asbestos, crystalline silica, hexavalent chromium, and polycyclic aromatic hydrocarbons (PAHs) as assessed by the DOM-JEM (Peters et al., 2011) in our main categorical exposure model for men. Heterogeneity in lung cancer ORs in men associated with ever EC exposure between 14 studies was measured using the p-value of the Cochran's Q statistic and as a percentage in I^2 (Higgins et al., 2003).

To assess the impact of various decisions during the development of the DEE-JEM, we also carried out multiple sensitivity analyses with different JEM configurations. In our male categorical cumulative EC exposure model, we tested the impact of including expert-assigned probabilities by using a JEM with no probabilities (i.e. all probabilities=1 for exposed job titles) and a JEM with no expert-assigned probabilities <1. We also tested the same model with a JEM with EC measurement data restricted in the respirable size fraction to see if this changes the findings obtained from the JEM with EC data in various size fractions.

To further explore lung cancer risks in women related to EC exposure, we limited our cumulative EC exposure model to women with lung cancer subtypes other than adenocarcinoma. Additional analysis to calculate lung cancer OR and 95% CIs associated with time-since-last-exposure (<10; 10–19; 20–29; 30–39; >39 years) for men and women separately, with similar adjustments as our main analyses. Trends were assessed using p-values from the respective indices of EC exposure as continuous variables for exposed subjects only.

SUPPLEMENTARY RESULTS

Sensitivity analyses

We observed associations between cumulative EC exposure and lung cancer in all stratified and restricted sensitivity analyses in men (Supplementary Tables 4.1-4.5). Associations were similar or stronger compared to our main models in models restricted to studies with population controls and models restricted to subjects who never worked in agriculture. Risk estimates were more attenuated and less precise in models restricted to studies with hospital controls, subjects who were blue-collar workers, workers employed after 1960, workers who were never-miners, as well as in the model with alternative control for exposure to other occupational lung carcinogens.

Heterogeneity was observed in the lung cancer ORs related to ever EC exposure in the 14 included studies (I^2=50%; Q=40; p<0.01). Significant reduction in heterogeneity was observed (I^2=18%; Q=24; p=0.13) in the remaining subgroup after excluding two studies: AUT and PARIS. Exposure-response patterns between lung cancer and cumulative EC exposure in this more homogeneous subgroup were attenuated, but the risk pattern was generally similar to those observed in the main analyses (Supplementary Table 4.5).

All analyses involving alternative JEM configurations produced results that were more attenuated than results from the main analyses; however elevated lung cancer ORs and exposure-response between EC exposure and lung cancer were observed in all three alternative models (Supplementary Tables 4.6-4.8).

For women with lung cancer subtypes other than adenocarcinoma, we observed elevated OR point estimates for all EC exposure categories compared with unexposed subjects (Supplementary Table 4.9). However the uncertainties around these estimates were large due to limited statistical power. Among women we observed an indication of increasing risk trend (p=0.04) with longer time since last exposure (Supplementary Table 5). No trends were observed in men.

SUPPLEMENTARY REFERENCES

Bakke, B., Ulvestad, B., Thomassen, Y., Woldbæk, T., & Ellingsen, D. G. (2014). Characterization of Occupational Exposure to Air Contaminants in Modern Tunnelling Operations. *The Annals of Occupational Hygiene*, *58*(7), 818–829. https://doi.org/10.1093/annhyg/meu034

Bott. (2010). *FIRE FIGHTER EXPOSURE TO DIESEL EXHAUST AT QFRS FIRE STATIONS (PDF Download Available)*. ResearchGate. http://dx.doi.org/10.13140/RG.2.1.3921.5601

Couch. (2015). *Evaluation of Diesel Exhaust Exposures at Multiple Fire Stations in a City Fire Department*. HHE.

Debia, M., Neesham-Grenon, E., Mudaheranwa, O. C., & Ragettli, M. S. (2016). Diesel exhaust exposures in port workers. *Journal of Occupational and Environmental Hygiene*, *13*(7), 549–557. https://doi.org/10.1080/15459624.2016.1153802

Elihn, K., Ulvestad, B., Hetland, S., Wallén, A., & Randem, B. G. (2008). Exposure to Ultrafine Particles in Asphalt Work. *Journal of Occupational and Environmental Hygiene*, *5*(12), 771–779. https://doi.org/10.1080/15459620802473891

Galea, K. S., Mair, C., Alexander, C., de Vocht, F., & van Tongeren, M. (2016). Occupational Exposure to Respirable Dust, Respirable Crystalline Silica and Diesel Engine Exhaust Emissions in the London Tunnelling Environment. *The Annals of Occupational Hygiene*, *60*(2), 263–269. https://doi.org/10.1093/annhyg/mev067

Health Council of the Netherlands. (2019, March 13). *Diesel Engine Exhaust: Health-based recommended occupational exposure limit*. https://www.gezondheidsraad.nl/binaries/gezondheidsraad/documenten/adviezen/2019/03/13/dieselmotoremissie/Diesel+Engine+Exhaust.pdf

Hewett, P., & Bullock, W. H. (2014). Rating locomotive crew diesel emission exposure profiles using statistics and Bayesian Decision Analysis. *Journal of Occupational and Environmental Hygiene*, *11*(10), 645–657. https://doi.org/10.1080/15459624.2014.899239

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal*, *327*(7414), 557–560.

HSE. (2019). *National Exposure Database*. https://data.gov.uk/dataset/afc1ad32-f7f6-4f54-a8fc-2a281a25b8ad/national-exposure-database

Lan, Q., Vermeulen, R., Dai, Y., Ren, D., Hu, W., Duan, H., Niu, Y., Xu, J., Fu, W., Meliefste, K., Zhou, B., Yang, J., Ye, M., Jia, X., Meng, T., Bin, P., Kim, C., Bassig, B. A., Hosgood, H. D., … Rothman, N. (2015). Occupational exposure to diesel engine exhaust and alterations in lymphocyte subsets. *Occupational and Environmental Medicine*, *72*(5), 354–359. https://doi.org/10.1136/oemed-2014-102556

Lee, K.-H., Jung, H.-J., Park, D.-U., Ryu, S.-H., Kim, B., Ha, K.-C., Kim, S., Yi, G., & Yoon, C. (2015). Occupational Exposure to Diesel Particulate Matter in Municipal Household Waste Workers. *PLOS ONE*, *10*(8), e0135229. https://doi.org/10.1371/journal.pone.0135229

Peters, S., Vermeulen, R., Cassidy, A., Mannetje, A. 't, Tongeren, M. van, Boffetta, P., Straif, K., & Kromhout, H. (2011). Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case–control study. *Occupational and Environmental Medicine*, *68*(2), 148–153. https://doi.org/10.1136/oem.2010.055608

Pronk, A., Coble, J., & Stewart, P. A. (2009). Occupational exposure to diesel engine exhaust: A literature review. *Journal of Exposure Science & Environmental Epidemiology*, *19*(5), 443–457. https://doi.org/10.1038/jes.2009.21

Sheesley, R. J., Schauer, J. J., Garshick, E., Laden, F., Smith, T. J., Blicharz, A. P., & Deminter, J. T. (2008). Tracking personal exposure to particulate diesel exhaust in a diesel freight terminal using organic tracer analysis. *Journal of Exposure Science and Environmental Epidemiology*, *19*(2), 172–186. https://doi.org/10.1038/jes.2008.11

Shih, T.-S., Lai, C.-H., Hung, H.-F., Ku, S.-Y., Tsai, P.-J., Yang, T., Liou, S.-H., Loh, C.-H., & Jaakkola, J. J. K. (2008). Elemental and organic carbon exposure in highway tollbooths: A study of Taiwanese toll station workers. *Science of The Total Environment*, *402*(2), 163–170. https://doi.org/10.1016/j.scitotenv.2008.04.051

Verma, D. K., Finkelstein, M. M., Kurtz, L., Smolynec, K., & Eyre, S. (2003). Diesel exhaust exposure in the Canadian railroad work environment. *Applied Occupational and Environmental Hygiene*, *18*(1), 25–34. https://doi.org/10.1080/10473220301386

Vermeulen, R., Coble, J. B., Yereb, D., Lubin, J. H., Blair, A., Portengen, L., Stewart, P. A., Attfield, M., & Silverman, D. T. (2010). The Diesel Exhaust in Miners Study: III. Interrelations between Respirable Elemental Carbon and Gaseous and Particulate Components of Diesel Exhaust derived from Area Sampling in Underground Non-metal Mining Facilities. *The Annals of Occupational Hygiene*, *54*(7), 762–773. https://doi.org/10.1093/annhyg/meq023

Supplementary Table 1: Description of the studies included in these analyses in the SYNERGY project

		Cases			Controls				
Study	**Country**	**Data collection**	**N**	**Response rate (%)**	**N**	**Response rate (%)**	**EC exposure**	**Control source**[b]	**Interview**[c]
AUT-Munich	Germany	90–95	3180	77	3249	41	31-95	P	S
CAPUA	Spain	00–10	559	91	512	96	26-10	H	S
EAGLE	Italy	02–05	1908	87	2065	72	32-05	P	S
HdA	Germany	88–93	1004	69	1002	68	26-93	P	S
ICARE	France	01–07	2739	63	3449	77	37-07	P	S & NOK
INCO	Czech Republic	99–02	304	94	452	80	37-02	H	S
INCO	Hungary	98–01	391	90	305	100	31-99	H	S
INCO	Poland	98–02	793	88	835	88	33-01	P & H	S
INCO	Romania	98–02	179	90	225	99	43-01	H	S
INCO	Russia	98–01	599	96	580	90	38-00	H	S
INCO	Slovakia	98–02	345	90	285	84	37-02	H	S
INCO/LLP	United Kingdom	98–05	441	78	916	84	34-04	P	S
LUCA	France	89–92	280	98	282	98	27-92	H	S
LUCAS	Sweden	85–90	1014	87	2307	85	23-90	P	S & NOK
MONTREAL	Canada	96–02	1176	85	1505	69	36-99	P	S & NOK
MORGEN[a]	Netherlands	93–97	43	N/A	115	N/A	45-94	P	S
PARIS	France	88–92	169	95	227	95	29-92	H	S
ROME	Italy	93–96	326	74	321	63	26-95	H	S
TORONTO	Canada	97–02	365	62	844	71	29-02	P & H	S
TURIN/VENETO	Italy	90–94	1086	79	1489	80	25-94	P	S
Overall	14 countries	85–10	16901	78%	20965	69%	23-10	P=79%	S=92.7 %

a Nested case-control study: 45% of invited participants to the original cohort completed the baseline questionnaire.
b P = population controls; H = hospital controls
c S = subject; NOK = Next-of-kin

Supplementary Table 2 Sensitivity analyses in men for the association between cumulative exposure to elemental carbon (EC) in decile groups and lung cancer

Cumulative EC exposure µg/m³-years	All studies		
	Cases/controls	OR	95%CI
Unexposed	5560/8270	1.0	Referent
>0–6.5	663/819	1.04	0.92–1.18
6.6–16	685/788	1.10	0.97–1.24
17–29	638/786	1.04	0.92–1.18
30–47	759/786	1.18	1.04–1.33
48–71	797/828	1.12	1.00–1.26
72–104	907/847	1.30	1.16–1.46
105–148	800/813	1.20	1.07–1.36
149–218	875/822	1.33	1.18–1.49
219–322	925/834	1.41	1.25–1.58
>322	996/858	1.42	1.27–1.59
Test for trend, p-value§		<0.01	
Excl. never exposed		<0.01	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

Supplementary Table 3: Odds ratios (OR) and 95% confidence intervals (CIs) of lung cancer for cumulative elemental carbon (EC) for male subjects with different smoking habits (packyears).

Smoking and EC exposure	Controls	Cases	OR	95% CI	
Never smokers					
Unexposed	2525	256	1.0	ref	
>0–22 µg/m³-years	453	66	1.33	0.99	1.77
23–70 µg/m³-years	445	41	0.88	0.61	1.23
71–178 µg/m³-years	485	55	1.10	0.80	1.48
179–218 µg/m³-years	95	17	1.60	0.91	2.66
219–261 µg/m³-years	112	11	0.89	0.45	1.61
262–322 µg/m³-years	115	10	0.79	0.39	1.47
>322 µg/m³-years	207	34	1.50	1.00	2.18
<10 pack-years					
Unexposed	1226	244	1.91	1.58	2.31
>0–22 µg/m³-years	336	81	2.12	1.60	2.79
22–70.6 µg/m³-years	302	76	2.28	1.70	3.02
70.6–178 µg/m³-years	305	89	2.59	1.97	3.39
178–218 µg/m³-years	69	18	2.28	1.29	3.82
218–261 µg/m³-years	54	30	5.07	3.14	8.05
261–322 µg/m³-years	61	18	2.60	1.47	4.40
>322 µg/m³-years	113	52	4.21	2.93	5.99
10-19 pack-years					
Unexposed	1245	494	3.78	3.20	4.47
>0–22 µg/m³-years	322	153	4.30	3.39	5.43
22–70.6 µg/m³-years	287	168	5.25	4.16	6.62
70.6–178 µg/m³-years	278	197	6.51	5.19	8.17
178–218 µg/m³-years	59	45	6.76	4.46	10.2
218–261 µg/m³-years	53	44	7.32	4.77	11.2
261–322 µg/m³-years	65	38	5.14	3.34	7.81
>322 µg/m³-years	114	76	6.21	4.49	8.54
20-39 pack-years					
Unexposed	1934	1904	13.7	12.0	15.8
>0–22 µg/m³-years	509	603	14.6	12.5	17.1
22–70.6 µg/m³-years	543	654	15.2	13.1	17.8
70.6–178 µg/m³-years	572	749	16.6	14.3	19.4
178–218 µg/m³-years	107	170	19.9	16.0	24.8
218–261 µg/m³-years	103	159	18.8	15.1	23.4
261–322 µg/m³-years	95	151	19.8	15.9	24.8
>322 µg/m³-years	236	316	18.5	15.5	22.1
>39 pack-years					
Unexposed	1340	2662	21.1	18.2	24.5
>0–22 µg/m³-years	382	781	20.5	17.1	24.6
22–70.6 µg/m³-years	428	919	21.8	18.3	26.1
70.6–178 µg/m³-years	434	1023	24.0	20.2	28.6
178–218 µg/m³-years	78	219	30.0	22.5	40.4
218–261 µg/m³-years	92	231	25.4	19.3	33.7
261–322 µg/m³-years	84	233	28.0	21.1	37.3
>322 µg/m³-years	188	518	28.1	22.7	34.9

OR is adjusted for study, age group, and List A job

Supplementary Table 4: Sensitivity analyses for the association between cumulative exposure to elemental carbon (EC) and lung cancer

