Université de Montréal

Occupational exposures and risk of lung cancer among women

Par
Mengting Xu

Département de médecine sociale et préventive
École de santé publique

Thèse présentée en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D)
en santé publique, option épidémiologie

Mai 2023

Université de Montréal

Département de Médecine Préventive et Sociale - École de santé publique

Cette thèse intitulée

Occupational exposures and risk of lung cancer among women

Présentée par

Mengting Xu

A été évaluée par un jury composé des personnes suivantes

Helen Trottier
Présidente-rapporteuse

Jack Siemiatycki
Directeur de recherche

Vikki Ho
Codirectrice de recherche

Stéphane Buteau
Membre du jury

Mieke Koehoorn
Examinatrice externene

Résumé

Contexte: Le cancer du poumon est la deuxième cause de décès par cancer chez les femmes mondialement. Peu d'études ont examiné les facteurs de risque professionnel potentiel de ce cancer chez celles-ci.

Objectif: Cette thèse vise à déterminer si certaines expositions professionnelles sont associées au risque de cancer du poumon chez les femmes. L'objectif principal est composé de trois sous-objectifs:

1) Étudier les associations entre les expositions professionnelles prévalentes telles qu'évaluées par des experts et le risque de cancer du poumon chez les femmes dans une étude cas-témoin Montréalaise.

2) Comparer la concordance des attributions d'exposition entre la matrice emploi-exposition canadienne (CANJEM) et l'évaluation d'experts pour les emplois occupés par des femmes.

3) Étudier les associations entre des expositions professionnelles prévalentes et le risque de cancer du poumon chez les femmes dans un ensemble de données internationales combinées de dix études cas-témoins de ce cancer, en utilisant CANJEM pour évaluer les expositions des femmes

Méthodes : Pour le sous-objectif 1, nous avons utilisé des modèles de régression logistique multivariée pour examiner les associations entre les expositions professionnelles prévalentes et le risque de cancer du poumon chez les femmes (361 cas et 521 témoins) dans l'étude de Montréal. Pour le sous-objectif 2, nous avons comparé la concordance des expositions entre CANJEM et les experts pour 69 expositions professionnelles en utilisant les données de la même étude montréalaise. Pour le sous-objectif 3, nous avons utilisé une approche méta-analytique pour examiner les associations entre des expositions professionnelles prévalentes et le risque de cancer du poumon chez les femmes (3040 cas et 4187 témoins) à partir de dix études cas-témoins menées en Europe, au Canada et en Nouvelle-Zélande. En l'absence de la disponibilité de données sur les expositions des femmes, nous avons appliqué CANJEM pour estimer les expositions professionnelles.

Résultats: Nous n'avons pas observé de risque accru de cancer du poumon chez les femmes exposées professionnellement à 22 agents prévalents évalués dans l'étude de Montréal. Nous avons constaté que la capacité de CANJEM à reproduire l'évaluation des expositions par des experts variait selon l'agent.

Compte tenu de ces résultats, nous avons sélectionné 15 agents susceptibles d'être évaluables à l'aide de CANJEM qui étaient également répandus dans l'ensemble de données internationales. Il n'y avait aucune association entre la plupart des agents examinés et le cancer pulmonaire; cependant, les expositions à la poussière métallique, aux composés de fer, à l'isopropanol et aux solvants organiques étaient associées à des risques légèrement élevés.

Conclusions: Cette thèse a estimé l'exposition professionnelle des femmes à un large éventail d'agents et a examiné leurs associations avec le cancer du poumon. Dans l'étude de Montréal, aucun des agents évalués par les experts n'était associé à ce cancer. CANJEM s'est avéré capable de reproduire des évaluations d'exposition similaires à celles des experts, bien que sa fiabilité dépende de l'agent. Les expositions professionnelles évaluées par CANJEM ont suggéré des risques plus élevés de cancer du poumon chez les femmes exposées à la poussière métallique, aux composés de fer, à l'isopropanol et aux solvants organiques.

Mots clés: cancer du poumon, femmes, expositions professionnelles, matrice emploi-exposition, évaluation des experts, métaux

Abstract

Background: Worldwide, lung cancer is the second leading cause of cancer death among women. Few studies have examined possible occupational risk factors for lung cancer in women.

Objective: This thesis aims to investigate whether selected occupational exposures are associated with lung cancer risk among women. The main objective consists of three sub-objectives:

1) To investigate associations between prevalent occupational exposures as assessed by experts and lung cancer risk among women in a Montreal lung cancer case-control study

2) To compare exposure assignment concordance between the Canadian Job-Exposure Matrix (CANJEM) and expert assessment for jobs held by women

3) To investigate associations between prevalent occupational exposures and lung cancer risk among women in a combined international dataset of ten lung cancer case-control studies, with exposure assessed by CANJEM.

Methods: For sub-objective 1, we used multivariate logistic regression models to examine the associations between prevalent occupational exposures as assessed by experts and lung cancer risk among women (361 cases and 521 controls) in the Montreal study. For sub-objective 2, we compared exposure assignment concordance between CANJEM and expert assessments for 69 occupational exposures using data from the same Montreal study. For sub-objective 3, we used meta-analysis to examine the associations between prevalent occupational exposures and lung cancer risk among women (3040 cases and 4187 controls) from ten case-control studies conducted in Europe, Canada, and New Zealand. In the absence of available expert-assessed exposures, we applied CANJEM to estimate occupational exposures.

Results: We did not observe a clearly increased risk of lung cancer among women occupationally exposed to 22 prevalent agents assessed by experts in the Montreal study. We found that CANJEM's ability to replicate expert assessment of exposures varied by agent and by specific configurations of CANJEM. Considering these findings, we selected 15 agents suitable to evaluate using CANJEM — which were also prevalent in the international dataset of ten studies — and examined their risks for lung cancer. There was no association between most agents examined in this analysis and lung cancer; however, exposures to metallic dust, iron compounds, isopropanol, and organic solvents were associated with suggestive higher risks.

Conclusions: This thesis estimated women's occupational exposure to a wide range of agents and examined their associations with lung cancer. In the Montreal study, none of the expert-assessed prevalent agents was associated with lung cancer. CANJEM was shown to be able to reproduce exposure assessments similar to those of the experts, although its reliability was agent dependent. Occupational exposures assessed by CANJEM in the international dataset of ten studies indicated slightly higher risks of lung cancer among women exposed to metallic dust, iron compounds, isopropanol, and organic solvents.

Key words: lung cancer, women, occupational exposures, job-exposure matrix, expert assessment, metals

Table of Contents

List of Tables

List of Figures

List of Abbreviations

95% CI	95% Confidence Interval
AdCa	Adenocarcinoma
BCME	Bis(chloromethyl)ether
CANJEM	Canadian Job-Exposure Matrix
CCDO	Canadian Classification and Dictionary of Occupations
CE	Cumulative Exposure
CMME	chloromethyl methyl ether
COPD	Chronic Obstructive Pulmonary Disease
CSI	Comprehensive Smoking Index
ECDF	Empirical Cumulative Distribution Function
ETS	Environmental Tobacco Smoke
FINJEM	Finnish Job-Exposure Matrix
GWAS	Genome-Wide Association Study
HPV	Human Papilloma Virus
IARC	International Agency for Research on Cancer
IRR	Incidence Rate Ratio
ISCO	International Standardized Classification
ISCO-68	International Standard Classification of Occupations, version 1968
JEM	Job-Exposure Matrix
K	Cohen's Kappa Coefficient
MAHs	Mononuclear Aromatic Hydrocarbons
NOC	National Occupational Classification
NOCCA	Nordic Occupational Cancer Study
NSCLC	Non-Small-Cell Lung Carcinoma
NZSEI	New Zealand Socio-Economic Index
OR	Odds Ratio
OSHA	Occupational Safety and Health Administration
PAHs	Polycyclic Aromatic Hydrocarbons

SCLC	Small-Cell Lung Carcinoma
SES	Socio-Economic Status
SIR	Standardized Incidence Ratio
SNPs	Single-nucleotide polymorphisms
SOC	Standardized Occupational Classification
SqCC	Squamous Cell Carcinoma
US	United States

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor Dr. Jack Siemiatycki and my co-supervisor Dr. Vikki Ho, for guiding me throughout my PhD journey. Your advice and support have been invaluable. The wisdom you share with me inspires me every day to do better.

Thank you to Lesley Richardson for your continued support and encouragement throughout the completion of my thesis. Your guidance and support enabled me to finish my thesis, and I hope I can leave a positive legacy for others like you have for me.

I am also grateful to all the members of our research team at the CRCHUM. A special thank you to Dr. Jérôme Lavoué for his assistance with CANJEM and for providing critical feedback on the manuscripts. Thanks also to Natasha Graham for her administrative support.

Thanks to all my friends and colleagues at the CRCHUM and ESPUM, I cherish the memories of those long hours spent studying, working, and partying!

The support and care my family provided me over the years have always been a source of inspiration and motivation for me. Special thanks to Romain, for your support and unconditional love during the countless highs and lows of my PhD journey. My lovely Mina, Lily, Dodo, and Jelly Bean spent hours sitting on my arms or walking over my keyboard as I typed out my thesis each day. Their moral support is greatly appreciated.

Lastly, I wish to thank the Fonds de Recherche du Québec - Santé (FRQS), the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST) and the Université de Montréal, who enabled my training and research through providing financial support.

This milestone of completing my thesis would not have been possible without all the help mentioned above. I thank each of them for playing a part in making this possible.

Chapter 1: Context of the thesis

1.1 General introduction

In 1898, Marie and Pierre Curie discovered radium, a chemical element that is often referred to at the time as “beautiful radium” because of its shimmering glow (1, 2). Not long after its discovery, radium was widely used as a glow-in-the-dark paint and there was a high demand in the U.S. during World War I to produce glowing watches and airplane instruments for U.S. soldiers. Radium companies preferred to hire women than men for detailed work such as watch dial painting due to their smaller hands compared to men’s. Young women working as watch dial painters, often called the radium girls, were exposed to a high level of radium due to constant close and direct contact with the agent, that they started to glow in the dark. Beautiful as it seemed, radium is also highly radioactive. Many workers started to fall ill within a few years of employment at the factories and bone cancer followed. More and more female workers died but the radium company dismissed the issue and attributed the deaths to the workers’ unhealthy lifestyles, it was not until the death of a male employee that the company finally hired a pathologist to investigate the link between excessive deaths and the dial painting occupation, and then it was concluded that radium was responsible for poisoning the factory workers. Due to a lack of proper worker’s health and safety protection, and compensation, the radium girls had to fight lengthy legal battles, even on their death beds, to demand occupational compensation. They finally succeeded in winning the case, and this has resulted in new safety standards being introduced at the workplace in the US to protect future watch dial painters, and the passing of the law which granted workers the rights to compensation for occupational illnesses. In addition, this legacy has ultimately led to the establishment of the Occupational Safety and Health Administration (OSHA) in the US (1, 2).

Fast forward to over half a century later, enormous progress has been made by researchers in occupational epidemiology and experimental carcinogenesis in identifying other occupational carcinogens. However almost all of the epidemiological work has been carried out among male workers. With the increasing participation of women in the workforce over the past 50 years, the paucity of research on occupational cancer among women represents a glaring absence. In this thesis, we will examine the role of selected common exposures among women at the workplace and their associated lung cancer risk.

