Annals of Work Exposures and Health, 2018, Vol. 62, No. 9, 1047–1063
doi: 10.1093/annweh/wxy080
Advance Access publication 18 September 2018
Review

Review

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

Calvin B. Ge[1*], Melissa C. Friesen[2], Hans Kromhout[1], Susan Peters[1,3], Nathaniel Rothman[2], Qing Lan[2] and Roel 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]Department of Neurology, University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands; [4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

*Author to whom correspondence should be addressed. E-mail: C.B.Ge@uu.nl

Submitted 18 May 2018; revised 7 August 2018; editorial decision 8 August 2018; revised version accepted 22 August 2018.

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.

© The Author(s) 2018. Published by Oxford University Press on behalf of the British Occupational Hygiene Society.

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.

Keywords: cancer epidemiology; case–control; expert judgement; exposure assessment; job-exposure matrix; reproducibility; self-reported exposure

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, available at *Annals of Work Exposures and Health* online). 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, available at *Annals of Work Exposures and Health* online).

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, available at *Annals of Work Exposures and Health* online). 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.

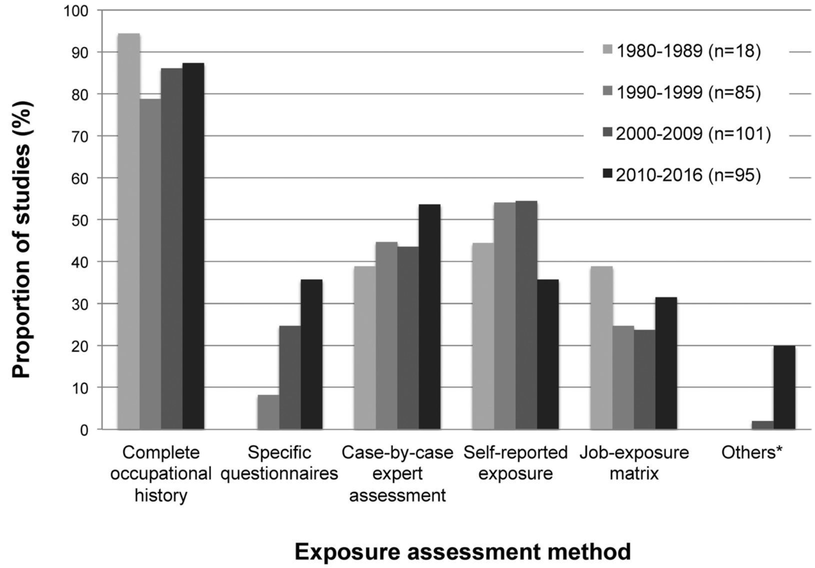

Figure 1. Use of various retrospective occupational exposure assessment methods in general population case–control occupational cancer studies (*Others: includes methods that are distinct from other major assessment methods, such as exposure assessment using expert-derived algorithms, measurement calibrated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

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.

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).

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 κ agreement for presence/absence of exposure between different expert assessors

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)

TBQ = task-based questionnaire.

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 κ 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 κ 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 κ values ranging from 0.06 to 0.30, with a median of 0.19. Adegoke *et al*. (2004), however, reported better agreement (κ 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 κ 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 κ 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 κ agreement between JEMs ranged from 0.25 to 0.51, with a median of ~0.46. In a multicentre European lung cancer study, Peters *et al*. (2011a) reported κ 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 κ 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 κ 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 κ 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).

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.

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 check-list 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

Table 4. *Continued*

Authors, year	Exposure	Assessment method	Comparison method	Reliability test	Results
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.

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, available at *Annals of Work Exposures and Health* online). 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.

Supplementary Data

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

Funding

Funding for this project was provided by the Intramural Research Program from the US National Cancer Institute to M.C.F., Q.L., and N.R.

Acknowledgements

M.C.F., Q.L., and N.R. acknowledge the support provided by the Intramural Research Program of the US National Cancer Institute. The authors declare no conflict of interest relating to the material presented in this article.