4.1 Analyses in men by type of controls

Cumulative EC exposure µg/m³-years	Studies with population controls*			Studies with hospital controls*		
	Cases/controls	OR	95%CI	Cases/controls	OR	95%CI
Unexposed	4001/6440	1.0	Referent	1571/1901	1.0	Referent
>0–22	1302/1615	1.10	1.00–1.21	317/286	1.22	1.01–1.48
23–70	1368/1512	1.14	1.03–1.25	427/397	1.06	0.89–1.25
71–178	1535/1516	1.30	1.18–1.43	513/432	1.17	1.00–1.38
>178	1723/1495	1.56	1.42–1.71	643/544	1.20	1.03–1.40
Test for trend, p-value§		<0.01			0.89	
Excl. never exposed		<0.01			0.37	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs
*Subjects from the INCO Poland and Toronto studies were included in both analyses, since both types of controls were used

4.2 Analyses restricted to male blue-collar workers and workers employed after 1960

Cumulative EC exposure µg/m³-years	Restricted to blue-collar workers			Restricted to workers employed after 1960		
	Cases/controls	OR	95%CI	Cases/controls	OR	95%CI
Unexposed	3559/4828	1.0	Referent	1682/2825	1.0	Referent
>0–22	1424/1567	1.01	0.92–1.11	342/460	0.98	0.82–1.17
23–70	1724/1762	1.00	0.92–1.10	443/466	1.25	1.05–1.48
71–178	1948/1830	1.10	1.01–1.21	403/429	1.17	0.98–1.39
>178	2242/1906	1.27	1.16–1.38	329/351	1.24	1.02–1.50
Test for trend, p-value§		<0.01			0.18	
Excl. never exposed		0.05			0.60	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

4.3 Analyses restricted to male non-agricultural and non-mining workers

Cumulative EC exposure µg/m³-years	Subjects never employed in the mining industry			Subjects never employed in the agriculture industry		
	Cases/controls	OR	95%CI	Cases/controls	OR	95%CI
Unexposed	5534/8220	1.0	Referent	5414/8044	1.0	Referent
>0–22	1620/1943	1.08	0.99–1.18	1406/1666	1.09	1.00–1.20
23–70	1729/1919	1.09	1.00–1.18	1434/1496	1.13	1.03–1.24
71–178	1928/1957	1.22	1.12–1.32	1467/1387	1.28	1.17–1.41
>178	2029/1881	1.38	1.27–1.50	1655/1405	1.43	1.31–1.57
Test for trend, p-value§		<0.01			<0.01	
Excl. never exposed		<0.01			<0.01	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

4.4 Analyses in men with alternative model adjustments for exposure to other occupational lung carcinogens

Cumulative EC exposure µg/m³-years	No co-exposure adjustment*			Alternative adjustment†		
	Cases/controls	OR	95%CI	Cases/controls	OR	95%CI
Unexposed	5560/8270	1.0	Referent	5560/8270	1.0	Referent
>0–22	1684/2002	1.11	1.02–1.21	1684/2002	1.00	0.91–1.09
23–70	1858/2005	1.13	1.04–1.22	1858/2005	0.99	0.90–1.08
71–178	2113/2074	1.25	1.16–1.36	2113/2074	1.09	0.99–1.19
>178	2390/2100	1.43	1.32–1.55	2390/2100	1.23	1.13–1.35
Test for trend, p-value§		<0.01			0.03	
Excl. never exposed		<0.01			0.06	

* OR is adjusted for study, age group, and smoking (pack-years, time-since-quitting smoking).
† (OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and ever exposure to silica, asbestos, polycyclic aromatic hydrocarbons, and chromium.

4.5 Analyses in men by homogenous group of studies (excluding AUT and PARIS)

Cumulative EC exposure µg/m³-years	All studies except AUT and PARIS		
	Cases/controls	OR	95%CI
Unexposed	4745/7056	1.0	Referent
>0–22	1235/1481	1.10	0.99–1.21
23–70	1406/1563	1.09	1.00–1.20
71–178	1606/1687	1.19	1.09–1.30
>178	1793/1746	1.29	1.18–1.40
Test for trend, p-value§		<0.01	
Excl. never exposed		0.20	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

4.6 Analyses in men with alternative JEM restricted to respirable EC data

Cumulative EC exposure µg/m³-years	All studies		
	Cases/controls	OR	95%CI
Unexposed	5560/8270	1.0	Referent
>0–24	1621/1971	1.07	0.98–1.17
25–73	1841/2024	1.09	1.01–1.19
74–193	2207/2071	1.30	1.20–1.41
>193	2376/2115	1.39	1.29–1.51
Test for trend, p-value§		<0.01	
Excl. never exposed		<0.01	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

4.7 Analyses in men with alternative JEM without any expert-assigned exposure probabilities

Cumulative EC exposure µg/m³-years	All studies		
	Cases/controls	OR	95%CI
Unexposed	5560/8270	1.0	Referent
>0–48	1735/2002	1.08	1.00–1.18
49–143	1982/2035	1.17	1.08–1.27
144–338	2093/2070	1.21	1.11–1.31
>338	2235/2074	1.42	1.31–1.53
Test for trend, p-value§		<0.01	
Excl. never exposed		<0.01	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

4.8 Analyses in men with alternative JEM restricted to jobs where expert-assigned exposure probabilities=1

Cumulative EC exposure µg/m³-years	All studies		
	Cases/controls	OR	95%CI
Unexposed	9020/12051	1.0	Referent
>0–25	900/1068	0.96	0.86–1.06
26–72	1186/1096	1.12	1.02–1.24
73–209	1223/1115	1.20	1.08–1.32
>209	1276/1121	1.27	1.16–1.40
Test for trend, p-value§		<0.01	
Excl. never exposed		0.02	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

4.9 Analyses in women with subtypes other than adenocarcinoma

Cumulative EC exposure µg/m³-years	Female subjects with cancer subtypes other than adenocarcinoma		
	Cases/controls	OR	95%CI
Unexposed	1624/4002	1.0	Referent
>0–22	110/179	1.18	0.87–1.59
23–70	74/162	1.18	0.84–1.65
71–178	38/99	1.06	0.67–1.63
>178	23/72	1.18	0.67–2.02
Test for trend, p-value§		0.33	
Excl. never exposed		0.74	

OR is adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

Supplementary Table 5: Lung cancer odds ratios (OR) in both sexes associated with time since last occupational elemental carbon (EC) exposure

Occupational EC exposure	Exposure category	Cases (%)	Controls (%)	OR*	95% CI
Men					
Time since last exposure					
(years) †	Never	5560 (40.9)	8270 (50.3)	1.0	Referent
	10–19	5007 (37.3)	4992 (30.3)	1.04	0.93–1.15
	20–29	1280 (9.4)	1367 (8.3)	1.04	0.94–1.16
	30–39	979 (7.2)	1005 (6.1)	1.17	1.04–1.31
	>39	709 (5.2)	817 (5.0)	1.00	0.88–1.13
Test for trend, p-value (excl. never exposed)				0.71	
Women					
Time since last exposure					
(years) †	Never	2910 (88.3)	4002 (88.7)	1.0	Referent
	10–19	144 (4.4)	210 (4.7)	0.81	0.56–1.17
	20–29	66 (2.0)	91 (2.0)	0.82	0.54–2.22
	30–39	81 (2.5)	78 (1.7)	1.44	0.99–1.08
	>39	95 (2.9)	133 (2.9)	0.92	0.49–1.27
Test for trend, p-value (excl. never exposed)				0.04	

*OR adjusted for study, age group, smoking (pack-years, time-since-quitting smoking), and List A jobs

†OR in "time since last exposure" is additionally adjusted for duration (continuous) of silica exposure. Trend test limited to exposed subjects.

Supplementary Table 6* – Occupational elemental carbon (EC) exposure measurements used in the diesel engine exhaust job-exposure matrix (DEE-JEM)

Description	Agent†	Duration	N	AM (SD)	Location	Year	Source
Drivers							
Drivers - bus	ECI	>4	4	2>LOD:11-20	US	1998	Pronk et al (2009)
Drivers - bus	ECR	>4	5	10	Estonia	2002	Pronk et al (2009)
Drivers - bus	ECR	>4	39	2 (1.3)	US	2002	Pronk et al (2009)
Drivers - bus and truck	ECI	>4	20	11	Sweden	2002-2004	Pronk et al (2009)
Drivers - local truck	ECNI	>4	4	5 (0.1)	US	1985	Pronk et al (2009)
Drivers - local truck	ECR	>4	5	7	US	1999	Pronk et al (2009)
Drivers - local truck	ECS	>4	56	5 (0.9)	US	1980s	Pronk et al (2009)
Drivers - local truck	ECS	>4	576	2 (2.3)	US	2001-2005	Pronk et al (2009)
Drivers - locomotive	ECNI	>4	76	5 (1.1-15.8)	Canada	1999-2000	Pronk et al (2009)
Driver - locomotive	ECR	>4	156	2.8	US	1996-2007	Hewett and Bullock (2014)
Drivers - locomotive	ECNI	>1	8	11.4 (5.5)	UK	1996	NEDB
Drivers - locomotive in tunnel construction	ECR	>4	2	24 (12)	UK	2014	Galea et al. (2015)
Drivers - locomotive in tunnel construction	ECR	>4	2	21 (4)	UK	2014	Galea et al. (2015)
Drivers - locomotive shunter	ECR	>4	19	20 (18.7)	Russia	2002	Pronk et al (2009)
Drivers - long haul truck	ECS	>4	21	1.55 (0.42)	US	2006?	Sheesley et al (2008)
Drivers - long haul truck	ECNI	>4	4	22 (13.2)	US	1985	Pronk et al (2009)
Drivers - long haul truck	ECR	>4	5	5	US	1999	Pronk et al (2009)
Drivers - long haul truck	ECS	>4	72	5 (0.4)	US	1980s	Pronk et al (2009)
Drivers - long haul truck	ECS	>4	349	1 (0.8)	US	2001-2005	Pronk et al (2009)
Drivers - taxi	ECI	>4	8	8	Sweden	2002-2004	Pronk et al (2009)
Drivers - truck	ECS	>4	18	2.71 (1.37)	US	2006?	Sheesley et al (2008)
Drivers - truck	ECI	1->4	3	10 (6)	US	1992	Pronk et al (2009)
Mechanics							
Mechanics - ambulance	ECR	>4	3	31	UK	2000	Pronk et al (2009)
Mechanics - bus	ECI	>4	4	ND	US	1998	Pronk et al (2009)
Mechanics - bus	ECR	>4	53	39	UK	2000	Pronk et al (2009)
Mechanics - bus	ECR	>4	15	39	Estonia	2002	Pronk et al (2009)
Mechanics - diesel engine testing	ECR	>4	54	48.5 (22.1)	US	2012-2013	Lan et al. (2015)
Mechanics - locomotive	ECNI	>1	22	31.4 (40.7)	UK	1994-2006	NEDB
Mechanics - motor vehicle	ECNI	>1	20	39.1 (45.1)	UK	1994-2006	NEDB
Mechanics - truck	ECS	>4	7	2.04 (1.02)	US	2006?	Sheesley et al (2008)
Mechanics - truck	ECR	>4	10	4	US	1999	Pronk et al (2009)
Mechanics - truck	ECS	>4	80	27 (4.1)	US	1980s	Pronk et al (2009)
Mechanics - truck and bus	ECI	>4	40	21	Sweden	2002-2004	Pronk et al (2009)
Mechanics - tunnel construction	ECR	>4	29	48.39	Norway	2010-2011	Bakke et al. (2014)
Firefighters							
Firefighters	ECI	>4	18	40 (20.3)	US	1995	Pronk et al (2009)
Firefighters	ECI	>4	12	10 (max)	US	1997	Pronk et al (2009)
Firefighters	ECI	<1	8	ND	US	1998	Pronk et al (2009)
Firefighters	ECI	>4	27	24 (max)	US	2002	Pronk et al (2009)
Firefighters	ECR	10	21	2	AUS	2010	Bott et al (2010)
Firefighters	ECR	>7	28	1.34 (0.56)	US	2016	Couch et al (2016)
Dockworkers							
Dockworker	ECS	>4	≥5		US	1990	Pronk et al (2009)
Dockworker	ECS	>4	54	24 (0.4-2.5)	US	1991	Pronk et al (2009)
Dockworker	ECI	>4	5	4 (1.8)	US	1992	Pronk et al (2009)
Dockworker	ECR	>4	12	9	US	1999	Pronk et al (2009)
Dockworker	ECR	>4	27	122	UK	2000	Pronk et al (2009)
Dockworker	ECS	>4	14	1.12 (0.41)	US	2006?	Sheesley et al (2008)
Dockworkers - ship loading	ECR	>4	20	49	UK	2000	Pronk et al (2009)
Dockworkers - ship loading	ECI	>4	168	6 (0.9-9.0)	US	2003-2005	Pronk et al (2009)

Miners							
Surface production	ECS	>4	23	23 (15-54)	US	1997	Pronk et al (2009)
Surface production	ECR	>4	164	13 (2-89)	US	2002	Pronk et al (2009)
Underground maintenance	ECS	>4	8	53 (46)	US	1997	Pronk et al (2009)
Underground maintenance	ECR	>4	269	144 (17-462)	US	2002	Pronk et al (2009)
Underground not specified	ECR	NI	7	66 (28)	UK	2004	Pronk et al (2009)
Underground not specified	ECNI	NI	27	27	Sweden	2006	Pronk et al (2009)
Underground production	ECR	>4	4	241	Estonia	2002	Pronk et al (2009)
Underground production	ECR	>4	15	637 (75-508)	US	1999	Pronk et al (2009)
Underground production	ECI	<1-4	12	538 (512)	US	2007	Pronk et al (2009)
Underground production	ECS	>4	38	219 (65-193)	US	1997	Pronk et al (2009)
Underground production	ECR	>4	343	202 (32-144)	US	2002	Pronk et al (2009)
Underground production	ECR	NI	6	148 (136)	UK	2004	Pronk et al (2009)
Underground production	ECS	1->4	13	163 (141)	US	2001-2002	Pronk et al (2009)
Others							
Asphalt road pavers	ECR	>4	3	3 (0.2)	Sweden	2005-2006	Elihn et al (2008)
Baggage handling	ECI	>4	72	11 (5.4)	US	2004	Pronk et al (2009)
Bus service workers	ECI	>4	4	2>LOD:15-30	US	1998	Pronk et al (2009)
Cleaners - general	ECNI	>1	4	22.4 (9.4)	UK	1994-2001	NEDB
Cleaners - locomotive	ECNI	>1	5	39.0 (9.4)	UK	1996-2006	NEDB
Concrete ring builders	ECR	>4	3	20 (8)	UK	2014	Galea et al. (2015)
Concrete ring segment lifters	ECR	>4	8	17 (8)	UK	2014	Galea et al. (2015)
Concrete sprayers	ECR	>4	7	57.41	Norway	2010-2011	Bakke et al. (2014)
Conductors - locomotive	ECNI	>1	14	21.9 (36.8)	UK	1994-2001	NEDB
Construction engineers	ECR	>4	6	20 (7)	UK	2014	Galea et al. (2015)
Construction workers	ECI	>4	22	13	Sweden	2002-2004	Pronk et al (2009)
Conveyor extension workers	ECR	>4	2	30 (3)	UK	2014	Galea et al. (2015)
Curbside waste collectors	ECR	>4	72	5.53	S. Korea	2014	Lee et al. (2015)
Customer service workers - bus station	ECNI	>1	1	28.4	UK	2006	NEDB