1.2 Organization of this thesis

This thesis is comprised of 9 chapters. The first chapter provides a general introduction and context for the thesis. An overview of the literature on the state of knowledge of lung cancer in women, occupational and other risk factors of lung cancer in women, and various occupational exposure assessment methods commonly used in epidemiologic studies is presented in Chapter 2, while Chapter 3 presents the main thesis objective and its three specific sub-objectives. This thesis includes three separate manuscripts, each addressing one of the three specific sub-objectives. Chapter 4 lays out the methodology used in each manuscript, including data source; design and study population; measures for exposures, covariates, and outcome; and statistical analysis strategies. Chapter 5 provides a brief introduction to the three manuscripts and some complementary information. The three manuscripts are presented in Chapters 6, 7, and 8, separately. Finally, Chapter 9 includes a summary of the main findings of this thesis, a description of the thesis contributions to epidemiology and public health, a discussion of the strengths and limitations of the research, and finally, a general conclusion.

Chapter 2: Overview of the literature

2.1 Lung cancer in women

2.1.1 Histology of Lung cancer

Primary lung cancer is a malignant tumor that originates from epithelium, bronchioles, alveoli, or bronchial mucous glands (3) in the lung. A new classification of lung tumors was published by the World Health Organization in 2015, in which lung cancer is divided into two main subtypes: small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC) (4). NSCLC accounts for approximately 85% of all lung cancer cases in North America in 2019 (5) and is further classified into: adenocarcinoma, squamous cell carcinoma, adenosquamous carcinoma, large cell carcinoma, and large cell neuroendocrine carcinoma. Adenocarcinoma, which forms from glandular cells of bronchial mucosa, has now surpassed squamous cell carcinoma to become the most common lung cancer subtype and represents about 40% of all lung cancers (6). Adenocarcinomas comprise a larger fraction of all lung cancers among women than among men (7, 8). It is thought that the increasing incidence of adenocarcinoma in women seen in recent years is related to the preference in the types of cigarettes smoked by women (including filtered and low-tar cigarettes) in addition to possible environmental exposures and genetic predisposition (9). Adenocarcinoma is also the most common lung cancer subtype diagnosed among people who have never smoked (6).

2.1.2 Incidence and mortality trends in lung cancer in women

Worldwide, lung cancer is the third most diagnosed malignant cancer and the second leading cause of cancer death in women, after breast cancer (10-12). It has been estimated that in 2022, 15 000 Canadian women would have been diagnosed with lung and bronchus cancer, and that 10 100 Canadian women would have died of this cancer (13).

Historically, lung cancer deaths were much higher in men than in women. Over time, the gap between sex-specific lung cancer death rates began to diminish, largely due to the narrowing gap of prevalence of smoking between men and women (14). In Canada, the lung cancer mortality rate started to level off in the late 1980s in men and has been declining ever since; while this rate was still increasing in women until 2006, and has only started to decrease since then (10).

On a global scale, the highest female lung cancer incidence rates have been observed in North America, Europe, Australia/New Zealand and Eastern Asia (12). Lung cancer incidence and mortality rates among women continue to increase worldwide, especially in developing countries (12). Figure 1 below presents the lung cancer age-standardized incidence rates by sex and region as of 2020 (12) and Figure 2 presents the time trends of lung cancer incidence rates in women by histological type and country from 1973 to 2002 (9).

Figure 1. Lung cancer incidence age-standardized rates by sex and region.

Figure adapted from Sung et al., 2021 (12). Permission was granted by the licensed content publisher John Wiley & Sons (license number: 5597930767881).

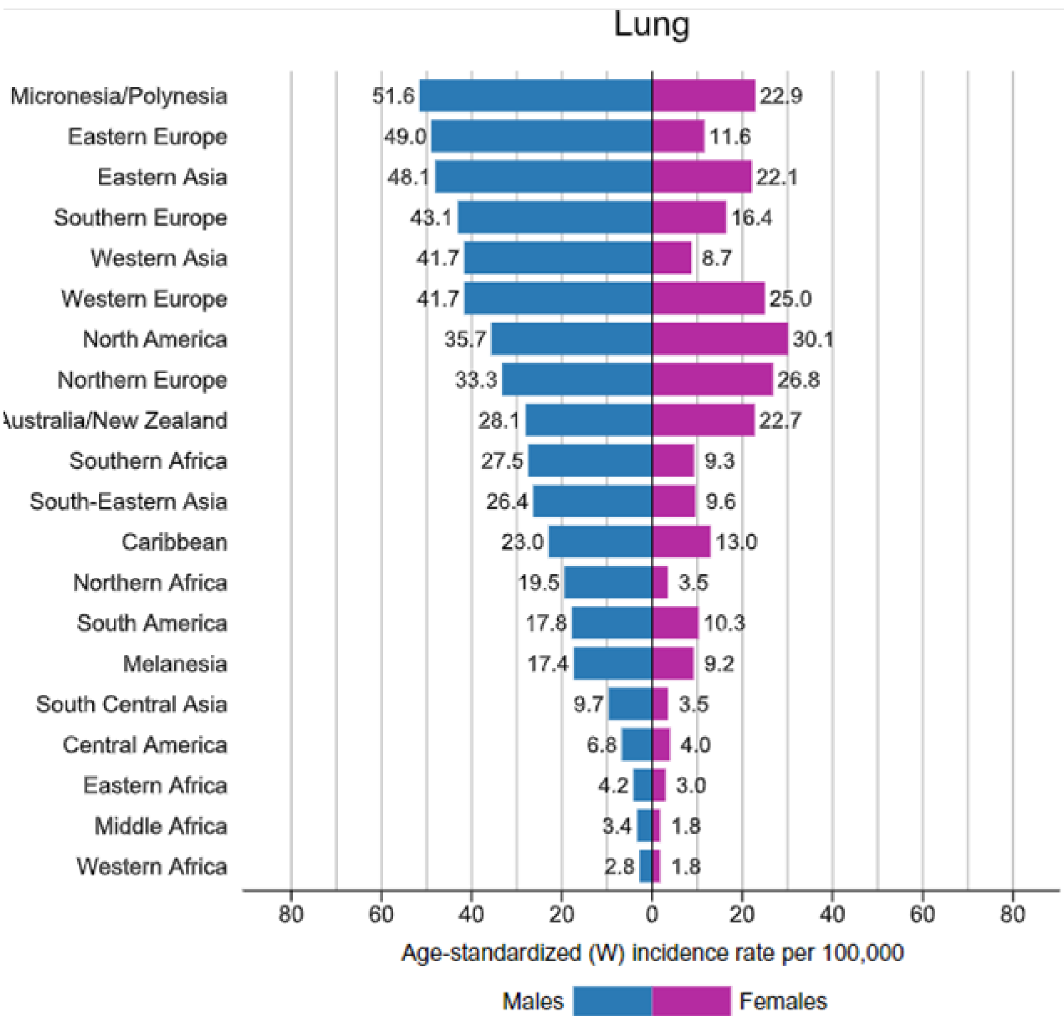
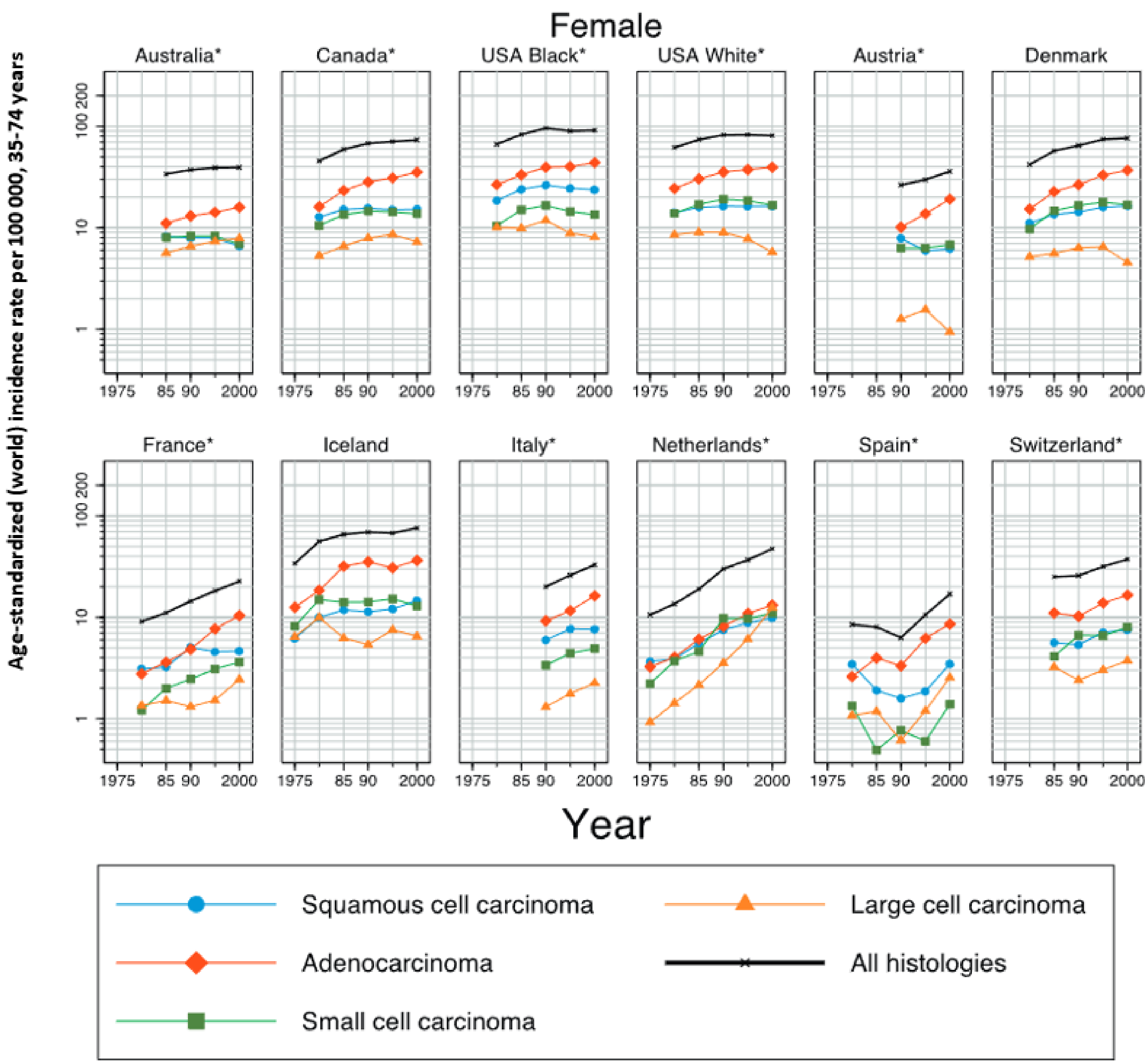


Figure 2. Lung cancer age-adjusted (world standard) incidence rates in women over time, by population and histological subtype, for ages 35–74.

Figure modified from Lortet-Tieulent et al. 2014 *(9)*. Permission was granted by the licensed content publisher Elsevier (license number: 5597940328156).