References

Adegoke OJ, Blair A, Ou Shu X *et al*. (2004) Agreement of job-exposure matrix (JEM) assessed exposure and self-reported exposure among adult leukemia patients and controls in Shanghai. *Am J Ind Med*; **45**: 281–8.

Armstrong BG. (1990) The effects of measurement errors on relative risk regressions. *Am J Epidemiol*; **132**: 1176–84.

Bhatti P, Newcomer L, Onstad L *et al*. (2011) Wood dust exposure and risk of lung cancer. *Occup Environ Med*; **68**: 599–604.

Black J, Benke G, Smith K *et al*. (2004) Artificial neural networks and job-specific modules to assess occupational exposure. *Ann Occup Hyg*; **48**: 595–600.

Bourgkard E, Wild P, Gonzalez M *et al*. (2013) 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*; **70**: 884–91.

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. *Int J Epidemiol*; **22** (Suppl. 2): S57–64.

Bouyer J, Hémon D. (1993b) Studying the performance of a job exposure matrix. *Int J Epidemiol*; **22** (Suppl. 2): S65–71.

Carey RN, Driscoll TR, Peters S *et al*. (2014) Estimated prevalence of exposure to occupational carcinogens in Australia (2011-2012). *Occup Environ Med*; **71**: 55–62.

Colt JS, Karagas MR, Schwenn M *et al*. (2011) Occupation and bladder cancer in a population-based case-control study in Northern New England. *Occup Environ Med*; **68**: 239–49.

Correa A, Min YI, Stewart PA *et al*. (2006) Inter-rater agreement of assessed prenatal maternal occupational exposures to lead. *Birth Defects Res A Clin Mol Teratol*; **76**: 811–24.

Daniels JL, Olshan AF, Teschke K *et al*. (2001) Comparison of assessment methods for pesticide exposure in a case-control interview study. *Am J Epidemiol*; **153**: 1227–32.

DellaValle CT, Purdue MP, Ward MH *et al*. (2015) Validity of expert assigned retrospective estimates of occupational polychlorinated biphenyl exposure. *Ann Occup Hyg*; **59**: 609–15.

Duell EJ, Millikan RC, Savitz DA *et al*. (2001) Reproducibility of reported farming activities and pesticide use among breast cancer cases and controls. A comparison of two modes of data collection. *Ann Epidemiol*; **11**: 178–85.

Févotte J, Dananché B, Delabre L *et al*. (2011) Matgéné: a program to develop job-exposure matrices in the general population in France. *Ann Occup Hyg*; **55**: 865–78.

Friesen MC, Coble JB, Lu W *et al*. (2012) Combining a job-exposure matrix with exposure measurements to assess occupational exposure to benzene in a population cohort in shanghai, china. *Ann Occup Hyg*; **56**: 80–91.

Friesen MC, Lan Q, Ge C *et al*. (2016a) Evaluation of automatically assigned job-specific interview modules. *Ann Occup Hyg*; **60**: 885–99.

Friesen MC, Lavoué J, Teschke K *et al*. (2015a) Occupational exposure assessment in industry- and population-based epidemiologic studies. In Nieuwenhuijsen MJ, editor. *Exposure assessment in environmental epidemiology*. Oxford, England: Oxford University Press.

Friesen MC, Locke SJ, Tornow C *et al*. (2014) Systematically extracting metal- and solvent-related occupational information from free-text responses to lifetime occupational history questionnaires. *Ann Occup Hyg*; **58**: 612–24.

Friesen MC, Pronk A, Wheeler DC *et al*. (2013) 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*; **57**: 470–81.

Friesen MC, Shortreed SM, Wheeler DC *et al*. (2015b) 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*; **59**: 455–66.

Friesen MC, Wheeler DC, Vermeulen R *et al*. (2016b) Combining decision rules from classification tree models and expert assessment to estimate occupational exposure to diesel exhaust for a case-control study. *Ann Occup Hyg*; **60**: 467–78.

Fritschi L, Friesen MC, Glass D *et al*. (2009) OccIDEAS: retrospective occupational exposure assessment in community-based studies made easier. *J Environ Public Health*; **2009**: 957023.