Customer service workers - locomotive platform	ECNI	>1	14	13.2 (6.9)	UK	1996-2001	NEDB
Drill and blast workers	ECR	>4	51	47.2	Norway	2010-2011	Bakke et al. (2014)
Electric utility installers	ECI	>4	120	4	US	1996-1997	Pronk et al (2009)
Electricians - railway station	ECNI	>1	7	39.5 (25.4)	UK	1994-2006	NEDB
Engineers - locomotive	ECI	1->4	49	6	US	1996-1998	Pronk et al (2009)
Forklift operators	ECR	>4	39	36	UK	2004	Pronk et al (2009)
Forklift operators - warehouse	ECNI	>1	25	82.2 (130.8)	UK	1994-2006	NEDB
Gate controllers in booth	ECR	>4	29	1.8	Canada	2013	Debia et al. (2016)
Grout pump operators	ECR	>4	1	110	UK	2014	Galea et al. (2015)
Grout pump operators	ECR	>4	5	19 (6)	UK	2014	Galea et al. (2015)
Heavy/highway construction workers	ECR	>4	261	13	US	1994-1999	Pronk et al (2009)
Highway toll booth workers	ECR	>4	63	6.1 (4)	China	2002	Shih et al (2008)
Hostlers - locomotive	ECNI	>4	5	4 (1.3)	Canada	1999-2000	Pronk et al (2009)
Hostlers - tractor	ECS	>4	4	1.3 (1)	US	2006?	Sheesley et al (2008)
Labourers - various industries	ECNI	>1	20	60.9 (63.7)	UK	1994-2006	NEDB
Lead miners	ECR	>4	1	12	UK	2014	Galea et al. (2015)
Loading and unloading workers - passenger ferries	ECNI	>1	10	46.2 (28.2)	UK	1994-2006	NEDB
Locomotive rail extension workers	ECR	>4	2	26 (6)	UK	2014	Galea et al. (2015)
Maintenance workers - locomotive	ECR	>4	64	39	UK	2000	Pronk et al (2009)
Maintenance workers - locomotive	ECNI	>4	48	5 (4.9-8.8)	Canada	1999-2000	Pronk et al (2009)
Material loaders	ECR	>4	18	38.43	Norway	2010-2011	Bakke et al. (2014)
Non-operating crew in trailing locomotive	ECI	>4	47	10 (12)	Canada	2003	Pronk et al (2009)
Parking booth attendants	ECR	>4	34	1.1 (0.6)	US	2002	Pronk et al (2009)
Pipe/walkway extension workers	ECR	>4	2	18 (2)	UK	2014	Galea et al. (2015)
Porters - warehouse	ECNI	>1	3	30.4 (35.1)	UK	1994-2006	NEDB
Security guards	ECNI	>1	2	60.9 (2.1)	UK	1994	NEDB
Support workers	ECR	>4	31	76.14	Norway	2010-2011	Bakke et al. (2014)

Toll attendants/cashiers	ECNI	>1	22	30.5 (22.3)	UK	1994-2006	NEDB
Tunnel boring machine operators	ECR	>4	5	17 (5)	UK	2014	Galea et al. (2015)
Tunnel construction workers	ECI	>4	10	314	Norway	1996-1999	Pronk et al (2009)
Tunnel construction workers	ECI	>4	12	132	Sweden	2002-2004	Pronk et al (2009)
Unloading equipment operator - shipyard	ECNI	>1	10	37.6 (24.9)	UK	2006	NEDB
Vehicle inspectors	ECNI	>1	2	21.8 (5.1)	UK	1999	NEDB
Vehicle testing workers	ECR	>4	11	11	UK	2000	Pronk et al (2009)
Waterproofing workers	ECR	>4	13	64.82	Norway	2010-2011	Bakke et al. (2014)
Workers in trailing locomotive	ECR	>4	22	11.1	US	1996-2007	Hewett and Bullock (2014)

*Table partially adapted from Tables 1-4 in review by Pronk and colleagues (Pronk et al., 2009).
†ECR: respirable elemental carbon; ECS: submicron elemental carbon; ECI: inhalable elemental carbon; ECNI: elemental carbon size fraction not indicated.

Supplementary Table 7: Job codes and descriptions for List A jobs included in study population

ISCO-68*	Job description
03810	mining technicians (general)
55130	janitor
55220	charworker
55290	other charworkers, cleaners and related workers
58110	fire fighter (general)
58190	other fire-fighters
62330	vineyard worker
70000	production supervisors and general foremen
70010	production supervisors and general foremen (general)
70020	supervisors and general foreman, mining, quarrying and well drilling
70030	supervisors and general foreman, metal processing
70050	supervisor and general foreman, manufacturing of machinery and metal products
70055	supervisor and general foreman, manufacturing and installation of electrical and electronic equipment
70075	supervisor and general foreman, construction work
70080	supervisor and general foreman, production and distribution of electricity, gas and water
70090	other production supervisors and general foremen
71100	miners and quarrymen
71105	miner (general)
71110	quarryman (general)
71120	cutting-machine operator (mine)
71130	drilling-machine operator (mine and quarry)
71150	shot-firer (mine and quarry)
71160	underground timberman
71170	sampler (mine)
71190	other miners and quarrymen
71250	jig tender
71290	other mineral and stone treaters
72000	metal processers
72100	metal smelting, converting and refining furnacemen
72120	blast furnaceman (ore smelting)
72130	open-hearth furnacemen (steel)
72170	furnaceman (non-ferrous metal converting and refining)
72190	other metal smelting, converting and refining furnacemen
72220	hot-roller (steel)
72230	continuous-mill-roller (steel)

72250	roller (non-ferrous metals)
72260	seamless pipe and tube roller
72290	other metal rolling-mill workers
72320	furnaceman (metal melting, except cupola)
72330	cupola furnaceman
72340	furnaceman (metal reheating)
72390	other metal melters and reheaters
72400	metal casters
72420	metal pourer
72440	die-casting-machine operator
72450	continuous rod-casting-machine operator (non-ferrous metal)
72490	other metal casters
72500	metal moulders and coremakers
72520	bench moulder (metal)
72530	floor and pit moulder
72540	moulder (machine)
72550	coremaker (hand)
72560	coremaker (machine)
72590	other metal moulders and coremakers
72620	annealer
72630	hardener
72700	metal drawers and extruders
72730	wire drawer (machine)
72740	seamless pipe and tube drawer
72750	extruder operator (metal)
72800	metal platers and coaters
72820	electroplater
72830	hot-dip plater
72840	wire-coating-machine operator
72890	other metal platers and coaters
72900	metal processers not elsewhere classified
72920	metal bluer
72930	casting finisher
72940	metal cleaner
72990	other metal processors
73210	sawmill sawyer (general)
74140	mixing-and blending-machine operator (chemical and related processes)
74230	roaster (chemical and related processes)

74290	other cookers, roasters and related heat-treaters
74390	other filter and separator operators
74920	coke burner
74925	coal-glass maker
74990	other chemical processors and related workers
75290	other spinners and winders
79630	vehicle upholsterer
79690	other upholsterers and related workers
81120	cabinetmaker
81230	wood turner
81920	coach-body builder
81925	cartwright
81935	wooden pattern maker
81955	wooden furniture finisher
81960	smoking-pipe maker
83000	blacksmiths, toolmakers and machine-tool operators
83100	blacksmiths, hammersmiths and forging-press operators
83110	blacksmith (general)
83120	hammersmith
83130	drop-hammer operator
83140	forging-press operator
83190	other blacksmiths, hammersmiths and forging-press operators
83220	tool and die maker
83240	metal pattern maker (foundry)
83250	metal marker
83290	other toolmakers, metal pattern makers and metal markers
83305	metalworking-machine setter (general)
83310	metalworking-machine setter-operator (general)
83320	lathe setter-operator
83330	milling-machine setter-operator
83350	boring-machine setter-operator
83370	precision-grinding-machine setter-operator
83390	other machine-tool setter-operators
83410	machine-tool operator (general)
83420	lathe operator
83430	milling-machine-operator
83440	planing-machine-operator
83450	boring-machine-operator

83460	drilling-machine-operator
83465	precision-grinding-machine-operator
83490	other machine-tool operators
83520	buffing- and polishing machine operator
83530	tool grinder, machine tools
83590	other metal grinders, polishers and tool sharpeners
83930	locksmith
83940	metal spinner
83950	metal former (hand)
83960	metal-press operator
83980	power-shear operator
83990	other blacksmiths, toolmakers and machine-tool operators
84100	machinery fitters and machine assemblers
84105	machinery fitter (general)
84110	machinery fitter-assembler (general)
84115	internal combustion engine fitter-assembler (except aircraft and marine engines)
84125	marine engine machinery fitter-assembler
84130	turbine fitter-assembler (except aircraft and marine engines)
84135	metalworking machine-tool fitter-assembler
84175	machinery erector and installer
84180	refrigeration and air-conditioning plan installer and mechanic
84190	other machinery fitters and machine assemblers
84230	precision -instrument maker and repairer
84320	automobile mechanic
84330	motor-truck mechanic
84390	other motor-vehicle mechanics
84410	aircraft engine mechanic (general)
84490	other aircraft engine mechanics
84900	machinery fitters, machine assemblers and precision instrument makers (except electrical) not elsewhere classified
84910	machinery mechanic (general)
84915	reciprocating steam-engine mechanic
84920	diesel engine mechanic (except motor vehicle)
84930	metalworking machine-tool mechanic
84970	plant maintenance mechanic
84980	oiler and greaser (except ships' engines)
84985	mechanical products inspector and tester
84990	other machinery fitters, machine assemblers and precision instrument makers (except electrical)

85110	electrical fitter (general)
85120	electrical motor and generator fitter
85130	electrical transformer fitter
85140	electrical switchgear and control apparatus fitter
85150	electrical instrument fitter
85190	other electrical fitters
85210	electronics fitter (general)
85250	electronics fitter (industrial equipment)
85260	electronic signalling systems fitter
85320	electrical equipment assembler
85340	coil winder (machine)
85420	radio and television mechanic
85510	electrician (general)
85520	building electrician
85530	aircraft electrician
85535	ship's electrician
85540	vehicle electrician
85560	maintenance electrician
85570	electrical repairman
85630	telephone and telegraph mechanic
85740	telephone and telegraph lineman
85920	electrical and electronic products inspector and tester
87000	plumbers, welders, sheet metal and structural metal preparers and erectors
87105	plumber (general)
87110	pipe fitter (general)
87130	marine pipe fitter
87190	other plumbers and pipe fitters
87200	welders and flame-cutters
87210	gas and electric welder (general)
87215	gas welder
87220	electric arc welder (hand)
87225	electric arc welder (machine)
87235	resistance welder
87245	brazer
87250	flame-cutter (hand)
87255	flame-cutter (machine)
87290	other welders and flame-cutters
87300	sheet-metal workers

87310	sheet-metal worker (general)
87320	sheet-metal marker
87330	coppersmith
87340	tinsmith
87350	boilersmith
87370	vehicle sheet-metal worker
87390	other sheet-metal workers
87420	structural metal maker
87430	structural steel worker (workshop)
87440	constructional steel erector
87450	metal shipwright
87455	ship plater
87460	hand riveter
87470	pneumatic riveter
87490	other structural metal preparers and erectors
90180	plastics products fabricator
90190	other rubber and plastic product makers (except tire makers and tire vulcanisers)
92200	printing pressmen
92700	photographic darkroom workers
93100	painters, construction
93120	building painter
93130	structural steel and ship painter
93190	other painters, construction
93900	painters not elsewhere classified
93920	brush painter (except construction)
93930	spray painter (except construction)
93940	hand dipper
93950	sign painter
93960	automobile painter
93990	other painters
94980	quality inspector
94990	other production and related workers nec
95120	bricklayer (construction)
95130	firebrick layer
95320	slate and tile roofer
95330	composition roofer
95340	asphalt roofer
95390	other roofers

95400	carpenters, joiners and parquetry workers
95410	carpenter (general)
95415	construction carpenter
95420	construction joiner
95440	wood shipwright
95445	ship joiner
95450	wooden boatbuilder
95455	ship's carpenter
95470	bench carpenter
95475	parquetry worker
95490	other carpenters, joiners and parquetry workers
95620	building insulator (hand)
95650	boiler and pipe insulator
95790	other glaziers
95900	construction workers not elsewhere classified
95910	housebuilder (general)
95940	scaffolder
95975	building exterior cleaner
95990	other construction workers
96190	other power-generating machinery operators
96910	stationary engine operator
96930	boiler fireman
96940	pumping-station operator
96980	heating and ventilation equipment operator
97100	dockers and freight handlers
97130	railway and road vehicle loader
97145	warehouse porter
97155	machine packer
97205	hoisting equipment rigger (general)
97230	ship rigger
97300	crane and hoist operators
97320	bridge- or gantry-crane operator
97327	tower-crane operator
97330	mobile-crane operator
97335	hoist operator (construction)
97345	mine cageman
97350	winch operator
97390	other crane and hoist operators

97445	road-grader and scraper operator
97450	road-roller operator
97460	tar-spreading machine operator
97920	lifting-truck operator
97990	other material handling equipment operators
98140	ordinary seaman
98190	other ships' deck ratings, barge crews and boatman
98290	other ships' engine-room ratings
98320	railway engine driver
98330	railway steam-engine fireman
98420	railway brakeman (freight train)
98550	lorry and van driver (local transport)
98560	lorry and van driver (long-distance transport)
98590	other motor-vehicle drivers
99910	labourer

*ISCO-68: five-digit International Standard Classification of Occupations (version 1968); dashes and dots omitted.

CHAPTER 8

General Discussion

During one of my presentations to fellow students I was asked a particularly pointed question: if diesel engine exhaust is already classified as a human carcinogen, why is this study important? At the time my response highlighted the importance of understanding the full range of exposure-response relationships. While truth-seeking and satisfying our innate curiosity about how things work are valid reasons to pursue advances in all scientific specialties, I believe that many occupational epidemiologists also strive to promote some real change to improve the health of the population through our work. If I were to answer the question again today, I would probably add the fact that the more evidence we have regarding diesel engine exhaust carcinogenicity, the better chance we have in making a real improvement in reducing the exposure in the workplace and beyond.