2.1.3 Lung cancer treatment and survival

Despite the advancement in our knowledge of lung cancer at the genetic and molecular levels, the prognosis for lung cancer survival remains very poor. Although women tend to have a higher chance of survival than men, the overall survival for lung cancer cases remains, nonetheless, very low in both sexes, with the five-year survival rate being 22% in women and 15% in men in Canada (10). Survival is better when lung cancer is diagnosed at an earlier stage. Among lung cancer patients, nonsmokers have better survival than smokers, possibly as a result of better response to therapy (15).

Treatment for lung cancer is generally ineffective at late stages and, unfortunately, most lung cancers are only diagnosed in later stages, due in part to the lack of effective screening techniques (16). In addition to cancer stage at diagnosis, treatment for lung cancer also depends on the histological subtype diagnosed. In general, Stage IA NSCLC only requires surgical re-section, stages IB to IIIA require surgical resection followed by adjuvant chemotherapy, and stage IV is treated with chemotherapy, and surgical resection in the presence of a solitary metastasis or for symptom control. SCLC is treated with platinum and etoposisde, either alone or combined with radiation therapy (11).

2.2 Non-occupational Risk factors for lung cancer in women

2.2.1 Tobacco smoking

Tobacco smoking is the most important risk factor for lung cancer among both women and men; with an observed relative risk in the order of 10 for all smokers and 20 to 50 among lifelong heavy smokers, when compared to never smokers (17). The strength of the association between smoking and lung cancer differs by histological subtype; stronger associations (i.e. higher relative risks) have been observed for squamous cell carcinoma and small cell carcinoma than for adenocarcinoma and large cell carcinoma (18, 19). Although risk of lung cancer starts to decline soon after quitting smoking, it does not fully return to the baseline lung cancer risk observed in never-smokers even decades after smoking cessation (19). In Canada, it was estimated that 86% of lung cancer cases can be prevented and that 72% are attributable to tobacco smoking (10).

In the past several decades, parallel to a drastic reduction in smoking, mainly in developed countries and among male smokers, there was a rapid reduction in lung cancer risk for small cell and squamous cell carcinomas — two lung cancer subtypes strongly associated with smoking; however, the reduction in adenocarcinoma risk was less rapid (20). Women are more likely to develop adenocarcinoma than men; the reason underlying this sex difference is not fully understood (21). Over the past decades, potentially due to a rising number of female smokers on a global scale and other occupational and environmental risk factors, the incidence of adenocarcinoma has increased dramatically and has since replaced squamous cell carcinoma as the most prevalent histological subtype of lung cancer in both sexes (22, 23).

Although tobacco smoking is one of the most important factors contributing to lung cancer, only a minority of smokers develop lung cancer, and lung cancer can occur among never-smokers. This suggests that smoking is not the only cause of lung cancer.

2.2.2 Environmental, genetic, hormonal, and other risk factors for lung cancer in women

There is evidence that some environmental, genetic, and hormonal factors, as well as viral infections may be associated with lung cancer in women.

Numerous environmental risk factors for lung cancer have been identified, including exposure to environmental tobacco smoke (ETS), asbestos, arsenic, radon, ionising radiation, outdoor air pollution, indoor emissions from household combustion of fuels including coal, gas, oil, and biomass (primarily wood) for heating and cooking, and emissions from high-temperature frying (17, 21, 22, 24, 25).

A history of lung cancer in the family appears to confer an increased level of risk for lung cancer, and it was reported that non-smoking women with a family history of lung cancer in a first degree relative have a greater risk of developing lung cancer when compared to non-smoking men with the same family history (15, 21, 26). In addition, genetic risk factors including the increased expression of the CYP1A1 gene, as well as the presence of the glutathione S-transferase M1 mutation, were also reported to be associated with carcinogenesis of the lung (21, 27). Genome-wide association studies (GWAS) have identified certain germline mutations differentially represented in lung cancer cases. Single-nucleotide polymorphisms (SNPs) at 6p21.33, 15p15.33, 15q15.2, 15q25.1, and 22q12 in European populations, and SNPs at 3q28, 5p15.33, 13q12.12, 17q24.3, and 22q12.2 in Asian populations, have been associated with elevated lung cancer risks (11).

Hormonal effects of estrogen have also been linked to tumorigenesis of the lung. *In vitro* studies have observed that estrogen promotes the growth of both healthy and malignant lung tissue and anti-estrogen treatments are able to suppress growth in some tumors (28). Female lung cancer risks associated with lifetime exposure to estrogen were examined in various publications with conflicting findings. Both an increased and a decreased risk of lung cancer has been observed in relation to age at menarche, age at menopause, parity, and use of hormone replacement therapy (21, 26, 27, 29-31).

There is evidence that lung cancer risk decreased with an increase in fruit consumption, irrespective of smoking status. In addition, among smokers, lung cancer risk also decreased with an increase in vegetable consumption (27, 32).

Human papilloma virus (HPV) infection was reported to be associated with an increased risk of lung cancer in studies of East Asian women. An increased expression of high-risk HPV haplotypes was found in lung tissue of women diagnosed with lung cancer (33). Two mechanisms were proposed to explain the presence of HPV in women's lung tissue. The first hypothesized mechanism posits that cervical infection of HPV leads to circulating virus, which then disseminates the HPV virus inside the lung tissue; and the second hypothesized mechanism posits that oral-genital sexual behaviour can lead to oral infection of the HPV virus followed by lung squamous cell infection (26, 34). HPV presence in the lung were much more prevalent in Asian populations versus in European, or North and South American populations. The reason for this large geographical heterogeneity in the presence of HPV DNA in lung tumor tissues is yet to be understood (35).

2.3 Occupational risk factors for lung cancer in women

Apart from smoking as the major risk factor for lung cancer, the most fruitful area for research on lung cancer risk factors has been the occupational environment. Since the mid-1900s, research has identified a large number of occupational lung carcinogens or occupations that lead to an excess risk of lung cancer (36). However, almost all the evidence was generated using data from industrial workforces in male-dominated occupations. Table 1 below presents a list of occupational exposures, occupations, industries, and manufacturing processes classified as Group 1 definite lung carcinogens by the International Agency for Research on Cancer (IARC).

Table 1. Occupational exposures, occupations, industries, and manufacturing processes classified as definite lung carcinogenic exposures (Group 1) by the IARC Monographs, Volumes 1–120 (37-39).

Agent
Ionizing Radiation
Bis(chloromethyl)ether (BCME) and technical-grade chloromethyl methyl ether (CMME)
Benzo[*a*]pyrene
Sulfur mustard
Coal-tar pitch
Soot
Diesel engine exhaust
Arsenic and Inorganic Arsenic Compounds
Beryllium and Beryllium Compounds
Cadmium and Cadmium Compounds
Chromium (VI) Compounds
Nickel Compounds
Asbestos (All Forms)
Silica Dust, Crystalline, in the Form of Quartz or Cristobalite
Tobacco smoke, secondhand
Hematite mining (underground)
Welding fumes
Occupations, industries, and manufacturing processes
Coal gasification
Coke production
Iron and steel founding
Aluminum production
Painting
Rubber production industry

There is little empirical evidence on occupational cancer risks incurred by women. The search for literature on occupational risk factors for lung cancer among women is also not straightforward, as most studies do not explicitly state in the title or the abstract whether their study population includes women. Among the limited studies identified that assessed risks specifically amongst women, a majority only examined occupational exposure at the job-title or industry levels, and lung cancer risk associated with specific occupational agents was rarely examined, often due to a lack of agent exposure assessment and low statistical power. In the following paragraphs, we present the estimated burden of lung cancer in women due to occupational exposures, and existing knowledge regarding occupations, industries and occupational agents associated with lung cancer in women. We also identify some gaps in knowledge of occupational risk factors for lung cancer in women in the literature.

2.3.1 Burden of lung cancer in women due to occupational exposures

Attributable fraction estimates are often used in epidemiology and public health to quantify the health burden of a disease. In 1981, as part of a landmark report estimating the proportional attributable fractions of US cancer mortality due to various environmental and lifestyle factors, Doll and Peto estimated that approximately 5% of lung cancer deaths in women could be attributed to occupational exposures to arsenic, asbestos, bischloromethyl ether, chromium, ionizing radiations, mustard gas, nickel, and polycyclic hydrocarbons in soot, tar and oil (40). Similar results were reported for occupationally-attributable lung cancer incidence risk among female workers in a study conducted in northern Germany in the 1990s (41). The occupational lung carcinogens selected in this study were based on a published list of jobs and industries with sufficient evidence for lung cancer as determined by the IARC at the time of the study. In 2003, as part of an assessment of the magnitude of US mortality due to selected causes of death associated with occupational exposures, Steenland et al. estimated a population attributable fraction of 2% for lung-cancer death in women due to occupational exposures to arsenic, asbestos, beryllium, cadmium, chromium, diesel fumes, nickel, silica, environmental tobacco smoke, and radon (42). A French study also concluded that lung cancer was the cancer site most impacted by ten recognized occupational carcinogens (asbestos, benzene, chromium VI, diesel engine exhaust, formaldehyde, nickel compounds, polycyclic aromatic hydrocarbons, silica dust, trichloroethylene, and wood dust), and accounted for 2.1% of all incident lung cancer cases among women in France in 2017 (43). These estimates likely underestimate the real burden of occupational risk factors for lung cancer in women, since risks were only estimated for a very limited number of known carcinogens.

2.3.2 Evidence of occupations and industries associated with lung cancer

Several occupational studies have examined associations between occupations and industries, and lung cancer risk among women. Depending on the country in which the studies were conducted, they were coded into different occupational and industrial classification systems. Some examples of common national occupational classifications include the Canadian Classification and Dictionary of Occupations (CCDO) version 1971 (44), the Canadian National Occupational Classification (NOC) version 2011 (45), and the United States Standardized Occupational Classification (SOC) version 2010 (46). To facilitate comparison of jobs across countries, many studies have also coded jobs into the International Standardized Classification (ISCO) version 1968 (47).

The Nordic Occupational Cancer (NOCCA) project, using data from a general population cohort of 15 million people in five Nordic countries with record linkage of census job titles and cancer data from national cancer registries, reported standardized incidence ratios (SIRs) for lung cancer by occupational category using 45-year follow-up data on cancer incidence (48). Individual information on smoking were not available in this study. The authors reported that, among women, occupations including engine operators, tobacco workers and "Other construction workers" were associated with the highest significant SIRs (SIRs > 2.00). Other occupations including painters, waitresses, beverage workers, transport workers, electrical workers, printers, welders, mechanics, packers, chemical process workers, drivers and glass makers also had significantly elevated SIRs (SIRs > 1.40). In addition, SIRs significantly above the null but below 1.40 were found for clerical workers, sales agents, shop workers, textile workers, smelting workers, wood workers, food workers, cooks and stewards, building caretakers, hair dressers, and launderers (48). On the other hand, occupations including farmers, gardeners, teachers (all three SIRs < 0.56), and nurses (SIR = 0.69) were found to be associated with a significantly below the null SIR in the NOCCA project. Economically inactive women also had a lower lung cancer risk than those in the workforce (48). To our knowledge, this is so far the largest study in terms of sample size and follow-up duration that reported associations between occupational job titles and lung cancer in women. This study used a modified version of the first large scale job exposure matrix, FINJEM, which had been modified by a team of hygienists to better reflect exposure to the 75 agents included in FINJEM as applied to the four participating countries. However, results of agent specific analyses have not yet been published. Another cohort study in Finland followed economically active Finns for lung cancer during 1971-95 (49). Participants' occupations were ascertained cross-sectionally using the 1970 census occupation data. In this study, elevated SIRs for lung cancer were found among women working in the transport industries. Considering that smoking data was only obtained between the mid- and end stages of the follow-up, the authors of this study stated that the smoking adjustment for women might be inaccurate. During the 1960s to 1970s, Finnish women's smoking habits changed from being a habit associated with high social class to being a habit associated with low education, implying that the smoking data gathered might not be an accurate proxy for determining women's prior smoking status, which is more etiologically relevant to tumor development (49).