Fritschi L, Glass DC, Tabrizi JS *et al*. (2007) Occupational risk factors for prostate cancer and benign prostatic hyperplasia: a case-control study in Western Australia. *Occup Environ Med*; **64**: 60–5.

Fritschi L, Nadon L, Benke G *et al*. (2003) Validation of expert assessment of occupational exposures. *Am J Ind Med*; **43**: 519–22.

Fritschi L, Siemiatycki J, Richardson L. (1996) Self-assessed versus expert-assessed occupational exposures. *Am J Epidemiol*; **144**: 521–7.

Gramond C, Rolland P, Lacourt A *et al*.; PNSM Study Group. (2012) Choice of rating method for assessing occupational asbestos exposure: study for compensation purposes in France. *Am J Ind Med*; **55**: 440–9.

Hardt JS, Vermeulen R, Peters S *et al*. (2014) A comparison of exposure assessment approaches: lung cancer and occupational asbestos exposure in a population-based case-control study. *Occup Environ Med*; **71**: 282–8.

Heid IM, Küchenhoff H, Miles J *et al*. (2004) Two dimensions of measurement error: classical and Berkson error in residential radon exposure assessment. *J Expo Anal Environ Epidemiol*; **14**: 365–77.

Hepworth SJ, Bolton A, Parslow RC *et al*. (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. *Occup Environ Med*; **63**: 267–72.

Hoar SK, Morrison AS, Cole P *et al*. (1980) An occupation and exposure linkage system for the study of occupational carcinogenesis. *J Occup Med*; **22**: 722–6.

International Labour Organization (ILO). (2010) ISCO-International Standard Classification of Occupations: brief history [WWW Document]. Available at http://www.ilo.org/public/english/bureau/stat/isco/intro2.htm. Accessed 20 July 2018.

Koh DH, Bhatti P, Coble JB *et al*. (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. *J Expo Sci Environ Epidemiol*; **24**: 9–16.

Kromhout H, Oostendorp Y, Heederik D *et al*. (1987) Agreement between qualitative exposure estimates and quantitative exposure measurements. *Am J Ind Med*; **12**: 551–62.

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.

Landis JR, Koch GG. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**: 159–74.

Lavoué J, Pintos J, Van Tongeren M *et al*. (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**: 465–71.

Lee DG, Lavoué J, Spinelli JJ *et al*. (2015) Statistical modeling of occupational exposure to polycyclic aromatic hydrocarbons using OSHA data. *J Occup Environ Hyg*; **12**: 729–42.

Lillienberg L, Zock JP, Kromhout H *et al*. (2008) A population-based study on welding exposures at work and respiratory symptoms. *Ann Occup Hyg*; **52**: 107–15.

Lynge E, Andersen A, Rylander L *et al*. (2006) Cancer in persons working in dry cleaning in the Nordic countries. *Environ Health Perspect*; **114**: 213–9.

Mannetje A, Fevotte J, Fletcher T *et al*. (2003) Assessing exposure misclassification by expert assessment in multicenter occupational studies. *Epidemiology*; **14**: 585–92.

Nam JM, Rice C, Gail MH. (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. *Am J Ind Med*; **47**: 443–50.

Neilson HK, Sass-Kortsak A, Lou WY *et al*. (2007) Personal factors influencing agreement between expert and self-reported assessments of an occupational exposure. *Chronic Dis Can*; **28**: 1–9.

Offermans NS, Vermeulen R, Burdorf A *et al*. (2012) Comparison of expert and job-exposure matrix-based retrospective exposure assessment of occupational carcinogens in The Netherlands Cohort Study. *Occup Environ Med*; **69**: 745–51.

Olsson AC, Vermeulen R, Schüz J *et al*. (2017) Exposure-response analyses of asbestos and lung cancer subtypes in a pooled analysis of case-control studies. *Epidemiology*; **28**: 288–99.

Orsi L, Monnereau A, Dananche B *et al*. (2010) Occupational exposure to organic solvents and lymphoid neoplasms in men: results of a French case-control study. *Occup Environ Med*; **67**: 664–72.

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

Parks CG, Cooper GS, Nylander-French LA *et al*. (2004) Comparing questionnaire-based methods to assess occupational silica exposure. *Epidemiology*; **15**: 433–41.