But what brings this real improvement and what can we collectively do to get more of it? The answer to the first half is relative straightforward in our society – legislations, when properly designed and enforced, reduce exposure by mandating an exposure limit or outright ban. We as occupational epidemiologists generate scientific knowledge, which underpin work exposure-related legislations. In order to create a good product, sufficient for informing well-designed legislations (and invalidating industry lobbyists [see Box 1]), our collective "product development" process involves multiple activities, which for the sake of discussion I have divided into two key areas:

i) Exposure assessment
ii) Epidemiological analysis

A good scientific product of course does not immediately and solely cause societal change. A topic with great consensus in the scientific community may remain controversial in economic, political, and ethical discussions. However, since any reasonable advances in occupational health must (hopefully) begin with scientific information as a foundation, I think it would be safe to assume that all occupational epidemiologists agree that improvements in the aforementioned key departments represent important ways forward both as advances in scientific knowledge and as catalyst for societal change in occupational and general health.

In this chapter, I will revisit the previous chapters to see how they fit within the larger context of improving our work in the specified key areas. Naturally, the work described in these chapters represents only a small incremental progress and the discussion will look beyond our current studies in an attempt to explore potential ways forward.

Box 1 Conflicting interests in occupational health research

In many areas, academia and industry have a synergistic relationship - scientific knowledge generated by the former are used and improved by the latter to create services and products that better human lives. Material science knowledge is woven into waterproof apparels; microbiological and biochemical insights become medications and vaccines; etc.

In occupational health, this synergism similarly promoted discoveries that are beneficial to scientists, business owners, and workers. All stakeholders are usually keen to cooperate on acute worker health issues leading to increased rates of absenteeism, injury, illness, and death at the workplace. This is because, in addition to causing pain and suffering among workers, these issues have direct negative impacts on business productivity and profit. One presently ongoing example of such opportunities for cooperation is COVID-19 transmission in slaughterhouses during the current pandemic. Preventing disease outbreaks in slaughterhouses benefits all directly involved parties: the scientists who want to understand SARS-CoV-2 transmission, the businesses that want to continue production, and the workers who want safe working conditions.

The situation becomes more complicated for chronic occupational health issues, for which causes are often multifactorial and onset usually occur years after exposure. Although occupational diseases one or two decades later certainly still affect the employees, they no longer necessarily decrease the employers' business profit. In fact, actions to promote long-term worker health and wellness that do not result in increased productivity are, by definition, business costs that reduce profit. It is therefore natural for businesses to act in ways that will reduce such costs. History is littered with examples of the concerted tactics employed by the industry to protect profit at the expense of long-term worker (and public) health. These examples include denying scientific evidence that an exposure is harmful, blaming diseases on other exposures or personal habits, manufacturing controversies by withholding research results and sponsoring the creation of inconsistent evidence, threatening scientists and governments with litigation, lobbying regulators against legislations on exposure controls and bans, and moving operations to jurisdictions with less protection for workers (Examples of these tactics used by the asbestos industry are described in the following: McCulloch 2006; Ruff 2014; Takahashi and Landrigan 2016).

Box 1 Continued

For topics within the current thesis, multiple sizeable industries have vested interests in the health effects of silica and diesel exhaust exposure, including construction, mining, oil and gas, automobile manufacturing, and transportation. It is perhaps no coincidence that there are often different conclusions on the lung carcinogenicity of silica and diesel exhaust between industry-sponsored versus non-sponsored studies (Specific examples for diesel exhaust and silica, respectively, are available in the following commentaries and their references: Silverman 2018; Manno et al. 2018).

Analogous to how some businesses have shifted the environmental costs of operation upon the planet and its population, many also externalized the health costs onto the public healthcare systems and individuals. While academic-industry partnerships can be productive in many research areas, we must realize that, regarding chronic occupational health conditions, there exists inherent conflicts of interests that cannot be directly reconciled. Given this fact, below are my opinions on possible ways forward depending on our goals.

If the goal is scientific truth-seeking and knowledge generation, we must remain independent from industry influence. On the research project level, we must be able to analyse data and report results in ways that are free of opinions and limitations from the industry. On the scientific discipline level, we must also actively seek out and exclude potential biases seeded by the industry from our evidence base for generating knowledge. For instance, in meta-analyses for chronic occupational health topics, results may be stratified by the presence of industry-sponsorship. This way, any difference between studies with and without potential conflicts of interest could be identified and evaluated. Although some may argue (validly) that certain industry-sponsored investigators perform good scientific work, it is naive to accept this as justification to treat all industry-sponsored work equally as non-sponsored work. This is because in addition to the inherent conflicts of interests, there is ample precedent of the industry's deliberate and concerted efforts to sabotage advances in science (e.g. asbestos industry on mesothelioma, tobacco industry on lung cancer, oil and gas industry on climate change). Although individual scientists have also been known to fabricate scientific results for personal gains, there are no instances of scientific and academic institutions colluding on a global scale to obscure research findings for more funding.

If our goal is to improve the health of workers and to hold the industry accountable for its activities, the actions required must take place beyond the scientific community and in the social and political realm. As such this is clearly well beyond the scope of the thesis, and I could only offer a few viewpoints. At the very least,

Box 1 Continued

industry-funded scientific results on chronic occupational disease must be excluded from policymaking. The trend of global governments to place the onus of worker health on employers is not in itself a problem – employers should be responsible for maintaining a healthy workplace and compensation for any occupational illnesses. The problem lies in accurately assessing the size of this responsibility and holding employers accountable through proper enforcement of fair and protective regulations. Currently in the Netherlands, if a work exposure causes 4 cancers per 1,000 lifetime workers, we deem this a risk that is tolerable (Health Council of the Netherlands, 2019). However, if there is a workplace with this particular exposure and 10,000 employees, my opinion is that it would be fair for the employer to be held financially responsible for the screening and treatment of these 40 cancers unless the exposure is reduced or eliminated. On an individual basis, it may be difficult to prove that a certain chronic disease is caused by a particular work exposure. However, this is possible on a population scale. Similar to how a carbon tax gives incentives for all businesses to reduce emission, a health tax will give incentives for all employers to reduce or eliminate known dangerous exposures in the workplace. In addition to promoting health and well-being for the working public, protective regulations and fair taxation of health hazards may upend the current cost and return dynamic of businesses. If chronic occupational health research could be seen by businesses as offering potential tax savings rather than potential revenue cuts, conflicts of interests may reduce and more productive collaboration opportunities may arise.

None of the aforementioned actions are easy. If we consider asbestos again, many colleagues will agree that it is in some sense the poster child of occupational health research for which we completed the entire scientific product cycle from hazard identification and causal inference to risk assessment and informing policymaking. However, decades after this work, asbestos is still being mined, traded, and used globally at volumes measured in millions of metric tons per annum (Takahashi & Landrigan, 2016). It is clear that much social, political, and ethical work remains to be done for one of our most exemplar scientific achievements. The storyline is unlikely to be very different for other exposures. We have much to fight for not only as scientists, but also as citizens, taxpayers, and workers.

Improving exposure assessment

Within the context of case-control studies within the general population, "assessing the exposures" refers to retrospective exposure assessment. Chapter 2 has provided a review of historical and current exposure assessment approaches used in case-control studies in the population and explored the reliability performance of these approaches. Although the literature review was limited to studies that focused on carcinogen exposure, many of these findings generalize to other chronic non-communicable diseases (NCD) in occupational epidemiology. In brief, exposure assessment approaches used today are largely similar to those used 30 years ago: the majority of current studies continue to rely on case-by-case expert assessment, job-exposure matrices (JEMs), and subject-reported exposures. Reliability of these methods appeared to be highly variable.

Innovations on a fundamental level, similar to the progression from case-by-case expert assessment to JEMs, is extremely challenging and hampered by the scarcity of exposure information in population-based studies. This is particularly the case for population-based case-control studies where exposures must be re-constructed retrospectively from a job title and, if one is lucky, some task descriptions.

Future changes in sampling technology may stimulate a new paradigm shift in occupational exposure assessment. For example, if mobile devices or passive samplers such as smartwatches, smart glasses and silicone wristbands enable the collection of exposure and contextual information on a personal level, more opportunities may open up for better reconstruction of past exposures. The combination of real-time personal monitoring, cloud data storage, and registry information may even allow the creation of a personal occupational passport, in which all work-related exposures are recorded and accessible to the worker, and if permitted, healthcare providers and researchers. A version of this personal exposure passport for documenting occupational radiation exposure in specific sectors is already implemented in Germany and proposed in other European countries (German Federal Office for Radiation Protection, 2020; IAEA, 2003).

For now, the few areas of innovation in exposure assessment highlighted in Chapter 2 represent incremental improvements such as the use of modular specific questionnaires, incorporation of measurement data with statistical models, and machine learning models to "mimic" expert assessment. These innovations tackle three central challenges in exposure assessment in all occupational studies: a) obtaining more reliable exposure-relevant information from subjects; b) incorporating exposure measurements in a systematic and reproducible manner; and c) applying expert judgment with better transparency, efficiency, and consistency.

Chapter 3 focused on the subject interview phase: we automatically assigned subjects more detailed questionnaires in real-time based on interview responses with the goal of improving exposure-related information extraction. Chapter 4 further explored the reliability of job coding between independent coding teams, which is an essential step in translating job information reported by subjects to estimated exposures. Study phases covered in Chapters 3 and 4, namely subject interview and job coding, represent the foundation of exposure assessment and the two stages are intricately linked. The following section will first discuss potential ways to improve job coding, and then explore how these advances may also directly benefit data collection during subject interview.

Job coding involves categorizing jobs in standardized groups. For population-based studies, the original job descriptions may originate from subject interviews or registries. Common categorization schemes include standardized coding systems such as the International Standard Classification of Occupations (ISCO) from the International Labour Organization or the Standard Occupation Classification (SOC) from the United States. In current occupational health studies, the task of translating job descriptions to a standardized code is almost always performed manually. Although the work is straightforward to describe and usually simple to execute in small collections, the sheer workload of coding all jobs in a modern population-based study is often daunting. Chapter 4 presented the job coding work involved in a case-control study in China involving 12,590 subjects and 34,353 total reported jobs. To ensure coding quality, all jobs were double coded and discordant codes (n=15,562) were further reviewed and recoded. Under the assumption that coders could assign or review one code every minute and that this average rate is maintained through tens of thousands of codes, the job coding work presented in Chapter 4 would have taken more than 1,400 hours (~36 working weeks) of manual coding. In the real-world setting, when combined with having separate phases of subject interviews, scheduling, delays, and other practical challenges of multi-centered international collaboration, the coding work in Chapter 4 took approximately four years from beginning to completion. This challenge is, of course, not exclusive to population-based case-control studies. The CONSTANCES prospective population-based study, for example, follows a cohort of adults in France starting in 2012 (Zins & Goldberg, 2015). As of December 2019, the cohort included more than 150,000 participants who reported more than 500,000 jobs; manual coding of occupation and industry for these jobs took approximately 10 full-time working years.

In addition to the enormous workload, the selection and training of the job coders also presents a challenge. Over the lengthy study period, staff changes may affect the coding team, resulting in the need for retraining and reevaluation of coding quality. Further, due to the large total coding workload, the work is often divided between different coders. Both scenarios introduce inter-rater differences in the job codes. Although, as shown in Chapter 4 and other studies (Koeman et al., 2013; Kromhout & Vermeulen, 2000), inter-rater differences at the job coding level tend to diminish once exposure is assigned, the practice undoubtedly introduces errors into the exposure assessment results. If the job coding task is divided among coders by job or industry (which is an intuitive and efficient way of job coding labour division), there is even a potential for these errors to be differential based on occupation.

As the field moves forward, the challenge in job coding will only exacerbate for at least three reasons: 1) study size, and thus coding workload, will become increasingly large; 2) the growing cost of manual coding; and 3) future subjects will likely have, on average, more jobs per person than past and current subjects due to changes in the job market and economy (Boyd, 2016). If occupational epidemiology studies, regardless of study design, were to proceed to increase their study size to hundreds of thousands or even millions of participants, the laborious job coding process must transform.

Automation of the job coding process is the ideal solution to the job coding problem in population-based studies. Multiple existing automatic coding systems and assistants have been created for job coding, and some show promise in replacing, either completely or partially, manual labour in job coding (Table 1). Although some automatic coding systems claimed accuracy rates of 80% to over 90% (Schuhl, 1996; Warwick Institute for Employment Research, 2018), little information is available regarding precisely how these performance results were obtained. Reliability results from peer-reviewed publications with software-human coding comparisons were similar compared to those obtained from comparing independent human coders – exact agreement at a high coding granularity level is around 40-50% (Koeman et al., 2013; Kromhout & Vermeulen, 2000).

Challenges for developing an automatic coding system that can outperform human coders in terms of both accuracy and efficiency may be roughly categorized into two types: technical and organizational. The technical challenges include ambiguous and overly succinct job descriptions, language differences in job descriptions and coding systems, dealing with errors in job descriptions, and use of non-standard abbreviations in job records. The organizational challenge has to do with the fact that different groups have been working separately to create algorithms to perform coding in the local language and/or coding systems. For instance, separate tools have been developed for automatically coding job descriptions in English to ISCO, US SOC, UK SOC, and the Canadian National Occupational Classification (NOC) coding systems (CDC, 2019; De Matteis et al., 2017; Rémen et al., 2018; Russ et al., 2016). While these algorithms were likely created as tools to alleviate the coding burden in particular studies (and may have succeeded as such), they represent missed opportunities in at least two aspects.

The first is in terms of efficiency and performance. A concerted effort in improving the performance of existing coding algorithms and adapting their use in different regional context is likely more efficient than reinventing the wheel each time. An ideal automatic coding system could take input in different languages and provide coding in different classification schemes, just as Google Translate automatically detects input language and provide the appropriate translation in the selected output language.

Table 1. Overview of existing automatic coding systems

Automatic coding system/assistant	Job description language	Job classification scheme	Agreement with manual coding
CAPS-Canada (Rémen et al., 2018)	EN, FR	ISCO 1968-2008 CCDO 1971-1989 CNOC 2011 US SOC 2010	
CAPS-France (French Public Health Agency & Bordeaux School of Public Health, 2013)	EN, FR	ISCO 1968-2008 PCS 1994/2003	
CASCOT (Warwick Institute for Employment Research, 2018)	EN, FR, NL, IT, ES, DE, PL	UK SOC 1990-2010 ISCO 2008	80%*
NIOCCS (CDC, 2019)	EN	US SOC 2010	
O*NET (National Center for O*NET Development, 2020)	EN	O*NET-SOC	
OSCAR (De Matteis et al., 2017)	EN	UK SOC 2000	3-digit: 52%
SOCcer (Russ et al., 2016)	EN	US SOC 2010	6-digit: 45%
SOCEye (Burstyn et al., 2014)	EN	US SOC 2010	6-digit: 51%
SICORE (Schuhl, 1996)	FR	PCS 1982	>90%*
SOIC (Patel et al., 2012)	EN	US BOC 1990	3-digit: 73%

%: Percent agreement;
*Not peer-reviewed; no testing details available.
Abbreviations: EN – English, FR – French, NL – Dutch, IT – Italian, ES – Spanish, DE – German, PL – Polish.
BOC: Bureau of Census Index of Industries and Occupations; CCDO: Canadian Classification and Dictionary of Occupations; CNOC: Canadian National Occupational Classification; ISCO: International Standard Classification of Occupations; PCS: Professions and Socio-professional Categories; SOC: Standard Occupation Classification.