Positive associations with lung cancer incidence were reported in two population-based case-control studies of women with long-term employment in the domestic service sector (50, 51); clerical-sales; service; and transportation-material handling occupations (50). A population-based case-control study in Germany reported positive associations between lung cancer incidence and ever employment in a variety of occupations including chemical processors and related workers; assemblers and unskilled metal workers; and stock clerks and related workers (52). Positive associations were also reported for ever employment in chemical and oil industries; pottery and glass; engine and vehicle building; paper, wood and print; and cleaning service, hairdressing, housekeeping or waste disposal (52). Lowered lung cancer incidence was found in women working in education, health, cultural, and sporting activity industries (52). Results from all these case-control studies were adjusted for smoking and other covariates.

In a study that examined occupational lung cancer mortality in US women who died between 1984 to 1998 in 27 States, the authors found significantly elevated proportionate mortality ratios for lung cancer among women working in manufacturing; transportation; retail trade; agriculture, forestry, and fishing; and nursing/personal care industries (53). The result of smoking adjusted analyses of high-risk occupations showed that women who were employed during their lifetime in industrial/blue collar and agricultural (including forestry and fishing) occupations were more likely to have died from lung cancer. In addition, technical, professional, administrative support, and managerial occupations were also associated with an increased proportionate lung cancer mortality among women in this study (53).

Summary of evidence on occupations / industries associated with lung cancer among women

There was some evidence to suggest that women working in industrial, manufacturing, transportation, and service occupations were at greater risk of developing lung cancer than women working in other occupations, whereas certain occupations in the health and teaching fields were associated with a lower risk of lung cancer.

Although the studies mentioned above provided some insights to identify potentially high-risk occupations or industries regarding lung cancer incidence or mortality among women, these studies are limited in their ability to identify specific exposures responsible for lung cancer because the occupational and industrial classifications are very general, and exposure within the same occupation or industry code may vary substantially. Confounding is another source of concern. Studies could suffer from confounding due to a lack of information on important confounders such as smoking (48), or from residual confounding due to inaccurate smoking adjustment (49). In addition, the use of proportionate mortality ratios is often criticized as a crude and potentially biased risk estimate, and thus has limited use to contribute to identifying causal occupational risk factors (54). This may partially explain why lung cancer is not consistently associated with any one occupation or industry across studies.

2.3.3 Evidence of occupational agents associated with lung cancer

Most occupational studies of lung cancer with data available for women only investigated lung cancer incidence or mortality associated with job titles or industry sectors (e.g., textile workers, dry-cleaners, and waiters/waitresses); moreover, not all studies provided female-specific risk estimates. Very few studies estimated the risk of lung cancer in women associated with specific occupational agent exposures. Among the limited number of studies that provided female-specific estimates for associations between occupational agent(s) and lung cancer, most included only one or very few agents and some used rather simplistic exposure assessment (i.e., self-reported exposure) and crude exposure parameterization (i.e., ever/never exposure). In the following paragraphs I summarize the epidemiological evidence concerning lung cancer risks of women exposed to selected agents for which there are multiple published studies. Agents not listed here may have evidence of carcinogenicity from studies of mainly male workers.

Textile dusts

Exposure to textile dusts is common among female workers as women comprise roughly half of the workforce employed in the textile manufacture or clothing production and sales industries, and are regularly exposed to various fabric dusts including natural textile materials such as cotton and wool and different sources of man-made synthetic textile dusts (55). While occupational exposure to textile dusts has been associated with certain non-neoplastic lung diseases including chronic obstructive pulmonary disease (COPD) and asthma (56), evidence regarding the effect of such dusts on lung cancer remains unclear. A meta-analysis examined occupational cotton dust exposure in women and reported a meta-RR of 0.77 (95%CI, 0.67-0.89) associated with lung cancer risk among women based on nine studies from Asia, North America, or Europe (57). Thus, it has been proposed that exposure to cotton dust is protective against lung cancer, and a mechanism has been proposed. Namely, it has been hypothesized that exposure to endotoxins, a class of pathogens that are found in organic textile products including cotton dust, may exert a protective effect against lung cancer (58, 59). Using data from a cohort of women textile workers in China, it was reported that textile worker's endotoxin exposure level was inversely associated with the risk of lung cancer after adjustment for age and smoking (59) and that occupational exposure to wool, silk, or synthetic fibre dusts, which do not contain cotton or endotoxin, is not associated with lung cancer (60). Increased lung cancer mortality was reported in a study of two US cohorts of female and male asbestos textile workers in North and South Carolina. Increased rates of lung cancer mortality were associated with cumulative chrysotile asbestos fibre exposure after model adjustment for age, sex, race, birth cohort and decade of follow-up (61).

Diesel and gasoline engine exhausts

IARC has concluded that there is *sufficient evidence* in humans that diesel engine exhaust causes cancer of the lung, and that there is *inadequate evidence* in humans for the carcinogenicity of gasoline engine exhaust (62). This conclusion is based largely on evidence from occupational studies of male workers. Among studies that reported risk estimates for female workers, null (63) or suggestively elevated (49) risks have been reported for Swedish or Finnish women occupationally exposed to diesel engine exhaust. Smoking was either unadjusted for or poorly adjusted for in these two studies. The same Finnish study also reported a positive association among women occupational exposed to gasoline engine exhaust but cautioned that this finding may be biased due to inaccurate smoking adjustment (49).

Formaldehyde

Formaldehyde has been classified as a group 1 carcinogen by IARC, based on sufficient evidence that it causes nasopharyngeal cancer and leukemia in humans; however, its association with lung cancer is unclear (64). Two case-control studies conducted in North America using either expert-assessed or self-reported occupational exposure to formaldehyde did not find an increase in lung cancer risk among exposed women (65, 66). Smoking was controlled for in both studies via either the inclusion of only non-smokers or model adjustment. One case-cohort study reported a statistically imprecise hazard ratio of 2.1 (95% CI 0.4 -11.0) after smoking adjustment among Chinese women with at least 10 years of exposure to formaldehyde working as textile workers (60).

Chlorinated solvents

A case-control study in France examined workers' occupational exposure to five chlorinated solvents (trichloroethylene, perchloroethylene, carbon tetrachloride, dichloromethane, and chloroform) using a job-exposure metrics named Matgéné, and observed a positive association between lung cancer and women ever exposed to perchloroethylene, a widely used chlorinated solvent, especially in the dry-cleaning sector. No associations were observed for the other chlorinated solvents (67). All models were adjusted for smoking and other confounders in this study.

Crystalline silica

Suggestive increased lung cancer risks among women associated with ever or over 10 years of occupational exposure to silica were reported in two studies, one conducted in China and another in multiple countries in Europe (60, 68), both studies have adjusted for smoking; however, the confidence intervals in these studies were very wide due to low statistical power.

Workplace Environmental tobacco smoke (ETS) in the workplace

Active tobacco smoking and exposure to environmental tobacco smoke (ETS) have long been proven as major risk factors for lung cancer in both women and men (17), and there is abundant literature showing an increased lung cancer risk among women who were exposed to tobacco smoke in the workplace. Most occupational studies assessed workplace ETS through self-report (65, 69-74). Studies conducted in North America, Europe, and Asia have all reported positive associations between women's ETS exposure in the workplace and lung cancer; a majority reported statistically significant increased lung cancer incidence or mortality by at least two-fold among women exposed to ETS in the workplace compared to unexposed women (65, 69-73). Higher lung cancer risks were observed among women with higher duration or intensity of exposure to ETS in the workplace (65, 69-74).

Summary of evidence on occupational agents associated with lung cancer among women

The evidence of lung carcinogenicity among women is quite strong for diesel engine exhaust and crystalline silica. There is weak evidence among women workers supporting the potential lung carcinogenicity of chlorinated solvents, and the potential protective effect of cotton dust. We did not find any clear evidence in previous studies in support of an increased risk among women occupationally exposed to gasoline engine exhaust or formaldehyde. There is strong evidence supporting that ETS is a risk factor for lung cancer in women, whether the exposure occurs in the workplace or elsewhere.

2.4 Occupational exposure assessment methods

Most epidemiologic research on occupational risk factors for cancer have either employed a retrospective industrial cohort design or a retrospective population-based case-control design. The methods of occupational exposure assessment employed in these studies largely depend on the study design (i.e., whether the study is of prospective or retrospective nature and whether it is industry-based or population-based) and the sample size of an epidemiologic study (i.e., it might be difficult to employ certain exposure assessment methods to a large study sample due to considerations of cost, practicality, or efficiency) (75). In prospective and cross-sectional studies, it is possible to directly measure some occupational exposures as they occur via biological monitoring or personal monitoring tools, however this is not an option for retrospective studies (unless such data have already been collected in the past, which is rarely the cases in most studies). In industry-based studies, which tend to employ an occupational cohort design or a nested case-control within an occupational cohort, workers' exposures are assessed based on their employment history within a particular occupational cohort, information on jobs held elsewhere is usually unavailable. Using this design, it might be possible to obtain past exposure measurements from historic surveillance data for a limited number of agents with high worker exposure prevalence. However, the incomplete coverage of a given worker's lifetime job history can be problematic for lifetime exposure assessment, particularly for workers in cohorts with high turnover rates, as this would lead to an underestimation of worker's lifetime cumulative exposure. In addition, it is often challenging to identify study subjects as the list of workers from company files might have been decades old, and thus many of the workers have retired, moved away, or died when the investigators tried to contact them. It is also difficult to collect information on important confounders such as one's lifetime smoking history and other factors. Health outcomes of investigation of such studies were often ascertained through linkages to databases such as the national mortality registry, which might not provide accurate cause of death. In population-based studies, which tend to recruit cases of a disease and controls from the general population residing in a specific geographic region, worker's lifetime occupational history and other relevant information including potential confounders can be collected through interviewing the participants, and exposures can be assessed using different indirect methods (more discussion on this in the next section); however, it is rarely feasible to obtain direct measurement of past occupational exposures using this study design (75).

The retrospective industrial cohort design is most appropriate when there is a clear hypothesis about a particular putative risk factor in a particular industry or occupation, while the retrospective population-based case-control design is most appropriate when the investigator wishes to explore a wide range of possible risk factors that can occur in different occupations or industries. A true prospective cohort design is rarely contemplated in this context because it implies the passage of many years or even decades before the research can reveal potential risks. The type of study we can contemplate for a study of occupational risk factors among women is the retrospective population-based case-control study design. Consequently, the sub-sections below will describe some of the issues regarding exposure assessment in that context.