Peters S, Glass DC, Milne E *et al*.; Aus-ALL consortium. (2014) Rule-based exposure assessment versus case-by-case expert assessment using the same information in a community-based study. *Occup Environ Med*; **71**: 215–9.

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

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

Peters S, Vermeulen R, Portengen L *et al*. (2011b) Modelling of occupational respirable crystalline silica exposure for quantitative exposure assessment in community-based case-control studies. *J Environ Monit*; **13**: 3262–8.

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

Pronk A, Stewart PA, Coble JB *et al*. (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. *Occup Environ Med*; **69**: 752–8.

Richiardi L, Mirabelli D, Calisti R *et al*. (2006) Occupational exposure to diesel exhausts and risk for lung cancer in a population-based case-control study in Italy. *Ann Oncol*; **17**: 1842–7.

Rocheleau CM, Lawson CC, Waters MA *et al*. (2011) Inter-rater reliability of assessed prenatal maternal occupational exposures to solvents, polycyclic aromatic hydrocarbons, and heavy metals. *J Occup Environ Hyg*; **8**: 718–28.

Segnan N, Ponti A, Ronco G, *et al*. (1996) Comparison of methods for assessing the probability of exposure in metal plating, shoe and leather goods manufacture and vine growing. *Occup Hyg*; **3**: 199–208.

Semple SE, Dick F, Cherrie JW; Geoparkinson Study Group. (2004) Exposure assessment for a population-based case-control study combining a job-exposure matrix with interview data. *Scand J Work Environ Health*; **30**: 241–8.

Sieber WK Jr, Sundin DS, Frazier TM *et al*. (1991) Development, use, and availability of a job exposure matrix based on national occupational hazard survey data. *Am J Ind Med*; **20**: 163–74.

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. *Am J Epidemiol*; **130**: 1236–46.

Stewart PA, Lees PS, Francis M. (1996) Quantification of historical exposures in occupational cohort studies. *Scand J Work Environ Health*; **22**: 405–14.

Teschke K, Olshan AF, Daniels JL *et al*. (2002) Occupational exposure assessment in case-control studies: opportunities for improvement. *Occup Environ Med*; **59**: 575–93; discussion 594.

Teschke K, Smith JC, Olshan AF. (2000) Evidence of recall bias in volunteered vs. prompted responses about occupational exposures. *Am J Ind Med*; **38**: 385–8.

Tielemans E, Heederik D, Burdorf A *et al*. (1999) Assessment of occupational exposures in a general population: comparison of different methods. *Occup Environ Med*; **56**: 145–51.

Tinnerberg H, Björk J, Welinder H. (2001) Evaluation of occupational and leisure time exposure assessment in a population-based case control study on leukaemia. *Int Arch Occup Environ Health*; **74**: 533–40.

Tinnerberg H, Heikkilä P, Huici-Montagud A *et al*. (2003) Retrospective exposure assessment and quality control in an international multi-centre case-control study. *Ann Occup Hyg*; **47**: 37–47.

Vlaanderen J, Straif K, Pukkala E *et al*. (2013) Occupational exposure to trichloroethylene and perchloroethylene and the risk of lymphoma, liver, and kidney cancer in four Nordic countries. *Occup Environ Med*; **70**: 393–401.

de Vocht F, Zock JP, Kromhout H *et al*. (2005) Comparison of self-reported occupational exposure with a job exposure matrix in an international community-based study on asthma. *Am J Ind Med*; **47**: 434–42.

Westberg HB, Hardell LO, Malmqvist N *et al*. (2005) On the use of different measures of exposure-experiences from a case-control study on testicular cancer and PVC exposure. *J Occup Environ Hyg*; **2**: 351–6.

Wheeler DC, Archer KJ, Burstyn I *et al*. (2015) Comparison of ordinal and nominal classification trees to predict ordinal expert-based occupational exposure estimates in a case-control study. *Ann Occup Hyg*; **59**: 324–35.

Wheeler DC, Burstyn I, Vermeulen R *et al*. (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. *Occup Environ Med*; **70**: 203–10.