The second, which is arguably much more important, is an opportunity to create a better classification system that is tailored to occupational exposure assessment. Current job classification systems were developed and updated for socio-economic rather than exposure assessment purposes. Job distinctions important for assessing the economy and labour market may at times coincide with those important for assessing occupational exposures, but sometimes they differ vastly. For instance, both underground and open-pit miners belong to the same job under the ISCO coding scheme (e.g. 7-11.05 Miner (general) in ISCO-1968 and 7111 Miners and quarry workers in ISCO-1988). Although these jobs might share similarities in terms of economic or labour statistics, in terms of occupational exposure an underground miner is likely to have much higher exposures to diesel engine exhaust or silica than a surface miner due to the enclosed working environment. If an automatic coding system has the accuracy and efficiency to be widely used in different occupational epidemiology studies in different countries, an opportunity will then emerge for a multinational collaborative effort to create a set of standardized job codes specifically for exposure assessment. The tailored job codes may further become dynamic based on exposure, geographical region, and time period to provide better job groups under different scenarios. New job codes could be created for novel occupations (e.g. social media content moderators) or for traditional occupations that have evolved into jobs with potentially different exposures (e.g. diagnostic laboratory technicians versus drive-thru testing center technicians). Alternatively, separate job codes may be merged if exposures do not differ under some scenarios (e.g. work exposure to wood dust is essentially zero for all administrative workers). A dynamic, multilingual job coding system will also eliminate the need for crosswalking between different coding schemes, which in turn allows the universal application of exposure assessment tools across nations and populations with similar exposure patterns (e.g. applying a Canadian JEM to an American population).

Although the technical problems of automatic job coding are far from trivial, creating and adopting an automatic coding system could directly provide some solutions. As we have shown in Chapter 4, one of the most challenging aspects of job coding for human coders is ambiguity in job descriptions. For instance, a "cleaner" may be cleaning the streets, an office, a factory, or any item along a production process. With such vague descriptions, no human or machine coder will be able to assign job codes accurately without additional information. However, if automatic coding systems were implemented real-time in the interview stage, study subjects may be given the opportunity to immediately clarify any ambiguity or errors in job history reporting. For instance, a "welding-shop owner" may be asked follow-up questions such as "Do you also weld?" to help distinguish between administrative and production roles. Real-time job-coding software could also improve the application of task-specific question modules during subject interview, similar to those presented in Chapter 3, which provide important task-based information for exposure assessment. Continuing from the previous example, after confirmation of welding activity, the welding shop owner may be asked, "On average, how many hours per week do you weld? What type of welding do you do?"

Recent advances in machine learning and natural language processing present a timely opportunity for the development of automated job coding systems. If more coordinated effort is focused on either developing new or improving current coding algorithms, in the near future manual coding may no longer burden occupational studies. In addition to cost and time savings, added benefits of coding automatically include better coding reliability, reduced error rate, relative ease of evaluation and further training to improve coding performance, better use of expert time, and a real potential for developing dynamic, specific job codes for exposure assessment.

Better epidemiological analysis

Chapters 5-7 focused on areas classically defined under "epidemiology" in occupational health research: association discovery and assessment of disease risks/ burden. Chapter 5 explored associations between occupational benzene exposure and a number of lymphohaematopoietic malignancies in a cohort that essentially represented the entire Swiss working population. Chapters 6 and 7 estimated the lung cancer risks and burden associated with silica and diesel exhaust exposure respectively in a pooled sample from 14 case-control studies from Europe and Canada. These three studies represent progress on two fronts. The first is an effort in recent decades to further increase the size and power of population-based studies by pooling separate studies or using registry-based information available for entire populations. This is particularly important for both increasing the precision of disease risk estimates and investigating associations for rare diseases and disease subtypes. Examples of other large pooled or registry-based studies include the pooled analyses of case-control studies for sinonasal (Luce et al., 2002) and laryngeal cancers (Hall et al., 2020), as well as the Danish registry-based cohort DOC*X (Flachs et al., 2019).

The second is the use of fully quantitative exposure estimates to calculate disease risks and burdens related to a unit of exposure. Historically, quantitative exposure measurements and disease estimates typically come from industry-based studies, which lack sufficient statistical power to investigate rare health outcomes. The higher levels of exposure usually observed in industry-based cohorts also sometimes raise the concern of generalizability to the general public with lower exposure levels. For population-based studies, expressing disease risk per unit of exposure allows for direct comparison with results from industry-based studies and for informing risk assessment and policymaking. The work in Chapters 5 and 6 built upon the foundation of quantitative JEMs created earlier, namely the BEN-JEM for benzene (Spycher et al., 2017) and the SYN-JEM for silica (Peters et al., 2011, 2012), enabling quantitative exposure assessment in population-based studies. In Chapter 7 we built a new quantitative JEM for diesel engine exhaust. Improved and quantitative exposure assessment allows for reporting of disease burden estimates that are quantifiable and dynamic based on different exposure scenarios, representing an important step forward as it ultimately increases the utility and impact of all occupational studies both in the general population and in industry-based cohorts.

In theory, moving exposure estimates from a categorical to a quantitative scale is also expected to result in a reduction in exposure misclassification between different occupations, which would ultimately lead to less attenuation in disease risk analyses. In practice, evidence for a notable gain in risk slopes with the use of quantitative exposure estimates in population-based studies is mixed. For instance, lung cancer risk estimates associated with quantitative diesel exhaust exposure from Chapter 7 were very similar to risks from earlier work with ordinal diesel exhaust exposure assignment (Olsson et al., 2011). Linear continuous risks from Chapter 7 were also approximately 2-3 times lower compared to those derived from industry-based cohorts (Vermeulen et al., 2014; Vermeulen & Portengen, 2016). For exposure to crystalline silica, the lung cancer risk slope calculated in Chapter 6 (1.05 for every unit increase in log-transformed cumulative silica exposure in mg/m^3-year) was more similar to those from industry-based cohorts (1.07 from a pooled analysis of 10 studies; 1.06 from a large cohort in China) (Liu et al., 2013; Steenland et al., 2001). The apparent difference in risk attenuation between lung cancer risks associated with silica and diesel exhaust may be due to the fact that 23,640 measurements were used to quantify silica exposure (Peters et al., 2011) whereas 4,417 measurements were used for diesel exhaust. In addition to indicating the obvious quantitative disparity in the number of available measurements to estimate exposure, this five-fold data difference also represents a qualitative difference in data coverage for exposed occupations. The silica SYN-JEM assigned exposure to 159 ISCO-68 occupations and had silica exposure measurements for 428 unique ISCO-68 occupations whereas the DEE-JEM had diesel exhaust measurements for only 39 unique job titles. To create the final DEE-JEM, which assigned exposures to 248 ISCO-68 occupations, significant gaps in data coverage must be filled using expert judgment and extrapolations between similar occupations. Further, the varying amounts of evidence for the carcinogenic effects of silica versus diesel exhaust may also be a factor – there has been decades of occupational health research on silica whereas diesel is a relatively new topic. More work is needed to confirm if quantitative exposure estimates result in attenuation reduction during risk modeling, and if so, the amount of data needed for this effect to be appreciable in magnitude.

Similar to most current occupational epidemiology studies on chronic diseases, Chapters 5-7 largely focused on characterizing the association between one well-known exposure and risks of one (or a closely related group of) well-defined health outcome. Although this general approach has thus far generated much of the current knowledge in occupational epidemiology, it also leaves several topics for potential improvements.

Multiple exposures

It is clear that real-life situations rarely involve only a single exposure leading to one disease. We know that many exposures are associated with higher risks of the same health outcomes, such as lung cancer from exposures to diesel engine exhaust, crystalline silica, asbestos, radon, hexavalent chromium, nickel, and PAHs. This becomes even more apparent when we move beyond the occupational setting and include environmental exposures and lifestyle habits, such as outdoor air pollution and active/passive cigarette smoke. Although the precise mechanisms of causing human disease are still unclear for most exposures, it seems reasonable to assume that certain commonalities and interactions exist among adverse outcome pathways for different exposures causing similar diseases. Some progress has already been made in accounting for other concurrent exposures in modern occupational studies. For instance, in population-based studies focusing on respiratory health outcomes, collection of information on cigarette smoking and its adjustment as a confounder has become a standard. Studies making model adjustments for contaminant occupational exposures, such as what we have done in Chapters 6 and 7, are also becoming more common. However, these advances are made mostly in the premise of confirming the existence and validity of a particular exposure-disease relationship rather than to understand the true combined effect of multiple exposures and their interactions on disease risks.

Epidemiological studies in environmental health offer some ideas for investigating the combined versus individual effects of multiple exposures. Research on air pollution and health, for instance, might compare results from single- versus multiple-pollutant models, report interactions between multiple air pollutants in pairs, and estimate the joint effect of exposure to a group of pollutants (Crouse et al., 2015; Jerrett et al., 2013; Klompmaker et al., 2019). For occupational exposures, we may follow or further expand beyond these approaches. Imagine a study where we are interested in characterizing the effect of work exposures to silica, diesel exhaust, and chromium on lung cancer. It is feasible to include subjects who were exposed to one, two, or all three substances for our study. Given a large population with enough exposure contrast, it may even be possible to compare the effects of specific exposure sequences among subjects with multiple exposures.

Venturing into the unknown

The complexity of occupational exposures naturally involves more than multiple exposures to known and probable work hazards. Exploration beyond known relationships also introduces multiple exposures and outcomes. Known substances or health outcomes may be screened for new associations. Unknown exposures and diseases may similarly be investigated through screening occupations and biomarkers. However, models with multiple simultaneous inference tests present another problem: multiple comparisons. The challenge of multiple comparisons is neither new nor unique to occupational epidemiology. Studies investigating the effect of diet or genetic factors, for example, have faced similar issues (Chen & Witte, 2007; Witte et al., 1994). Adjustments for multiple comparisons may be made purely based on a mathematical basis (e.g. via the Bonferroni method), which assumes that all tests have the same low prior. However, in occupational epidemiology this is generally not favored, because it has been argued that observed associations between exposures and outcomes do not merely represent random chance (Perneger, 1998; Rothman, 1990). If the observed associations in a study were a mixture of true associations and random error, adjustments that consider only random error may eliminate true associations, leading to increased probability of missing true associations and reporting false-negative findings (i.e. more type II errors).

Although some have argued that adjustments for multiple comparisons are unnecessary in occupational epidemiology (Rothman, 1990), others have proposed different approaches of reducing potential type I errors to better highlight associations that may be true. These approaches typically apply different shrinkage statistics to effect betas and variations, such as adjustment using either empirical or semi-Bayes methods (Corbin et al., 2008, 2012), hierarchical models (Corbin et al., 2012), and machine learning models (Mieth et al., 2016). Another potential solution is to cluster the covariates into groups, then running the analysis with clustered groups instead of individual covariates to reduce the number of parallel tests performed (Molitor et al., 2010). The suitability of the different approaches and appropriateness of the various amounts of shrinkage depend on the specific aims of the study. If we were planning to commit significant resources for a new study and need to set priorities on what

to focus on, a more conservative model increases our chances of getting it right and prevents wasting resources on false leads. On the other hand, if we want to explore the different exposures/jobs for new ideas and insight, a more lenient model may allow for the retention of more interesting information. For instance, in a recent agnostic exploration of occupations as risk factors for lung cancer among the SYNERGY population using lasso penalized regressions, more lenient models highlighted that being a teacher in a range of education levels have protective effects on lung cancer (Ge et al., 2020).

Temporality of exposure

The ubiquitous use and acceptance of the lifetime cumulative exposure parameter for describing total exposure suggest another set of generalizations typically made in occupational studies regarding the temporal variabilities of exposure. As the product of exposure intensity and duration, cumulative exposure accounts for one temporal aspect of exposure (i.e. duration) but ignores many others, including starting age, time since exposure cessation, and exposure rate. These temporal factors, however, are expected to affect disease development and risks. It is very unlikely that a worker exposed to low concentrations since adolescence would have the same disease risk as another exposed to high concentrations close to retirement age, even if they score similarly on the cumulative exposure scale. This hypothesis is confirmed by studies on cigarette smoking, which showed that given the same cumulative exposure (i.e. pack-years of cigarette smoking), differential lung cancer risks are associated with different smoking duration, intensity, and time-since-quitting (IARC, 2007; Vlaanderen et al., 2014). Temporal exposure characteristics may even change how an exposure is related to different diseases. For instance, recent studies on occupational benzene exposure showed that distant exposure (>10 year) were more related to risks of non-Hodgkin's lymphoma than recent exposure (<10 year), yet recent exposure was more related to risks of acute myeloid leukemia than distant exposures (Hayes et al., 1997; Linet et al., 2019).

In Chapters 6 and 7, we have included exposure duration and time since last exposure in our risk models to individually assess how these temporal variables relate to disease. Further work that combines cumulative exposure and exposure rate may follow the approach developed by Lubin and colleagues, where the cumulative exposure parameter in a linear model was further modified by additional factors such as exposure intensity and time since last exposure (Lubin et al., 2007; Lubin & Caporaso, 2006; Vlaanderen et al., 2014). Other more complex methods to model exposure-time-response, including the multistage models (Freedman & Navidi, 1990), two-stage clonal expansion models (Hazelton et al., 2005), and compartmental models (Chadeau-Hyam et al., 2014) may also be explored.

Summary and future perspectives

As in all scientific specialties, “product development” in occupational health research is a continuous and iterative process. My work was built upon the efforts of earlier research and it will hopefully now serve as foundations for future work.

Future developments in different areas in exposure assessment and epidemiology are valuable because, in addition to addressing the specific issues, they collectively promote better scientific products that drive societal change. For instance, further work on multiple exposures and exposure-time-response models may aid in deriving disease risk estimates that are more accurate in various subgroups (e.g. young workers, women, workers with consecutive or concurrent exposures), leading to better burden of disease estimates for targeted intervention and regulatory action.

Population-based studies, particularly large pooled studies and registry-based studies, represent a unique testing ground for many new ideas. To investigate multiple exposures and exposure-time-response, for example, a study would need access to a population with adequate contrast in the different exposures and detailed exposure assessment across all relevant exposures. For agnostic exploration of occupation-disease associations, a population with sufficient sample size in each occupation is needed.