2.4.1 Methods for occupational exposure assessment

In population-based case-control epidemiological studies, common methods used for retrospectively assessing occupational exposure include: subjects' self-report of their industry or job history; subjects' self-report of their occupational exposure to chemical or physical agents; expert-assessment of occupational exposures based on subject's self-reported job history; and exposure assessment using a job-exposure matrix (JEM) based on subject's self-reported job history (76). Self-reported exposure has been identified as the most frequently used exposure assessment method in a review by Ge et al. on assessment method trends from 1975 to 2016 in occupational cancer case-control studies in the general population; followed by expert-assessment and JEMs (77). Overall, the use of these methods has remained rather stable over the past four decades. Since 2010 and beyond, the use of self-reported exposure has declined and the use of expert-assessment and JEMs has increased; most studies assessed in the review by Ge et al. relied on one of the aforementioned methods to assess occupational exposure (77). Other exposure assessment methods, including modeling of exposure using historical measurement data, and using machine-learning algorithms to link self-reported questionnaire responses to expert-assigned exposure estimates, have also emerged in the past decade (77).

2.4.2 Validity of occupational exposure assessment methods

Retrospective assessment of occupational exposures is a major challenge in occupational epidemiology. This is particularly problematic in the context of population-based case-control studies of diseases with long latency and induction periods, where lifetime occupational exposure needs to be estimated for each subject to better represent the cumulative exposure to occupational agents. It is very challenging to accurately assess lifetime exposure to any occupational agent in a sample of workers drawn from a cross-section of occupations and industries, as it is quite common for subjects to work in different occupations in different industries across different time periods. The nature of retrospective exposure assessment requires information on past workplace processes and exposures. However, data such as historical exposure measurements across the wide spectrum of workplaces are very scarce, except for a few exposures such as asbestos, benzene, crystalline silica, and ionizing radiation, that were commonly found in traditional blue-collar manual jobs held by men. Consequently, retrospective population-based occupational studies often employ qualitative or semi-quantitative methods, including self-reported occupation, self-reported occupational exposure, case-by-case expert assessment, and JEMs, to assess past occupational exposures (77, 78).

Self-reported occupation title histories have been shown to be reasonably accurate; and the analysis of job titles has contributed useful leads in identifying at-risk situations (76). However, the analysis of risk in relation to job titles does not necessarily provide good leads in identifying the particular chemical agents that may be responsible for an observed risk, and it fails to bring together all the workers exposed to a given agent in different occupations (79). Some studies assess chemical exposures through subject self-report. Although providing a relatively cheap and fast method to assess occupational chemical exposure, this method suffers from considerable error in terms of a given subject's ability to correctly identify specific agents used in the workplace due to lack of knowledge of the presence of specific chemicals contained in products or due to the complexity of the work environment and tasks (80). Another drawback associated with this method is the potential for differential recall bias in case-control studies: controls do not have the same incentive as cases who may be attempting to deduce exposures that may have contributed to their disease (76).

Initially developed in the early 80s, expert assessment of occupational exposure to chemicals and physical agents is a method by which experts in fields such as occupational hygiene, chemistry and engineering, assess each job held and reported by each subject, to assign a level of exposure for each job to a selected list of agents based on their expertise and relevant literature (79, 81). Compared to subjects' self-reported occupational exposure, this method generally provides more accurate and valid estimates of exposure, and it is "usually the best approach" for retrospective occupational exposure assessment in case-control studies (76). This approach requires detailed occupational histories to be collected, preferably in face-to-face interviews, using well-designed questionnaires and interview techniques. Further, the expert-assessment method is very costly and time consuming, making it almost impossible to implement in large-scale population-based occupational studies, particularly those involving multiple exposures of interest (82). A less costly and time-prohibitive method, which has gained popularity since the 1990s, is to use a JEM built on either expert assessment or existing routinely collected measurement data to assess lifetime exposure to specific occupational agents based on the subject's job history. JEMs are basically cross-tabulations with at least two axes: one occupational axis containing a list of occupation or industry titles, and an agent axis containing a list of agents (83). Some JEMs also include a time period axis with two or more time periods. For each combination of an occupation/industry and an agent, and when available a time period, JEMs can provide information on one or more exposure estimates including: the intensity, frequency and/or the probability of exposure. However, as JEMs provide aggregated estimates of exposure, they are unable to account for exposure heterogeneity within one specific job title and/or time period (76).

Ge et al. reviewed 34 reliability studies that compared concordance agreement between two or more occupational exposure assessment methods, namely self-reported exposure, expert-assessed exposure, JEM-assessed exposure, and quantitatively measured exposure (77). Cohen's kappa statistic (k) was often used by reliability studies included in the review to evaluate concordance agreement between categorical measures of exposure (e.g., presence/absence of exposure) (84).

All studies included in this review compared candidate assessment method(s) against one or more assessment method(s) selected as the comparison method(s) (77). In addition to kappa, some studies also reported agent sensitivity and specificity values using their selected comparison method as the reference method. Case-by-case expert assessment was the most frequently used assessment method among the reviewed reliability studies and was often used as the reference comparison method against other exposure assessment methods (since expert-assessment is often considered as the "best-available approach" for retrospective exposure assessment in the absence of a true gold standard method). Some studies compared exposure assessments between different experts. Among the studies that compared agreements between various exposure assessment methods, concordance results vary greatly depending on the choice of agent(s) to examine and the exposure assessment methods being compared. Overall, using expert-assessment as the reference comparison method, Ge et al. found slightly higher median kappa agreement between different experts (k=0.6) than between expert-assessment and self-reported exposures (k=0.5) or JEMs (k=0.4).

A few of the reliability studies included in Ge et al.'s review compared reliability between expert-assessed exposures and some of the emerging exposure assessment methods, which employ various statistical learning/clustering machine-learning algorithms to predict exposure based on self-reported questionnaire responses (77). These studies compared reliability agreement for very few agents and a majority focused on predicting exposure to diesel engine exhaust, followed by asbestos and polycyclic aromatic hydrocarbons (PAHs). Reported kappa values comparing exposure probabilities estimated with machine-learning algorithms vs. expert-assessment ranged between 0.49 and 0.82 in three of the studies that assessed reliability for diesel engine exhaust, asbestos, or PAHs. Agreement between these algorithm-based methods tended to be higher for agents with high probability of exposure in a few occupations (e. g., diesel engine exhaust) compared to agents with low probability of exposure across many occupations (e.g., PAHs).

Most of the reliability studies identified in the current literature focus on occupational exposures that are known or suspected carcinogens, and most of these studies validate exposure reliability for male workers. There is a lack in literature regarding information on the reliability of different exposure assessment methods for occupational agents that are commonly encountered in jobs held by women.

2.5 Gaps in knowledge on occupational risk factors for lung cancer in women

Almost all the evidence on occupational risk factors for lung cancer has been generated among male workers. In part, this is due to the historic job markets and occupational profiles of men and women and/or due to the choices made by researchers (85). During most of the period from 1950s to 1990s when much of the occupational epidemiological research took place, there were proportionately fewer women compared to men who were employed for long durations of time in industries/occupations that involved heavy industrial work. This meant that there was very low statistical power to detect elevated risks among women even if there had been elevated risks. Given the limited resources available for occupational research, earlier studies tended to focus on the more numerous and heavily exposed (to chemicals) occupations where male workers predominated, in an effect to improve statistical power to detecting risks. There were some occupational lung cancer studies that included female workers, but exposure assessment methods and statistical analyses of female data tended to be of inferior quality (i.e., less detailed occupational exposure assessment, less in-depth statistical analysis) compared to many analyses of male data (85-87).

The lack of empirical evidence regarding occupational cancer risks among women is an important concern. As the historic differential in incidence of lung cancer between men and women has diminished (88), and as the historic differential in workforce participation has likewise diminished (89), it becomes more important to evaluate workplace influences on lung cancer among women, without assuming that the effects are identical among men and women. Women may have different biological responses to the same occupational exposures that have been investigated in men; this may be due to different genetic, hormonal, or environmental co-factors, or it might be due to differences in the ways in which men and women are exposed. A 2017 study of toxicokinetics has shown that absorption rate, metabolism and bioavailability of chemicals are likely to differ by sex (90) and previous studies have observed a sex difference in the level of particle deposition in human lungs (91, 92).

Aside from sex differences, gender differences may also influence the tasks performed by men and women, leading to possible differences in exposure within the same occupation (93). In addition, due to the different occupational/industrial profiles of men and women, there are likely agents to which women are more likely to be exposed than men. For example, it has been reported that women were 30% more likely to report occupational exposure to disinfectants, hair dyes and textile dust when compared to men; while men were two to four times more likely to report occupational exposure to dust and chemical substances, loud noise, and vibrating tools when compared to women (93). Indeed, previous occupational cancer studies tended to focus on exposures in jobs that have historically been mostly occupied by men (i.e., jobs in the construction, mining, forestry and heavy manufacturing settings) while exposures in jobs mostly occupied by women such as those in the service sector, which have employed increasing numbers of women over the past 70 years, have rarely been studied. Scarselli et al. examined gender differences in occupational exposure among Italian workers (94). Exposed female workers were more likely than male workers to have higher exposure levels to formaldehyde, vinyl chloride monomer, propylene oxide, trichloroethylene, cadmium, nickel, and chromium IV compounds, after adjusting for age at exposure. The risk of having high or medium exposure levels was significantly higher in exposed women (versus that in exposed men) employed as blue collar workers, hired either in small (10–19 workers) or medium-large firms (> 50 workers), and working in firms manufacturing fabricated metal products, construction, and retail trade (mainly furniture and hardware) (94). The presence of high-exposure groups of female workers in many industrial sectors makes it vital to conduct further research in this field.

It has always been a recognized challenging endeavor to validate retrospective exposure assessments in occupational epidemiological studies. It is even more problematic for studies attempting to study female occupational risk factors for diseases since data on the validity of exposure assessment tools for women's exposure are very sparse. Previous validation studies were often limited in available resources and thus were only able to evaluate a few agents, and most of those were agents commonly found in male-dominated occupations.

To improve these identified research gaps, this thesis was conceptualized to further our knowledge on occupational risk factors of lung cancer among women using detailed occupational exposure assessment methods to agents and large databases of lung cancer case-control studies of women. The associations between lung cancer risk and exposure to a large list of agents that are prevalent in jobs held by women were assessed.

Chapter 3: Thesis objective

The overarching objective of the proposed thesis is to investigate whether selected occupational exposures are associated with lung cancer risk among women.

To address this research question, we used two lung cancer datasets of women. There are different strengths and limitations associated with each dataset regarding its exposure assessment tool and its statistical power to detect lung cancer risks. Namely, we have access to a Montreal-based case-control study of lung cancer, where female workers' occupational exposures were estimated using expert assessment, which is considered the best available retrospective exposure assessment method; however, this study is somewhat limited in its power to detect lung cancer associations due to its relatively small number of female participants. We also have access to a multicenter, international case-control dataset of lung cancer with a very large sample size of women, which was created by combining the women from ten individual case-control studies. All of those individual studies collected job histories of the women, but there were no available individual expert-assessed occupational exposure histories. To estimate occupational exposures, we decided to use a Canadian Job-Exposure Matrix (CANJEM) which had been developed by our team. In order to use CANJEM optimally, we wished to focus on those agents in CANJEM that provide similar exposure estimates to that of the expert assessment. This led to an attempt to quantify the ability of CANJEM to reproduce expert assessment, which resulted in an investigation of concordance between CANJEM and expert assessment.

Within this context, the main objective of this thesis is further divided into three specific sub-objectives:

1) To investigate associations between prevalent occupational exposures as assessed by experts and risk of lung cancer among women in the Montreal lung cancer case-control study.
2) To compare exposure assignment concordance between the Canadian Job-Exposure Matrix (CANJEM) and expert assessment for jobs held by women in the Montreal lung cancer case-control study.

3) To investigate associations between CANJEM-assessed prevalent occupational exposures and risk of lung cancer among women in a combined dataset of ten lung cancer case-control studies from Europe, Canada, and New Zealand.

Chapter 4: Methodology

The overarching aim of this thesis was to explore the role of occupational exposures in lung cancer risk among women. Three manuscripts were produced, and each addressed one of the three sub-objectives of the thesis. The first manuscript examined the associations between prevalent occupational exposures assessed by experts and risk of lung cancer among women in the Montreal lung cancer case-control study. The second manuscript compared exposure assignment concordance between CANJEM and expert assessment for jobs held by women in the Montreal study. Finally, the last manuscript examined the associations between CANJEM-assessed prevalent occupational exposures and risk of lung cancer among women in a combined dataset of ten lung cancer case-control study centers from Europe, Canada, and New Zealand. The methodologies for these analyses are presented below.

4.1 Sub-objective 1 —Occupational exposures assessed by experts and lung cancer risk among women in the Montreal study

The first specific sub-objective of the thesis was to examine potential associations between prevalent occupational exposures as assessed by a team of experts and lung cancer risk in women, using data from a Montreal-based case-control study.

4.1.1 Data source

<u>The Montreal lung cancer case-control study</u>

The population-based Montreal lung cancer case-control study was conducted from 1996 to 2001, with the primary objective of studying the associations between a wide range of occupational, environmental and lifestyle factors and lung cancer etiology (95). Incident lung cancer cases were ascertained from all major hospitals in the Montreal area during the study period. Cases (n=1,203; 781 men and 422 women) included subjects aged 35-75 years, with histologically confirmed incident primary lung cancer, who were Montreal residents and Canadian citizens. Controls (n=1,513; 936 men and 577 women) were also Montreal residents and Canadian citizens and were randomly selected from the population-based electoral lists and were frequency matched to cases by age, sex and area of residence. Response proportions were 82% and 69% for female cases and population controls, respectively.

Data collection for cases and controls involved in-person interviews with trained interviewers. Next-of-kin provided responses for 34% of female cases and 5% of female controls that were too ill or had died. Information on the subject's socio-demographic characteristics, lifestyle, medical history, family history of cancer, and detailed lifetime smoking history was collected using a structured questionnaire. Then a semi-structured questionnaire was administered to obtain a lifetime job history; namely, a detailed description for each job ever held by the subject, including information about the company, its products or main activities, the nature of the work site, the primary and subsidiary tasks of the subject and additional information on the use of protective equipment, equipment maintenance, and activities of co-workers. Supplementary specialized questionnaires with detailed technical probing questions were also administered for many specific occupations (e.g., sewing machine operators, nurses, cooks, construction workers, and welders).

Ethics approval was obtained from the Institutional Review Boards of the Institut Armand-Frappier, McGill University, the Université de Montréal and all participating hospitals (see Appendix 1); all study participants provided informed consent.

4.1.2 Design and study population

This study used a population-based case-control design to examine lung cancer risk in women. The data were available for both male and female participants in a Montreal lung cancer case-control study. In this study, we limited the analysis to women who had ever worked, which resulted in 361 female incident lung cancer cases and 521 female population-based controls that were frequency-matched to the age distribution of cases.

4.1.3 Outcome measure

The outcome of interest in this study was histologically confirmed incident primary lung cancer and cases were ascertained from all major hospitals in the Montreal area during the study period.

4.1.4 Exposure assessment method

<u>Expert assessment of occupational exposures</u>

The research group directed by Dr. Jack Siemiatycki has contributed to the methodological development of occupational exposure assessment for community-based case-control studies over the past several decades. In the 1980s, they developed an approach to assessing occupational exposures that involves a team of experts reviewing each job described by interviewed cases or controls in their job histories, an approach which became known as the expert assessment method (96-98). Detailed lifetime occupational histories include job title, performed tasks, work environment, products or equipment used for any job that lasted for more than six months. Duration of each job was derived from the reported start and end date provided by the subjects during in-person or telephone interview. A team of trained experts in chemistry and industrial hygiene, blinded to subject's case-control status, coded each job according to standardized occupational and industrial coding systems and reviewed each job to determine possible occupational exposure to a list of 294 chemical, biological and physical agents. It is noteworthy that the exposure assignment to an agent was based not only on the reported job titles, but also on the unique characteristics of the workplace and the tasks reported.

Three experts participated in the expert assessment of occupational exposures in the Montreal lung study. A consensus approach was adopted during the coding process. Each job was reviewed by at least two experts: the first one would conduct an in-depth assessment of potential exposure to all 294 agents on the list and one or two other experts would review and validate the assessments. Disagreements on the presence of an exposure or the level of exposure to an agent would be resolved by discussion until a consensus was reached. For each agent to which the experts considered the subject had been exposed in a given job, three indicators of the assessment were recorded: the degree of reliability or confidence in the assessment (possible, probable, or definite); the presumed concentration or intensity of exposure (low, medium, or high); and the presumed frequency of exposure during a typical workweek (the number of hours a subject was exposed). The presumed intensity of exposure was assigned in reference to certain *a priori* benchmarks based on the background level of exposure to each agent in the general population. It is established that for the agents under study, the level of exposure in the workplace would need to exceed that in the general environment to be considered as "exposed". A low intensity represents a slightly higher level of exposure to an agent above what one would be exposed to in a non-occupational setting during the relevant time period. A high intensity represents the highest intensity of exposure one would be exposed to in the workplace during the relevant time period. A medium intensity represents an intensity of exposure in between low and high, on a relative scale. The varying level of intensity (low, medium, high) can be transformed into numerical values depending on the approximate ratios of concentration that the experts had in mind when coding intensity for a particular agent. Naturally, the concentration of exposure differs by agent, thus the intensity of exposure level is not meant to be compared between agents. However, within each agent, the intensity levels can be compared across jobs.

A number of trials were carried out throughout the years by the team to assess reliability and validity of this expert assessment method. A previous inter-rater reliability study involving the same team of experts revealed high agreement between experts, with a kappa value around 0.8 (99). An exposure validity study was also carried out to compare the exposure assessments from the same team of experts with previously recorded industrial hygiene air sampling measurements in Australia (gold-standard measurement) for 19 occupational agents, and the average sensitivity was 73% for experts to detect true presence of exposures (100). Among agents coded as present by the experts, the raters were also quite accurate in rating the relative concentration and frequency of exposure (100). From 1980 to 2011, this expert assessment method was used in five cancer case-control studies in Canada, mainly in the Montreal area; one study on multiple cancer sites (101), one study on lung cancer (102), one study on brain cancer (103), and two studies on breast cancer (1996-1997 and 2008-2011) (104 , 105), with a stable set of experts in chemistry and industrial hygiene as the exposure assessors. In aggregate, the team of experts has spent over 50 expert-years in assessing exposures in these five studies.

4.1.5 Selection of occupational agents and exposure variables of interest

The number of exposed women in this study, as in most population-based case-control studies, is limited. To ensure a reasonable level of power to detect any association, if it exists, we established the following criteria for an agent to be included. Included agents needed to satisfy two conditions among the cases in our study sample: at least 30 cases exposed, where exposure was rated as 'probable' or 'definite'; and at least 10% of those exposed cases were substantially exposed (i.e., exposed at medium or high concentration and over 2 hours of exposure per week, and over 5 years of duration). This led to the selection of the following 22 agents for analysis: cotton dust, wool fibres, synthetic fibres, polyester fibres, treated textile fibres, cellulose, ammonia, formaldehyde, cooking fumes, isopropanol, toluene, synthetic adhesives, organic solvents, volatile organic liquids, alkanes C5-C17, aliphatic alcohols, aliphatic aldehydes, polycyclic aromatic hydrocarbons (PAHs), mononuclear aromatic hydrocarbons (MAHs), cleaning agents, biocides, and micro-organisms.

For each agent, we estimated lung cancer risk associated with ever exposure and cumulative exposure (CE) to that agent. The reference unexposed category for computation of the ORs for a given agent comprised women who were never exposed to that agent. Women's CE was calculated using the formula: $CE = \sum_{i=1}^{d} \frac{I_i}{25} \times \frac{F_i}{40}$ where *i* represents the *i*th year, *d* represents the total number of years exposed, *Ii* represents the intensity of exposure in year *i*, and *Fi* represents the number of hours exposed per week in year *i*. The values of *Ii* were transformed from (low, medium, high) to (1, 5, 25) as these were the approximate ratios of intensity that the experts had in mind when coding intensity for most agents. This formula for CE assigns equal weights to the concentration and frequency of exposure through dividing each measure by their highest value. Because we do not assume that the risk of lung cancer is linearly associated with each unit of increase in CE for every agent, the continuous measure of CE was further divided into three categories (never, ≤ median CE, > median CE), based on median distribution of CE values for each agent among exposed controls.

4.1.6 Selection of occupations and exposure variables of interest

Although the main purpose of the analysis was to examine the association between selected occupational agents and lung cancer risk among women, we also took the opportunity of the available data to analyze the association between selected prevalent occupations and lung cancer risk among women in this study. Six occupations categorized at the three-digit level of the International Standard Classification of Occupations, version 1968 (ISCO-68) satisfied the criterion of a minimum of 30 exposed cases and were selected for the job title analysis. These occupations include stenographers, typists and teletypists; bookkeepers and cashiers; correspondence and reporting clerks; waitresses, bartenders and related workers; maids and related housekeeping service workers not elsewhere classified; and sewers and embroiderers. For each selected occupation, we analyzed lung cancer risk associated with ever exposure, and with over 10 years in that occupation, using women who had never worked in that occupation as the reference unexposed category.

4.1.7 Analyses

Odds ratios (ORs) and 95% confidence intervals (CIs) between exposure to each selected occupational agent, or an occupation, and lung cancer risk among women were estimated using separate unconditional logistic regression models, adjusted for selected covariates.

The selection of covariates to include in the models was based on the following considerations. We were limited to variables that had been collected in the original case-control study conducted in the late 1990's. Because the number of participants in this study was rather small, we avoided over-loading the list of covariates in the statistical models. In every model we forced three variables that are strongly associated with lung cancer and that differ by exposure status in our study population: age (continuous), ethnicity (French Canadian, Others) and smoking. The dataset included a rich lifetime smoking history. Previous methodologic work by our team (106) showed that a single continuous variable composed of different dimensions of the smoking history (incorporating smoking status, duration, time since cessation and intensity), dubbed the comprehensive smoking index (CSI), provides as good a basis for controlling confounding by smoking as including the different dimensions separately in the model. This CSI variable has previously been used in this Montreal lung cancer case-control study and has been shown to be a valid and parsimonious alternative to conventional modeling of different aspects of smoking history (106). For this thesis, I conducted a sensitivity analysis to explore the impact of using more conventional tactics instead of the CSI, namely including three separate smoking dimensions as covariates (smoking status, cigarette pack-year, and time since quitting smoking).