One ongoing effort to create larger pooled studies is the Exposome Project for Health and Occupational Research (EPHOR) project, which aims to construct a mega-cohort with over 20 million study subjects by combining over 40 existing European population- and occupational-cohort studies (EPHOR, 2020). An open access online inventory of study cohorts and exposure assessment tools for occupational exposures is being assembled as part of OMEGA-NET (Kogevinas et al., 2020; Peters et al., 2020). Projects of this magnitude do not merely open the door to innovations, they demand them. Job coding, for instance, must be reformed and automated at least to some degree, because none of us could manually assign/crosswalk job codes for tens of millions of subjects. Future work in developing and comparing novel methods within these and other projects will provide evidence for determining the advantages and disadvantages of various approaches, and what additional benefits they bring over current standard methodology in occupational epidemiology.

REFERENCES:

Boyd, L., Carol. (2016). *The life of American workers in 1915: Monthly Labor Review: U.S. Bureau of Labor Statistics*. https://www.bls.gov/opub/mlr/2016/article/the-life-of-american-workers-in-1915.htm

Burstyn, I., Slutsky, A., Lee, D. G., Singer, A. B., An, Y., & Michael, Y. L. (2014). Beyond crosswalks: Reliability of exposure assessment following automated coding of free-text job descriptions for occupational epidemiology. *The Annals of Occupational Hygiene*, *58*(4), 482–492. https://doi.org/10.1093/annhyg/meu006

CDC. (2019). *NIOSH Industry and Occupation Computerized Coding System (NIOCCS)*. https://wwwn.cdc.gov/nioccs3/Default.aspx

Chadeau-Hyam, M., Tubert-Bitter, P., Guihenneuc-Jouyaux, C., Campanella, G., Richardson, S., Vermeulen, R., De Iorio, M., Galea, S., & Vineis, P. (2014). Dynamics of the risk of smoking-induced lung cancer: A compartmental hidden Markov model for longitudinal analysis. *Epidemiology (Cambridge, Mass.)*, *25*(1), 28–34. https://doi.org/10.1097/EDE.0000000000000032

Chen, G. K., & Witte, J. S. (2007). Enriching the Analysis of Genomewide Association Studies with Hierarchical Modeling. *American Journal of Human Genetics*, *81*(2), 397–404.

Corbin, M., Maule, M., Richiardi, L., Simonato, L., Merletti, F., & Pearce, N. (2008). Semi-Bayes and empirical Bayes adjustment methods for multiple comparisons. *Epidemiologia E Prevenzione*, *32*(2), 108–110.

Corbin, M., Richiardi, L., Vermeulen, R., Kromhout, H., Merletti, F., Peters, S., Simonato, L., Steenland, K., Pearce, N., & Maule, M. (2012). Hierarchical regression for multiple comparisons in a case-control study of occupational risks for lung cancer. *PloS One*, *7*(6), e38944. https://doi.org/10.1371/journal.pone.0038944

Crouse, D. L., Peters, P. A., Hystad, P., Brook, J. R., van Donkelaar, A., Martin, R. V., Villeneuve, P. J., Jerrett, M., Goldberg, M. S., Pope, C. A., Brauer, M., Brook, R. D., Robichaud, A., Menard, R., & Burnett, R. T. (2015). Ambient PM2.5, O_3, and NO_2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environmental Health Perspectives*, *123*(11), 1180–1186. https://doi.org/10.1289/ehp.1409276

De Matteis, S., Jarvis, D., Young, H., Young, A., Allen, N., Potts, J., Darnton, A., Rushton, L., & Cullinan, P. (2017). Occupational self-coding and automatic recording (OSCAR): A novel web-based tool to collect and code lifetime job histories in large population-based studies. *Scandinavian Journal of Work, Environment & Health*, *43*(2), 181–186. https://doi.org/10.5271/sjweh.3613

EPHOR. (2020). *EPHOR Work packages*. EPHOR Project. https://www.ephor-project.eu/work-packages

Flachs, E. M., Petersen, S. E. B., Kolstad, H. A., Schlünssen, V., Svendsen, S. W., Hansen, J., Budtz-Jørgensen, E., Andersen, J. H., Madsen, I. E. H., & Bonde, J. P. E. (2019). Cohort Profile: DOC*X: a nationwide Danish occupational cohort with eXposure data – an open research resource. *International Journal of Epidemiology*, *48*(5), 1413–1413k. https://doi.org/10.1093/ije/dyz110

Freedman, D. A., & Navidi, W. C. (1990). Ex-smokers and the multistage model for lung cancer. *Epidemiology (Cambridge, Mass.)*, *1*(1), 21–29. https://doi.org/10.1097/00001648-199001000-00006

French Public Health Agency, & Bordeaux School of Public Health. (2013). *CAPS : Codage Assisté des Professions et Secteurs d'activité*. https://ssl3.isped.u-bordeaux2.fr/CAPS-FR/Langue.aspx

Ge, C., Portengen, L., Olsson, A., Brüning, T., Kromhout, H., Straif, K., & Vermeulen, R. (2020). *Agnostic exploration of employment in various occupations as risk factors for lung cancer*. Search Results Web results Exposome Symposium 2020, New York.

German Federal Office for Radiation Protection. (2020). *The radiation passport*. Federal Office for Radiation Protection; BfS. https://www.bfs.de/EN/topics/ion/radiation-protection/occupation/methodology/radiation-passport.html

Hall, A. L., Kromhout, H., Schüz, J., Peters, S., Portengen, L., Vermeulen, R., Agudo, A., Ahrens, W., Boffetta, P., Brennan, P., Canova, C., Conway, D. I., Curado, M. P., Daudt, A. W., Fernandez, L., Hashibe, M., Healy, C. M., Holcatova, I., Kjaerheim, K., … Olsson, A. (2020). Laryngeal Cancer Risks in Workers Exposed to Lung Carcinogens: Exposure-Effect Analyses Using a Quantitative Job Exposure Matrix. *Epidemiology (Cambridge, Mass.)*, *31*(1), 145–154. https://doi.org/10.1097/EDE.0000000000001120

Hayes, R. B., Yin, S. N., Dosemeci, M., Li, G. L., Wacholder, S., Travis, L. B., Li, C. Y., Rothman, N., Hoover, R. N., & Linet, M. S. (1997). Benzene and the dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine—National Cancer Institute Benzene Study Group. *Journal of the National Cancer Institute*, *89*(14), 1065–1071.

Hazelton, W. D., Clements, M. S., & Moolgavkar, S. H. (2005). Multistage carcinogenesis and lung cancer mortality in three cohorts. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *14*(5), 1171–1181. https://doi.org/10.1158/1055-9965.EPI-04-0756

Health Council of the Netherlands. (2019, March 13). *Diesel Engine Exhaust: Health-based recommended occupational exposure limit*. https://www.gezondheidsraad.nl/binaries/gezondheidsraad/documenten/adviezen/2019/03/13/dieselmotoremissie/Diesel+Engine+Exhaust.pdf

IAEA. (2003). *OCCUPATIONAL RADIATION PROTECTION: PROTECTING WORKERS AGAINST EXPOSURE TO IONIZING RADIATION*. INTERNATIONAL ATOMIC ENERGY AGENCY.

IARC. (2007). *IARC Handbooks of Cancer Prevention: Tobacco Control* (Volume 11). https://publications.iarc.fr/Book-And-Report-Series/Iarc-Handbooks-Of-Cancer-Prevention/Tobacco-Control-Reversal-Of-Risk-After-Quitting-Smoking-2007

Jerrett, M., Burnett, R. T., Beckerman, B. S., Turner, M. C., Krewski, D., Thurston, G., Martin, R. V., van Donkelaar, A., Hughes, E., Shi, Y., Gapstur, S. M., Thun, M. J., & Pope, C. A. (2013). Spatial Analysis of Air Pollution and Mortality in California. *American Journal of Respiratory and Critical Care Medicine*, *188*(5), 593–599. https://doi.org/10.1164/rccm.201303-0609OC

Klompmaker, J. O., Janssen, N. A. H., Bloemsma, L. D., Gehring, U., Wijga, A. H., van den Brink, C., Lebret, E., Brunekreef, B., & Hoek, G. (2019). Associations of Combined Exposures to Surrounding Green, Air Pollution, and Road Traffic Noise with Cardiometabolic Diseases. *Environmental Health Perspectives*, *127*(8), 87003. https://doi.org/10.1289/EHP3857

Koeman, T., Offermans, N. S. M., Christopher-de Vries, Y., Slottje, P., Van Den Brandt, P. A., Goldbohm, R. A., Kromhout, H., & Vermeulen, R. (2013). JEMs and incompatible occupational coding systems: Effect of manual and automatic recoding of job codes on exposure assignment. *The Annals of Occupational Hygiene*, *57*(1), 107–114. https://doi.org/10.1093/annhyg/mes046

Kogevinas, M., Schlünssen, V., Mehlum, I. S., & Turner, M. C. (2020). The OMEGA-NET International Inventory of Occupational Cohorts. *Annals of Work Exposures and Health*, *64*(6), 565–568. https://doi.org/10.1093/annweh/wxaa039

Kromhout, H., & Vermeulen, R. (2000). Long-term trends in occupational exposure: Are they real? What causes them? What shall we do with them? *The Annals of Occupational Hygiene*, *44*(5), 325–327. https://doi.org/10.1093/annhyg/44.5.325

Linet, M. S., Gilbert, E. S., Vermeulen, R., Dores, G. M., Yin, S.-N., Portengen, L., Hayes, R. B., Ji, B.-T., Lan, Q., Li, G.-L., Rothman, N., Ding, C., Dores, G. M., Gao, Y., Gilbert, E. S., Hayes, R. B., Ji, B.-T., Lan, Q., Li, G.-L., … Zhou, J.-S. (2019). Benzene Exposure Response and Risk of Myeloid Neoplasms in Chinese Workers: A Multicenter Case–Cohort Study. *JNCI: Journal of the National Cancer Institute*, *111*(5), 465–474. https://doi.org/10.1093/jnci/djy143

Liu, Y., Steenland, K., Rong, Y., Hnizdo, E., Huang, X., Zhang, H., Shi, T., Sun, Y., Wu, T., & Chen, W. (2013). Exposure-Response Analysis and Risk Assessment for Lung Cancer in Relationship to Silica Exposure: A 44-Year Cohort Study of 34,018 Workers. *American Journal of Epidemiology*, *178*(9), 1424–1433. https://doi.org/10.1093/aje/kwt139

Lubin, J. H., & Caporaso, N. E. (2006). Cigarette smoking and lung cancer: Modeling total exposure and intensity. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *15*(3), 517–523. https://doi.org/10.1158/1055-9965.EPI-05-0863

Lubin, J. H., Caporaso, N., Wichmann, H. E., Schaffrath-Rosario, A., & Alavanja, M. C. R. (2007). Cigarette smoking and lung cancer: Modeling effect modification of total exposure and intensity. *Epidemiology (Cambridge, Mass.)*, *18*(5), 639–648. https://doi.org/10.1097/EDE.0b013e31812717fe

Luce, D., Leclerc, A., Bégin, D., Demers, P. A., Gérin, M., Orlowski, E., Kogevinas, M., Belli, S., Bugel, I., Bolm-Audorff, U., Brinton, L. A., Comba, P., Hardell, L., Hayes, R. B., Magnani, C., Merler, E., Preston-Martin, S., Vaughan, T. L., Zheng, W., & Boffetta, P. (2002). Sinonasal cancer and occupational exposures: A pooled analysis of 12 case–control studies. *Cancer Causes & Control*, *13*(2), 147–157. https://doi.org/10.1023/A:1014350004255

Manno, M., Levy, L., Johanson, G., & Cocco, P. (2018). Silica, silicosis and lung cancer: What level of exposure is acceptable? *La Medicina Del Lavoro*, *109*(6), 478–480. https://doi.org/10.23749/mdl.v109i6.7928

McCulloch, J. (2006). Saving the Asbestos Industry, 1960 to 2006. *Public Health Reports*, *121*(5), 609–614.

Mieth, B., Kloft, M., Rodríguez, J. A., Sonnenburg, S., Vobruba, R., Morcillo-Suárez, C., Farré, X., Marigorta, U. M., Fehr, E., Dickhaus, T., Blanchard, G., Schunk, D., Navarro, A., & Müller, K.-R. (2016). Combining Multiple Hypothesis Testing with Machine Learning Increases the Statistical Power of Genome-wide Association Studies. *Scientific Reports*, *6*(1), 36671. https://doi.org/10.1038/srep36671

Molitor, J., Papathomas, M., Jerrett, M., & Richardson, S. (2010). Bayesian profile regression with an application to the National Survey of Children's Health. *Biostatistics (Oxford, England)*, *11*(3), 484–498. https://doi.org/10.1093/biostatistics/kxq013

National Center for O*NET Development. (2020). *O*NET OnLine*. https://www.onetonline.org/

Olsson, A. C., Gustavsson, P., Kromhout, H., Peters, S., Vermeulen, R., Brüske, I., Pesch, B., Siemiatycki, J., Pintos, J., Brüning, T., Cassidy, A., Wichmann, H.-E., Consonni, D., Landi, M. T., Caporaso, N., Plato, N., Merletti, F., Mirabelli, D., Richiardi, L., … Straif, K. (2011). Exposure to diesel motor exhaust and lung cancer risk in a pooled analysis from case-control studies in Europe and Canada. *American Journal of Respiratory and Critical Care Medicine*, *183*(7), 941–948. https://doi.org/10.1164/rccm.201006-0940OC

Patel, M. D., Rose, K. M., Owens, C. R., Bang, H., & Kaufman, J. S. (2012). Performance of automated and manual coding systems for occupational data: A case study of historical records. *American Journal of Industrial Medicine*, *55*(3), 228–231. https://doi.org/10.1002/ajim.22005

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, *316*(7139), 1236–1238. https://doi.org/10.1136/bmj.316.7139.1236

Peters, S., Turner, M. C., Bugge, M. D., Vienneau, D., & Vermeulen, R. (2020). International Inventory of Occupational Exposure Information: OMEGA-NET. *Annals of Work Exposures and Health*, *64*(5), 465–467. https://doi.org/10.1093/annweh/wxaa021

Peters, S., Vermeulen, R., Olsson, A., Van Gelder, R., Kendzia, B., Vincent, R., Savary, B., Williams, N., Woldbæk, T., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Dahmann, D., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Kromhout, H. (2012). Development of an exposure measurement database on five lung carcinogens (ExpoSYN) for quantitative retrospective occupational exposure assessment. *The Annals of Occupational Hygiene*, *56*(1), 70–79. https://doi.org/10.1093/annhyg/mer081

Peters, S., Vermeulen, R., Portengen, L., Olsson, A., Kendzia, B., Vincent, R., Savary, B., Lavoué, J., Cavallo, D., Cattaneo, A., Mirabelli, D., Plato, N., Fevotte, J., Pesch, B., Brüning, T., Straif, K., & Kromhout, H. (2011). *Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies*. *13*(11), 3262–3268. https://doi.org/10.1039/C1EM10628G

Rémen, T., Richardson, L., Pilorget, C., Palmer, G., Siemiatycki, J., & Lavoué, J. (2018). Development of a Coding and Crosswalk Tool for Occupations and Industries. *Annals of Work Exposures and Health*, *62*(7), 796–807. https://doi.org/10.1093/annweh/wxy052

Rothman, K. J. (1990). No Adjustments Are Needed for Multiple Comparisons. *Epidemiology*, *1*(1), 43–46.

Ruff, K. (2014). Asbestos: A continuing failure of ethics by McGill University. *International Journal of Occupational and Environmental Health*, *20*(1), 1–3. https://doi.org/10.1179/1077352513Z.000000000102

Russ, D. E., Ho, K.-Y., Colt, J. S., Armenti, K. R., Baris, D., Chow, W.-H., Davis, F., Johnson, A., Purdue, M. P., Karagas, M. R., Schwartz, K., Schwenn, M., Silverman, D. T., Johnson, C. A., & Friesen, M. C. (2016). Computer-based coding of free-text job descriptions to efficiently identify occupations in epidemiological studies. *Occupational and Environmental Medicine*, *73*(6), 417–424. https://doi.org/10.1136/oemed-2015-103152

Schuhl, P. (1996). *SICORE, THE INSEE AUTOMATIC CODING SYSTEM*.

Silverman, D. T. (2018). Diesel Exhaust and Lung Cancer-Aftermath of Becoming an IARC Group 1 Carcinogen. *American Journal of Epidemiology*, *187*(6), 1149–1152. https://doi.org/10.1093/aje/kwy036

Spycher, B. D., Lupatsch, J. E., Huss, A., Rischewski, J., Schindera, C., Spoerri, A., Vermeulen, R., & Kuehni, C. E. (2017). Parental occupational exposure to benzene and the risk of childhood cancer: A census-based cohort study. *Environment International*, *108*(Supplement C), 84–91. https://doi.org/10.1016/j.envint.2017.07.022

Steenland, K., Mannetje, A., Boffetta, P., Stayner, L., Attfield, M., Chen, J., Dosemeci, M., DeKlerk, N., Hnizdo, E., Koskela, R., & Checkoway, H. (2001). Pooled exposure–response analyses and risk assessment for lung cancer in 10 cohorts of silica-exposed workers: An IARC multicentre study. *Cancer Causes & Control*, *12*(9), 773–784. https://doi.org/10.1023/A:1012214102061

Takahashi, K., & Landrigan, P. J. (2016). The Global Health Dimensions of Asbestos and Asbestos-Related Diseases. *Annals of Global Health*, *82*(1), 209–213. https://doi.org/10.1016/j.aogh.2016.01.019

Vermeulen, R., & Portengen, L. (2016). Is diesel equipment in the workplace safe or not? *Occupational and Environmental Medicine*, *73*(12), 846–848. https://doi.org/10.1136/oemed-2016-103977

Vermeulen, R., Silverman, D. T., Garshick, E., Vlaanderen, J., Portengen, L., & Steenland, K. (2014). Exposure-response estimates for diesel engine exhaust and lung cancer mortality based on data from three occupational cohorts. *Environmental Health Perspectives*, *122*(2), 172–177. https://doi.org/10.1289/ehp.1306880

Vlaanderen, J., Portengen, L., Schüz, J., Olsson, A., Pesch, B., Kendzia, B., Stücker, I., Guida, F., Brüske, I., Wichmann, H.-E., Consonni, D., Landi, M. T., Caporaso, N., Siemiatycki, J., Merletti, F., Mirabelli, D., Richiardi, L., Gustavsson, P., Plato, N., … Vermeulen, R. (2014). Effect Modification of the Association of Cumulative Exposure and Cancer Risk by Intensity of Exposure and Time Since Exposure Cessation: A Flexible Method Applied to Cigarette Smoking and Lung Cancer in the SYNERGY Study. *American Journal of Epidemiology*, *179*(3), 290–298. https://doi.org/10.1093/aje/kwt273

Warwick Institute for Employment Research. (2018). *Cascot: Computer Assisted Structured Coding Tool*. https://warwick.ac.uk/fac/soc/ier/software/cascot/details/

Witte, J. S., Greenland, S., Haile, R. W., & Bird, C. L. (1994). Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology (Cambridge, Mass.)*, *5*(6), 612–621. https://doi.org/10.1097/00001648-199411000-00009

Zins, M., & Goldberg, M. (2015). The French CONSTANCES population-based cohort: Design, inclusion and follow-up. *European Journal of Epidemiology*, *30*, 1317–1328. https://doi.org/10.1007/s10654-015-0096-4

Appendices

SUMMARY

Occupational health research in the population

For most people, exposures at the workplace comprise a large portion of total environmental exposures. Generally, two types of studies are used to investigate how work exposures affect health: industry-based studies and population-based studies. As the name implies, industry-based studies involve participants from specific industries or occupations (e.g. aviation workers, pilots). Because subjects from industry-based studies typically work in similar workplaces, it is relatively straightforward to measure or estimate occupational exposures. Industry-based research may be the only feasible way to study some rare exposures if they occur only in limited specific workplaces. As an example, aviation jet fuel exposure only occurs for those working with jet fuels and engines. However, the size of industry-based research is limited by the size of the industry. Imagine a study where we successfully recruit all current and past jet plane pilots, study size will still be quite small since there are just not that many jet pilots around. A study with a few dozen or hundred jet pilots is unlikely to be very helpful for studying rare diseases such cancers and neurodegenerative diseases. Industry-based studies also place more emphasis on exposures occurring within the specific industry – potentially important additional factors such as exposures from previous jobs and personal lifestyle habits are sometimes unavailable.

Population-based studies, also sometimes called community-based studies, draw subjects from the general population. Recruiting subjects from the general population has notable advantages that complement the shortcomings in industry-based studies. One of the most important benefit is that the size of population-based studies is, in theory, limited only by the size of the general population. Although practical limitations affecting all scientific studies (e.g. funding, time) still apply, population-based studies have been able to achieve study sizes up to tens of thousands and even millions of participants. This large sample size allows us to study diseases or subtypes of diseases that are uncommon. Compared to industry-based studies, population-based studies also typically collect more information regarding total occupational history across multiple jobs and additional lifestyle factors. This additional information is useful for better understanding the relationships between work exposures and disease.

In order to study whether exposure X is related to disease Y, a study must include subjects who are exposed and unexposed to X, as well as subjects with and without disease Y (also called cases and controls). Based on how a study is designed, population-based studies may be described as either prospective or retrospective. A prospective study recruits subjects before disease occurs, follows the subjects until some become sick with the disease, then compare the exposures that occurred among those who are healthy versus sick. A retrospective study enrolls subjects known to be with or without the disease then compare the past exposures between these two groups. There are many pros and cons of prospective versus retrospective study design, the most obvious perhaps being a difference in efficiency. If we needed 200 cases of disease Y for a particular study, it would be much quicker to assemble 200 people already sick with Y and another 200 who do not suffer from Y than to round up a number of healthy people and wait for 200 to eventually develop disease Y. This is particularly true if Y is both relatively uncommon and takes a long time to develop, such as for instance pancreatic cancer. However, not everything is easier in retrospective studies. In a prospective study, up to date and more accurate occupation information may be obtained through regular interviews conducted while we are waiting for enough disease cases to occur. In turn, we can use this information to estimate work exposures experienced by study subjects (we call this exposure estimation "exposure assessment"). In retrospective studies, especially for chronic diseases, important work exposures related to disease may have occurred decades in the past. This makes exposure assessment challenging, because now we have to estimate work exposures that occurred decades earlier based on the full work history information the subjects gave during an interview. Subjects may or may not accurately recall jobs performed a long time ago. Workplaces and work tasks in previous decades may have ceased to exist today.

This thesis

The theme of my thesis revolves around these challenges in retrospective population-based occupational health studies: how can we do a better job in obtaining work history information from study subjects, how can we perform better exposure assessment with this information, how can we better calculate disease risks with the assessed exposures, and how can we make population-based studies both more useful in understanding the exposure-disease relationship and more impactful in improving worker and public health.

After Chapter 1 as a general thesis introduction, Chapter 2 reviewed methods and approaches used in exposure assessment in retrospective population-based occupational cancer studies. This review is the first to systemically review exposure assessment methods used in population-based studies in recent decades and its findings form the foundation in further exploration of potential innovations in exposure assessment. Specifically, Chapter 2 found that the majority of current population-based studies continues to rely on exposure assessment approaches developed decades earlier and highlighted potential areas of new development,

including how to: a) obtain more reliable job information from subjects; b) make better use of exposure measurements; and c) apply expert judgment with more transparency, efficiency, and consistency.

Chapter 3 evaluated the performance of a computer-algorithm based software for occupational history interviews in population-based studies. The algorithm was applied in a lymphoma and leukemia population-based case-control study in Asia (the AsiaLymph study) with 11,409 reported jobs. Based on different job titles and job tasks reported by the subject, the computer system automatically assigns different interview questions in real-time to better capture information relevant to work exposures. Compared to human expert selections, the computer algorithm selected the identical questions for 57.7% of reported jobs, and 86% of the computer-selected questions would collect the same amount of exposure-relevant information. The study demonstrated that computer algorithms can reliably assign work interview questions automatically in real-time and similar tools have the potential in improving the quality of job information collected from study participants.

Chapter 4 focused on the process of classifying subject reported jobs into standardized groups, also known as job coding, which is crucial for exposure assessment. The chapter investigated the difference between job coding assignments on the same jobs performed by different expert coders, and how this difference affects subsequent work in exposure assessment. This work was also carried out within the AsiaLymph study and included 34,353 jobs reported from 12,590 study subjects. Job coding agreement between different coders ranged from 51.0-77.1% depending on different coder pairs and level of detail in job coding. The study also found that this degree of difference in job coding generally becomes smaller in subsequent steps in exposure assessment, because not all job code disagreements result in different estimated exposures. Chapter 4 represents one of the largest job coding performance evaluation studies performed to date, where job coding was performed independently by two teams and reviewed by a third expert, with additional subsequent differences assessed in the assigned exposure level.

Chapter 5 estimated the mortality risks of lymphohaematopoietic cancers related to occupational exposure to benzene in a population-based cohort of 2.97 million Swiss citizens. The study found that occupational exposure to benzene is associated with elevated mortality risks for acute myeloid leukemia, diffuse large B-cell lymphoma, and possibly follicular lymphoma. The most important aspect of this study is in demonstrating the feasibility in using large census- and registry-based data to study rare diseases and disease subtypes in the population.

Chapters 6 and 7 estimated the lung cancer disease risks associated with occupational exposures to respirable crystalline silica and diesel engine exhaust, respectively. These two studies were both performed in a population that was obtained from combining subjects from 14 different population-based case-controls studies from Europe and Canada with 37,866 subjects. Both chapters are innovative in terms of assessing exposure for a large study population in a fully quantitative manner (i.e. estimating exposure in numeric values instead of categorical groups such as low, medium and high exposure). For silica, because historical exposure measurements were abundant, a model was created to enable estimating work exposures for different jobs in distinct regions and time periods. For diesel exhaust, available exposure data for diesel exhaust was more limited and quantitative exposure assessment was only possible by occupation. Chapter 6 reported consistent increases in lung cancer risks in relation to increasing work exposure to silica. Similarly, Chapter 7 found that increasing work exposure to diesel engine exhaust was associated with increasing lung cancer risk in male subjects. In both studies, the observed exposure-disease associations were consistent in subjects who were smokers as well as in those who never smoked.

Chapter 8 reviewed the main findings from each content chapters and discussed potential ways of improving occupational exposure assessment and disease risk analysis in population-based studies. The work presented in this thesis represents continual progress to improve our methods and approaches for estimating work exposures and their associated disease risks. Grounded in (and indebted to) earlier works, my work points towards a few directions for future progress: 1) Automation of the job coding process holds promise for improving efficiency and reliability, and may potentially create improved job classifications systems for exposure assessment; 2) Quantitative exposure assessment is possible in various study settings and for different exposures, and doing so elevates the utility and impact of occupational health studies; 3) Due to its large scale and efficiency, population-based studies are a unique testing ground for new ideas in occupational health research, such as accounting for multiple exposures and various temporal exposure characteristics.

NEDERLANDSE SAMENVATTING

Arbeidsgezondheidsonderzoek bij de bevolking

Voor de meeste mensen vormen blootstellingen op de werkplek een groot deel van hun totale blootstelling aan omgevingsfactoren. Over het algemeen worden twee soorten onderzoeksdesigns gebruikt om te onderzoeken hoe blootstelling op het werk de gezondheid beïnvloedt: industriespecifieke studies en algemene populatie studies. Zoals de naam al aangeeft, kijken we in industriespecifieke studies naar deelnemers uit een specifieke bedrijfstak of beroep (bijv. luchtvaartpersoneel, piloten). Omdat deelnemers in industriespecifieke studies doorgaans in vergelijkbare omgevingen werken, is het relatief eenvoudig om beroepsmatige blootstelling te meten of in te schatten. Industriespecifieke studies zijn vrijwel de enige haalbare manier om zeldzame blootstellingen te bestuderen zeker als deze blootstellingen alleen voorkomen op een beperkt aantal specifieke werkplekken. Blootstelling aan vliegtuigbrandstof komt bijvoorbeeld alleen voor bij degenen die werken met deze brandstoffen en vliegtuigmotoren. De omvang van industriespecifieke studies wordt echter beperkt door de omvang van de industrie. Stel we hebben een studie waarbij we met succes alle huidige en vroegere straaljagerpiloten rekruteren, dan nog zal de omvang van de studie vrij klein zijn omdat er nu eenmaal niet zoveel straaljagerpiloten zijn. Een studie met enkele tientallen of honderd straaljagerpiloten zal niet erg nuttig zijn voor het bestuderen van zeldzame ziekten zoals kanker en neurodegeneratieve aandoeningen. Industriespecifieke studies leggen ook vooral nadruk op blootstellingen binnen die specifieke sector - potentieel belangrijke aanvullende factoren zoals blootstelling in eerdere beroepen en persoonlijke leefstijlgewoonten zijn niet altijd beschikbaar.

Bij algemene populatie studies komen de deelnemers uit de algemene bevolking. Het werven van deelnemers uit de algemene bevolking heeft een aantal voordelen die de tekortkomingen van industriespecifieke studies kunnen aanvullen. Eén van de belangrijkste voordelen is dat de omvang van algemene populatie studies in theorie alleen wordt beperkt door de omvang van de algemene bevolking. Hoewel praktische beperkingen van alle wetenschappelijke studies (bijv. financiering, tijd) nog steeds van toepassing zijn, hebben zulke studies groottes tot tienduizenden en zelfs miljoenen deelnemers kunnen bereiken. Zo'n grote steekproefomvang stelt ons in staat om zeldzame ziektebeelden of subtypes hiervan te bestuderen. In vergelijking met de industriespecifieke studies, wordt in algemene populatie studies doorgaans ook meer informatie over de totale beroepsgeschiedenis en leefstijlfactoren verzameld. Deze aanvullende informatie helpt om de relaties tussen blootstelling op het werk en ziekte beter te begrijpen.