In addition to the core covariates, we explored whether adding the following additional covariates to the model modified the estimates of relative risk of the examined occupational agents with lung cancer: education, family income, and domestic exposure to traditional (wood, coal, biomass, etc.) heating and cooking fuels.

Finally, the proportion of proxy respondents were higher in cases than in controls in this study, which can result in differential exposure misclassification if the quality of the information provided by proxy respondents differ from that of self-respondents. To examine this, we conduced additional sensitivity analyses restricted to self-respondents only.

Statistical analyses were conducted using R (V3.5.3).

4.2 Sub-objective 2 — The concordance of exposure assignment between CANJEM and Expert assessment

The second specific sub-objective of the thesis was to compare exposure assignment concordance between the Canadian Job-Exposure Matrix (CANJEM) to expert assessment for jobs held by women, using data from the Montreal lung cancer case-control study (described in section 4.1). This methodological investigation was conducted to aid in the selection of agents suitable for use in CANJEM for exposure assessment, in an international combined dataset of case-control studies of lung cancer among women, without expert-assessed exposure histories (sub-objective 3).

4.2.1 Occupational exposure data source

The Montreal lung cancer case-control study

Details of occupational exposure data from this study have already been described previously in section 4.1.1. In total, 3403 jobs were held by female participants in this study.

4.2.2 Exposure assessment methods

In this methodological investigation, we compared the concordance of assignment of exposure for prevalent occupational agents among jobs held by women in the Montreal study using two exposure assessment methods: expert assessment, which is considered as the best available method for retrospective exposure assessment in case-control studies, and CANJEM, a Canadian general-population JEM proposed as an alternative to expert assessment of occupational exposures.

4.2.2.1 Expert assessment

Details of the expert assessment method have been described previously in section 4.1.4.

4.2.2.2 The Canadian Job Exposure Matrix (CANJEM)

CANJEM: Background

CANJEM was developed by the team directed by Dr. Siemiatycki and Dr. Lavoué, using expert assessment data on occupational exposures from multiple case-control studies (one of which is the Montreal lung cancer study described previously in section 4.1.1) (107). At the time of the analysis for manuscript two, CANJEM exposure source data came from four Montreal-based case-control studies (i.e., the study of multiple cancers, lung, brain and breast (1996-1997)) which provided a total of 31,673 jobs from 8,912 subjects (107). Expert assessment of occupational exposure data from a second breast cancer case-control study (2008-2011) was later incorporated into CANJEM estimates and an updated CANJEM was then applied in manuscript three, but this data was not ready at the time of manuscript two. The data collection procedures were the same for all studies that provided source exposure data for CANJEM: in-person interviews with cases and controls to obtain detailed job histories which were then evaluated by a team of specially trained exposure experts to assign exposure using a checklist of up to 294 agents, as described above in section 4.1.4. They also coded the jobs according to various occupational and industrial classification systems.

The team of experts evaluated about 9000 subjects' work histories, and this comprised about 32,000 distinct jobs reported by the participants in the four case-control studies mentioned above. Given the huge database and rich information on the presence or absence of each agent on the checklist for each job, as well as indicators of the level of confidence in the assessment and the intensity and frequency of exposure assessed by the experts, and the dates of the job, the team took advantage of all this information and created a general-population job exposure matrix called CANJEM (http://www.canjem.ca/) (107, 108).

Construction of CANJEM: source data

CANJEM was built using the information from the expert assessment of jobs in four Canadian case-control studies. Table 2 below presents the study period, number of cases and controls, subject age range and the number of jobs assessed by the experts in these four studies. Studies 1, 2 and 3 were conducted in the Greater Montreal area and Study 4 was conducted in the greater Montreal, Ottawa and Vancouver areas, although the expert assessment of occupational exposures was only carried out for participants from the Montreal and Ottawa centers. The same protocol for obtaining subject's occupational history was used in all studies (in-depth subject interviews with trained interviewers), and the same expert assessment approach was carried out in all studies.

Table 2. Information on the case-control studies included in the CANJEM source data

Study	**Study period**	**No. cancer cases**	**No. population controls**	**Age range (yrs)**	**No. jobs**
Study 1 (multi-site cancer) (101)	1979-1986	3,726 men	533 men	35-70	15,067 (men)
Study 2 (lung cancer) (102)	1996-2001	739 men & 466 women	925 men & 616 women	35-75	10,371 (6,877 men & 3,494 women)
Study 3 (breast cancer) (104)	1996-1997	608 women	667 women	50-75	3,510 (women)
Study 4 (brain cancer) (103)	2000-2005	124 men & 121 women	198 men & 216 women	30-59	2,725 (1,461 men & 1,264 women)

The agents that were evaluated in each of the studies varied slightly; consequently, 36 out of the 294 agents were excluded from the CANJEM database as exposure assessments for those agents were not available in all four studies. In the end, CANJEM provides exposure estimates for 258 agents. The full list of agents and their descriptions can be found at: https://expostats.ca/chems/.

Construction of CANJEM: harmonization of source data

As mentioned previously, the same protocol for expert assessment of occupational exposures was used in all four studies, although there was a minor difference in how frequency of exposure was recorded in each study. In Study 1, the frequency of exposure to an agent in a given job was recorded in categories (exposed for: 0%, <5%, 5-30%, and ≥30% per week, assuming a 40-h workweek). While in Studies 2 through 4, the frequency of exposure was recorded as the number of hours exposed to an agent in a given job. In order to harmonize the data from all studies together, it was decided that a quantitative measure of frequency of exposure would be created for Study 1. This was achieved by applying the computed agent-specific median value of frequency of exposure (in number of hours) in Study 2 for each category of frequency of exposure in Study 1, and to assign these median values to the matching combination of agent, job and frequency category in Study 1. The median exposure values were computed from Study 2 because this is the largest study among Studies 2, 3 and 4 and the most similar to Study 1 in terms of the study population characteristics.

Construction of CANJEM: three-axes structure

There are three axes available in CANJEM: an occupational code axis, a time period axis, and an agent axis. These axes form the framework of CANJEM. Each cell within the matrix provides exposure metrics for a given agent within a specific combination of occupational code and time period.

The occupational code axis

CANJEM offers linkage via one international and three North American occupational classification systems, and one international and two North American industrial classification systems. The list of occupational and industrial coding systems available in CANJEM, as well as the coding resolutions and the number of groups in each classification can be found in Table 3 below.

Table 3. Available occupational and industrial coding systems in CANJEM.

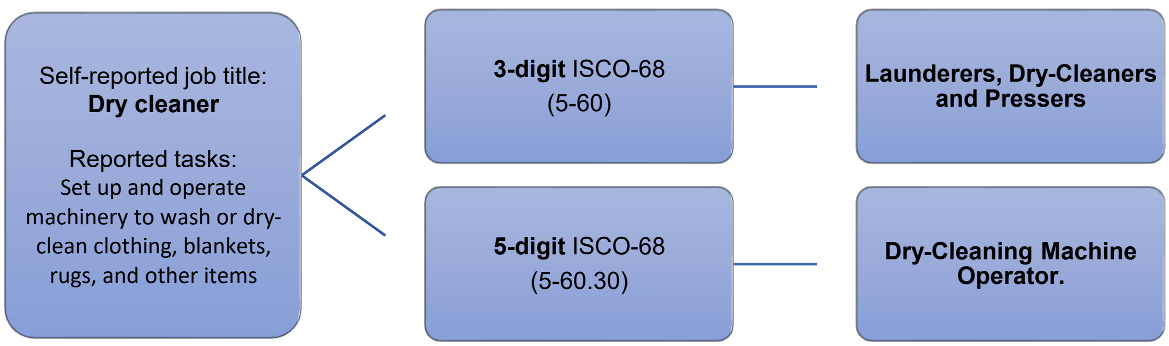
Table adapted from Sauvé et al. 2018 (109).

Classification	Resolution	Level	Number of groups in classification
(A) Occupations			
International Standard Classification of Occupations (ISCO), 1968[a,b]	1 digit	Major group	8
	2 digits	Minor group	81
	3 digits	Unit group	282
	5 digits	Occupation	1504
Canadian Classification and Dictionary of Occupations (CCDO), 1971[c]	2 digits	Major group	23
	3 digits	Minor group	81
	4 digits	Unit group	500
	7 digits	Occupation	7907
Canadian National Occupational Classification (NOC), 2011[d]	1 digit	Division	10
	2 digits	Major group	40
	3 digits	Minor group	140
	4 digits	Unit group	500
United States Standardized Occupational Classification (SOC), 2010[e]	2 digits	Major group	23
	3 digits[f]	Minor group	97
	5 digits	Broad occupation	461
	6 digits	Detailed occupation	840
(B) Industries			
International Standard Industrial Classification (ISIC) revision 2, 1968[g,h]	1 digit	Major division	9
	2 digits	Division	33
	3 digits	Major group	71
	4 digits	Group	159
Canadian Standardized Industrial Classification (SIC), 1980[i]	1 digit	Division	18
	2 digits	Major group	76
	3 digits	Minor group	318
	4 digits	Unit group	860
North American Industry Classification System (NAICS), 2012[j]	2 digits	Sector	20
	3 digits	Subsector	102
	4 digits	Group	323
	5 digits	Industry	711
	6 digits	Canadian industry	922

[a] International Labour Office (ILO) (1969).
[b] Includes Armed Forces as a category in each level of resolution.
[c] Dominion Bureau of Statistics (1970).
[d] Statistics Canada (2012a).
[e] U.S. Bureau of Labor Statistics (2014).
[f] Level includes two 4-digit codes: 15-11 (computer occupations) and 51-51 (printing workers).
[g] Major division 0 (Activities not Adequately Defined) and nested subgroups omitted.
[h] United Nations (1971).

Each system has its own hierarchical structure and differs in terms of resolution (number of digits for the codes) and the total number of occupational codes within each resolution. In all systems, lower resolution (fewer digits) job codes aggregate multiple higher resolution job codes into one broader job code. As the number of digits in the job code resolution increases, the job description gets more precise. As illustrated in Figure 3 below, a subject who had reported working as a dry cleaner and reported main job tasks of setting up and operating machinery to dry-clean clothing and other items in one of her jobs would have this job coded by the experts under different resolutions within each occupational coding system. Using the ISCO-68 system as an example, a 3-digit code (5-60) would only be able to define this job as "Launderers, dry-cleaners and pressers" while a higher resolution 5-digit code (5-60.30) would provide a more precise description of the job as being "Dry-cleaning machine operator".

Figure 3. Example of job description precision under different resolutions of job code.



Although higher resolutions provide more detailed job descriptions, there is a trade-off. CANJEM estimates for higher resolution job codes were generated from fewer source jobs compared to estimates generated from lower resolution job codes. When the number of source jobs is very low, the generated estimates of exposure might not be as accurate or representative as when a lower resolution was used.