Om te onderzoeken of blootstelling X verband houdt met ziekte Y, moet een onderzoek deelnemers hebben die wel en niet zijn blootgesteld aan X. Bovendien moet een aantal van de deelnemers ziekte Y hebben of krijgen (de patiënten) waarbij de rest dan als controlepersoon dient. Op basis van hoe het onderzoek is opgezet, kunnen populatie studies worden omschreven als prospectief of retrospectief. Een prospectieve studie werft deelnemers voordat de ziekte zich voordoet, volgt de deelnemers totdat sommigen ziek worden, en vergelijkt vervolgens de blootstellingen die optraden bij patiënten versus controlepersonen. Een retrospectieve studie begint met deelnemers waarvan bekend is of ze de ziekte hebben of niet en vergelijkt vervolgens de eerdere blootstellingen tussen deze twee groepen. Er zijn veel voor- en nadelen aan zowel de prospectieve als de retrospectieve onderzoeksopzet. Het meest voor de hand liggende is het verschil in efficiëntie. Stel we hebben 200 patiënten met ziekte Y nodig om een bepaalde onderzoeksvraag te kunnen beantwoorden. Het zou dan veel sneller zijn om 200 patiënten en 200 controlepersonen te vinden, dan om een grote groep mensen samen te stellen en te wachten tot 200 mensen ziekte Y ontwikkelen. Dit geldt met name als Y relatief zeldzaam is en het lang duurt voordat het zich openbaart, zoals bijvoorbeeld alvleesklierkanker. Bij retrospectieve studies is echter niet alles eenvoudiger. In een prospectieve studie kan er actuele en nauwkeurigere beroepsinformatie worden verkregen door herhaalde interviews die worden afgenomen terwijl we wachten op voldoende ziektegevallen. Bovendien kunnen we deze informatie gebruiken om de beroepsmatige blootstelling van deelnemers beter te schatten en te beschrijven (we noemen dit "karakterisering van blootstelling"). In retrospectieve studies, vooral voor chronische ziekten, kunnen er in het verleden decennialang belangrijke beroepsmatige blootstellingen hebben plaatsgevonden die verband houden met ziekte. Dit maakt het beoordelen van de blootstelling een uitdaging, omdat we in retrospectieve studies de beroepsmatige blootstellingen moeten schatten die decennia eerder plaatsvonden op basis van de volledige werkgeschiedenis die we hebben verkregen uit één enkel interview. Deelnemers kunnen zich taken die lang geleden zijn uitgevoerd veelal niet nauwkeurig herinneren. Sommige werkplekken en taken zijn mogelijk zelfs verdwenen over de tijd.

Dit proefschrift

Het thema van mijn proefschrift draait om de uitdagingen in retrospectieve algemene populatie onderzoeken op het gebied van arbeid en gezondheid: hoe kunnen we een betere werkgeschiedenis verkrijgen van deelnemers, hoe kunnen we beter de blootstelling inschatten met deze informatie, hoe kunnen we beter gezondheidsrisico's berekenen met de geschatte blootstellingen, en hoe kunnen we algemene populatie studies effectiever inzetten om de relatie tussen blootstelling en ziekte te begrijpen en de gezondheid van werknemers en de algemene volksgezondheid te verbeteren.

Na Hoofdstuk 1 als een algemene inleiding van het proefschrift, bespreek ik in Hoofdstuk 2 de methoden en benaderingen die gebruikt zijn bij het bepalen van de blootstelling in retrospectieve populatie studies naar beroepsgerelateerde kanker. Dit literatuuroverzicht is belangrijk omdat het de eerste is die systematisch de methoden voor blootstellingsbeoordeling evalueert die de afgelopen decennia zijn gebruikt in populatieonderzoeken. Bovendien vormen de bevindingen ervan de basis voor verdere verkenning van mogelijke innovaties in blootstellingsbeoordeling. Uit het literatuuroverzicht blijkt dat het merendeel van de huidige algemene populatie onderzoeken nog steeds gebaseerd is op methoden voor de blootstellingsbeoordeling die tientallen jaren eerder zijn ontwikkeld. Verder geven we potentiële nieuwe ontwikkelingsgebieden aan, inclusief hoe: a) betrouwbaardere beroepsinformatie van deelnemers kan worden verkregen; b) beter gebruik gemaakt kan worden van blootstellingsmetingen; en c) de expertbeoordeling toegepast kan worden met meer transparantie, efficiëntie en consistentie.

In hoofdstuk 3 evalueer ik de prestaties van op computer-algoritmen gebaseerde software voor interviews met beroepsgeschiedenis in algemene populatie studies. Het algoritme werd toegepast in een patiënt-controle onderzoek naar lymfomen en leukemie in Azië (de AsiaLymph-studie) met 11.409 gerapporteerde banen. Op basis van verschillende functietitels en functietaken die door deelnemers worden gerapporteerd, wijst het computersysteem automatisch (en in real-time) verschillende interviewvragen toe. Zo kan relevante informatie over de beroepsmatige blootstelling beter worden vastgelegd. Vergeleken met de menselijke deskundige selecties, selecteerde het computeralgoritme identieke vragen voor 57,7% van de gerapporteerde banen, en 86% van de door de computer geselecteerde vragen zou dezelfde hoeveelheid blootstellingsrelevante informatie opleveren. Het onderzoek toonde aan dat de computeralgoritmen op betrouwbare wijze interviews voor de beroepsgeschiedenis automatisch in real-time kunnen toewijzen en dat vergelijkbare tools de kwaliteit kunnen verbeteren van de beroepsinformatie die van de deelnemers wordt verzameld.

Hoofdstuk 4 richtte zich op het proces van het indelen van door deelnemers gerapporteerde banen in gestandaardiseerde groepen, ook wel bekend als functiecodering, wat cruciaal is voor de blootstellingsbeoordeling. In dit hoofdstuk is het verschil onderzocht tussen het toewijzen van taakcoderingen voor dezelfde taken die werden uitgevoerd door verschillende deskundige codeerders, en hoe dit verschil latere blootstellingsbeoordeling beïnvloedde. Dit werk werd ook uitgevoerd binnen de AsiaLymph-studie en omvatte 34.353 banen gerapporteerd door 12.590 proefpersonen. De overeenkomst in de functiecodering varieerde van 51,0-77,1% tussen verschillende codeurs, afhankelijk van de verschillende codeerkoppels en de mate van detail in de functiecode. Uit de studie bleek ook dat deze verschillen in functiecodering in het algemeen kleiner worden in de vervolgstappen van de blootstellingsbeoordeling, omdat niet alle verschillen in functiecode leiden tot verschillende blootstellingsschattingen. Hoofdstuk 4 vertegenwoordigt één van de grootste evaluatiestudies voor functiecodering die er tot nu toe zijn uitgevoerd, waarbij de taakcodering onafhankelijk werd uitgevoerd door twee teams en beoordeeld door een derde expert, en waar vervolgens de verschillen in het toegewezen blootstellingsniveau zijn beoordeeld.

In Hoofdstuk 5 is het mortaliteitsrisico van lymfohematopoëtische kankers gerelateerd aan beroepsmatige blootstelling aan benzeen in een algemene populatie cohort van 2,97 miljoen Zwitserse burgers geschat. Uit de studie bleek dat beroepsmatige blootstelling aan benzeen gepaard gaat met een verhoogd sterfterisico van acute myeloïde leukemie, diffuus grootcellig B-cellymfoom en mogelijk folliculair lymfoom. Het belangrijkste aspect van deze studie is het aantonen van de haalbaarheid van het gebruik van grote census- en registratiegegevens om zeldzame ziekten en ziektesubtypen in de populatie te bestuderen.

In Hoofdstuk 6 en 7 is een schatting gemaakt van de risico's van longkanker in relatie met beroepsmatige blootstelling aan respectievelijk silica en dieselmotor-emissies. Deze onderzoeken werden beide uitgevoerd in een populatie die werd verkregen door proefpersonen uit 14 longkanker patiënt-controle onderzoeken uit Europa en Canada te combineren, met in totaal 37.866 deelnemers. Beide studies zijn innovatief in termen van het volledig kwantitatief beoordelen van de blootstelling voor een grote onderzoekspopulatie (d.w.z. het schatten van de blootstelling in numerieke waarden in plaats van categorische groepen zoals lage, gemiddelde en hoge blootstelling). Voor silica, omdat er veel historische blootstellingsmetingen beschikbaar waren, werd een model gemaakt om de beroepsmatige blootstelling voor verschillende banen in verschillende regio's en tijdsperioden te schatten. Voor dieseluitlaatgassen waren de beschikbare blootstellingsgegevens beperkter en was een kwantitatieve blootstellingsbeoordeling alleen mogelijk op basis van beroep. Hoofdstuk 6 rapporteert een consistente toename van het risico op longkanker in relatie tot blootstelling aan silica op het werk. Op dezelfde manier werd in hoofdstuk 7 vastgesteld dat een toenemende blootstelling aan dieselmotoremissies in verband staat met een toename van longkanker bij mannen. In beide studies waren de waargenomen blootstellings-ziekte associaties consistent tussen rokers en niet-rokers.

Hoofdstuk 8 bespreek ik de belangrijkste bevindingen van elk hoofdstuk en mogelijke manieren om de karakterisering van beroepsmatige blootstelling en gezondheids-risicoanalyse in populatie studies te verbeteren. Het werk dat in dit proefschrift wordt beschreven, draagt bij aan de voortdurende vooruitgang in het verbeteren van onze methoden en benaderingen voor het schatten van werkblootstellingen en de bijbehorende gezondheidsrisico's. Gebaseerd op (en met dank aan) eerder werk, geeft mijn werk een aantal aanwijzingen voor toekomstige verbeteringen: 1) Automatisering van het beroepscoderingsproces is veelbelovend voor het verbeteren van de efficiëntie en betrouwbaarheid in algemene populatie studies, en kan mogelijk verbeterde beroepsclassificatiesystemen creëren voor de blootstellingsbeoordeling; 2) Kwantitatieve blootstellingsbeoordeling is mogelijk in verschillende onderzoeks-settings en voor verschillende blootstellingen, en dit vergroot het nut en de impact van arbeidsgezondheidsonderzoeken; 3) Vanwege de grote schaal en efficiëntie zijn algemene populatie studies een unieke proeftuin voor nieuwe ideeën in arbeidsgezondheidsonderzoek, zoals het bestuderen van het effect van meervoudige blootstellingen en verschil in chronologie van blootstellingskenmerken in risicoanalyses.

Acknowledgement

Coming to the Netherlands for a PhD is one of my best life decisions. It has been an incredible journey. I owe everything to those around me for my personal and professional growth.

Roel: When we sat down for one of our first conversations at IRAS you said, "Doing a PhD is a bit like swimming across a huge lake. At times, often in the middle, you may feel lost and far away from everything. But if you keep at it, you will eventually reach the other side." I remember making a distinct mental note, stashing away the quote for the seemingly inevitable day I become lost. That day never came. This foreshadowing remains unfulfilled largely thanks to you. Under your guidance, I always felt that I was propelling myself, but never doing it alone. I am most grateful for your trust and the freedom I am given to explore ideas and make mistakes. Had no idea what I was getting myself into when I interrupted the conversation between you and Melissa at EPICOH 2014, but I am glad I did. I could not have asked for a better teacher, boss, and mentor – your passions for knowledge, excellence, and social justice are contagious.

Susan: We only started working more closely together in the past couple years, but you have made a big impact on all chapters in this thesis. In fact, your PhD thesis seldom left my desktop after Roel handed it to me soon after I started. The only occasions it did leave my desk were times when I brought it home with me as a must-have reference. (In retrospect I am not sure why I didn't just download a copy.) Much of my work would not have been possible without the foundation you built in exposure assessment, in SYNERGY and SYN-JEM, your attention to detail, and your willingness to repeatedly review (and translate) my writing. Thank you. I look forward to continue working with you in EPHOR and other projects.

Hans: My teacher's teacher. Your door is always open for me to ask quick questions and those quick questions sometimes turn into long discussions. Your experience, tips, and perspective are always appreciated. Your work on occupational exposure assessment, JEMs, SYNERGY yielded many fruits and we are grateful for the bounty. Thank you for all the input on many papers in this thesis, particularly the literature review. It was a challenging task and you really helped straightened the course.

Lützen: Usually I am quite happy if I exit your office with >70% comprehension of our conversation. The remainder would then click into place days or months later, if at all. It has been a privilege to learn from you in both statistics and programming. Thank you for patiently guiding a relative novice. To me you have the best job at IRAS. I hope to be one day as cool as you and as helpful to others.

Nat and Qing: Thank you for the incredible work in AsiaLymph, which funded this very PhD position. Traveling and working with you in Asia are highlights of my PhD, and they were eye-opening in terms of scientific collaboration on a multi-continental and multi-cultural scale.

Melissa: Thank you for your guidance throughout my journey, for work both within and beyond AsiaLymph. Although you are thousands of kilometres away, when I am in need you always come to my aid swiftly as if you were just down the hall.

319b: I may be biased, but we have the most awesome office with the coolest gang. Aileen, Amine, Anran, Daniel, Edith, Ilse, Joseph, Jules and Marije, you have been great colleagues at work and awesome friends at large. Thank you for putting up with the daily coffee grinding and indulging me on the coffee club.

Ingrid, Christina, Petra: Thank you for all your help since day 0, even before I set foot in the Netherlands for this position. Your institutional knowledge has been invaluable for myself and countless other students and staff. I appreciate all of you.

He Nan: The fact that you made it on the acknowledgements on my thesis may seem like a bit of an irony: meeting you most definitely limited my long nights in the office. But this booklet would not have been the same without you, from the cover design to the author. Thank you for your companionship, encouragement, and positivity. Utrecht is home because of you.

To my family: None of my achievements could have been done without your hard work, sacrifices, and unwavering support. Thank you. I love you.

Curriculum Vitae

Calvin was born in Shenzhen, China on December 22, 1982. After graduating from the University of British Columbia (UBC) majoring in microbiology and immunology in 2006, Calvin joined a private veterinarian practice specialised in bovine embryo transfer services. In 2011, Calvin graduated from the master program at the School of Population and Public Health in UBC. From 2010 to 2014, Calvin worked as an occupational hygienist at CAREX Canada, a research project investigating the prevalence and level of carcinogen exposures among Canadians. In 2015, Calvin started pursuing a doctorate degree at the Institute for Risk Assessment Sciences (IRAS) at Utrecht University, producing the works presented in this thesis. Since 2020, Calvin has started his postdoctoral research at IRAS on the Exposome Project for Health and Occupational Research (EPHOR) with a focus on developing and improving methods for exposure assessment, epidemiology, and impact assessment.

Utrecht University
Institute for Risk Assessment Sciences