The time period axis

At the time of analysis for manuscript two, CANJEM provided estimates of exposure for jobs reported from 1930 to 2005. Across the large time span, technology and industrial regulations have changed and consequently exposures to certain agents within a given occupation could have varied with time. Depending on the era when the population under study was economically active and the agent(s) being studied, CANJEM users have the freedom to customize their time interval for the time period axis of CANJEM in order to obtain exposure estimates that they judge to be the most suitable for their study population.

The agent axis

A list of the 258 agents available in CANJEM can be found on the CANJEM website at http://canjem.ca/. Descriptive statistics for each of these agents are available from http://expostats.ca/chems. The list of agents includes specific chemicals, broader mixtures, physical agents, and agents grouped together based on use. On the CANJEM website, the agents are presented in a hierarchical order, under the five main categories: inorganic substances, organics substances, inorganic and organic mixtures, general categories, and ionizing radiation, electric and magnetic fields. It is important to note that some of the agents could partially overlap with other agent(s) due to the hierarchical agent structure. For example, benzene is part of the agent mononuclear aromatic hydrocarbons, which in itself is part of the agent organic solvents.

Construction of CANJEM: available metrics

For each combination of occupational code, time period, and agent, CANJEM provides information on close to 50 variables (see Appendix 2). Among them, the most essential variables for epidemiologists encompass the probability of exposure, the certainty of exposure, and if the probability of exposure is above zero, the frequency and intensity of exposure.

Probability of exposure

The probability of exposure is a continuous measure ranging from 0-100%. It was derived by calculating the proportion of jobs that were classified as exposed by the experts to a given agent at any frequency or intensity among all jobs with the same occupational code in CANJEM. Unlike the expert assessment method which attributes different exposure assessments to each job on an individual basis, CANJEM, as any other JEM, would attribute the same level of exposure to an agent in all jobs with the same occupational code. This would not be an issue when the probability of exposure is either 0% or 100%, indicating that none or all of the assessed jobs sharing an occupational code were unexposed or exposed to an agent in a given time period. However, when the probability of exposure is somewhere in between, the exposure status for those jobs becomes less clear. In order to classify jobs into the exposed or un-exposed group, a threshold percentage for the probability of exposure variable needs to be chosen. The selection of probability cutpoints to define exposure status for epidemiologic analysis will be discussed later in section 4.2.4.

Frequency of exposure

CANJEM provides the median or mean frequency of exposure to an agent within an occupation code. The frequency of exposure is presented as the number of exposed hours per 40-hour workweek. To minimize the effect of outliers when computing the frequency, it is recommended that the median frequency of exposure be used in epidemiological studies (110).

Intensity of exposure

CANJEM provides the median intensity and mean frequency-weighted intensity of exposure to an agent within an occupation code. Intensity of exposure has three categories of low, medium and high, and is represented using four ratios (1:2:3, 1:3:9, 1:5:25, and 1:10:100). For epidemiological studies, it is recommended that the median intensity of exposure be used with the 1:5:25 ratio, unless otherwise specified. The median value of intensity is recommended over the mean to minimize the effect of outliers and the ratio of 1:5:25 is recommended because it is judged by the experts to be the best estimate of the low, medium and high concentration for the majority of agents in CANJEM.

Sample partial output of CANJEM for a select cell

Table 4 below provides an example of a partial CANJEM output representing the combination of ISCO-68 occupational code of 7-55.30 ‘Knitting-Machine Operator (Hosiery),’ for the agent fabric dust, during the time period of 1950-2005.

Table 4. Selected CANJEM output representing the combination of ISCO-68 code 7-55.30 Knitting-Machine Operator (Hosiery), for the agent fabric dust, covering the time period from 1950 to 2005.

Variable	Value
Occupational agent code (ISCO-68)	7-77.30
Occupational agent name	Fabric dust
Number of jobs within an occupation code	17
Number of exposed job within an occupation code	11
Probability of exposure	64.71
Median intensity of exposure using a ratio of 1:5:25	5
Median frequency of exposure (hours/week)	40

CANJEM: various versions

Users have the options to configurate different versions of CANJEM based on their choice. Different versions can be constructed depending on the selected occupational and industrial classification systems and the selected time periods to estimate exposures for agents. For example, within a given occupational classification system, users can choose multiple occupational code resolution(s) (e.g., users can download CANJEM's estimates for selected agents linked to the ISCO-68 three-digit and the ISCO-68 five-digit resolutions). For agents with a clear change in exposure profile during a specific time period, users can stratify by time periods and obtain CANJEM's estimates for those agents for each selected period. Users can also create different versions of CANJEM based on their preference in how exposure estimates with different levels of certainty should be treated. The expert assessment of the original five case-control studies that provided the exposure data for CANJEM assigned each exposure with a notation of their confidence of exposure (No, Possible, Probable or Definite exposure). CANJEM allows the user to select any level of confidence as a threshold for defining exposure. Namely, the following versions of CANJEM can be created based on user configuration: 1) possible exposure treated as exposed, 2) possible exposure treated as *not* exposed, 3) possible exposure removed from CANJEM estimates, 4) possible and probable exposures treated as *not* exposed, and 5) possible and probable exposures removed from CANJEM estimates. Finally, users can also customize the minimum frequency, intensity, number of jobs and number of subjects required to generate estimates in CANJEM.

4.2.2 Application of CANJEM: linkage strategies

Ideally, estimates of agents generated with the highest resolution within each occupational and industrial system and within a more specific time period should provide the best estimates of exposure, when there are enough jobs and subjects to populate each cell of CANJEM. When this is not the case, as it is in most situations, users need to come up with a strategy to maximize the linkage of subjects' jobs to CANJEM and to link those jobs at the highest resolution as far as possible.

A previous investigation examining various CANJEM linkage strategies (110) suggested that it is best to first prioritize linking jobs to the highest resolution within a chosen occupational system using a wider interval of time period, and then to link the un-linked jobs using a lower resolution of the occupational code, again using a wider interval of time period.

As mentioned earlier, CANJEM estimates for higher resolution job codes were generated from fewer source jobs compared to estimates generated from lower resolution job codes. When the number of source jobs is very low, the generated estimates of exposure might not be as accurate or representative as when a lower resolution is used. For manuscript two of the thesis, we decided to construct CANJEM estimates using the following two rules: 1) Source jobs from the four case-control studies with at least probable or definite confidence of exposure and with a minimal input of 10 jobs per cell coming from a minimum of 3 subjects were used to calculate estimates of exposure in a given cell (jobs with possible exposure were treated as *not* exposed). 2) Estimates of exposure were calculated within the time period from 1950 to 2005, as this roughly represents the period where the women in the study population were economically active. Once estimates of CANJEM were constructed, we linked women's jobs from the Montreal case-control study to CANJEM using the following procedure: Jobs were linked to CANJEM using the ISCO-68 classification system. Each job was first linked at the highest resolution (ISCO-68 five-digit); the unlinked jobs were then linked at the second highest resolution (ISCO-68 three-digit). Any jobs that could not be linked using these two resolutions were excluded from the analysis. This algorithm allowed us to maximize the job linkage while ensuring the accuracy of the estimates for the linked jobs. Using this strategy, we were able to link 99.95% of all jobs in our study population to CANJEM.

4.2.3 Selection of agents for exposure concordance comparison

For manuscript two on concordance between expert assessment and CANJEM in assessing exposures, a large list of prevalent exposures among jobs held by women were selected for comparison. Among the 258 agents that are available in CANJEM, we selected for this analysis those agents that had been listed by expert assessment as exposed in at least 30 of the 3403 jobs held by women in the Montreal case-control study of lung cancer. A total of 69 agents were selected based on this criterion. We evaluated the concordance between the CANJEM-derived assessment and the expert-derived assessment for each agent. The expert assessment method assigned a binary exposure status to an agent in a given job, whereas CANJEM provides the user with a continuous probability of exposure, and users can select a cutpoint of probability of exposure to dichotomize the exposure status. To reduce the influence of trivial exposures, for both methods, ever exposure to an agent was limited to jobs with probable or definite exposures (jobs with a possible exposure were considered unexposed), and that exposed jobs need to reach a minimum exposure frequency of 0.5 h per week and a frequency-weighted intensity (a continuous index that combines intensity and frequency of exposure) (108) corresponding to at least 2 h per week at low intensity.

4.2.4 Selection of CANJEM probability of exposure cutpoint

To define exposure status to an agent within a given job title, unlike the binary (exposed vs not exposed) exposure estimate output from the expert assessment, the output from CANJEM is a probability of exposure ranging from 0 to 100%. Although users have the option to configurate different versions of CANJEM as described in the section above, the output of exposure estimates for agents would always remain a probability of exposure. In epidemiologic studies, it is useful to categorize the exposure status as exposed or not exposed to an occupational agent for estimations of relative risks of diseases associated with that agent. The classification of a binary exposure status to an agent within a job held by a worker can be achieved using the continuous probability of exposure provided by CANJEM and choosing a probability of exposure cutpoint, below which we would consider the worker as not exposed, and above which we would consider the worker as exposed.

There is no universally established probability of exposure cutpoint to define ever exposure in occupational epidemiologic studies, this decision needs to be made by users of CANJEM for their individual projects. Ideally, if all the probabilities of exposure to an agent are either 0% or 100%, then we could simply categorize the 0% as not exposed and the 100% as exposed. However, the reality is that the probabilities fall along the range from 0% to 100%. If the empirical frequency distribution of the probability of exposure is bimodal, then this would lead to a natural "valley" between the two peaks where a cutpoint can be chosen; while if the empirical frequency distribution is not bimodal, then it is more arbitrary to decide on where to set the cutpoint. Figures 4 and 5 show examples in CANJEM of two agents that exemplify those two scenarios. Figure 4 provides a demonstration of a good scenario in which setting the probability of exposure cutpoint to around 25% would differentiate quite well occupations exposed and unexposed to cotton dust; while Figure 5 provides a less optimal scenario with far more occupations having a probability of exposure around the cut-off threshold, making it harder to correctly classify occupations exposed and unexposed to carbon monoxide.

Figure 4. Frequency distribution of probability of exposure to cotton dust in all ISCO-68 jobs with a probability > 0% assigned by CANJEM.

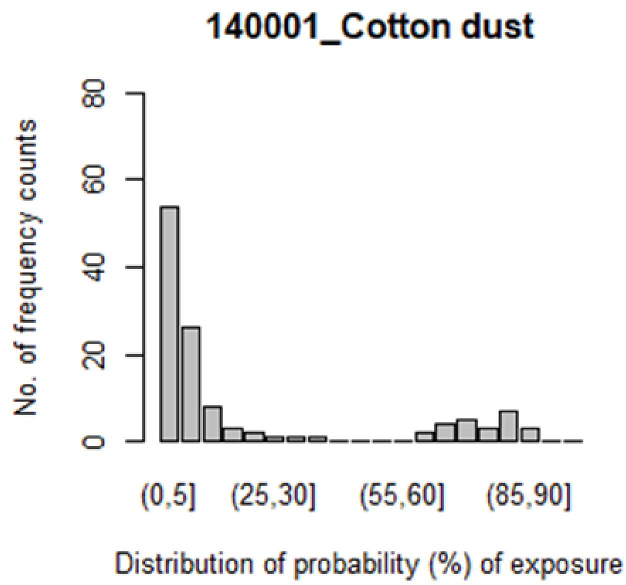
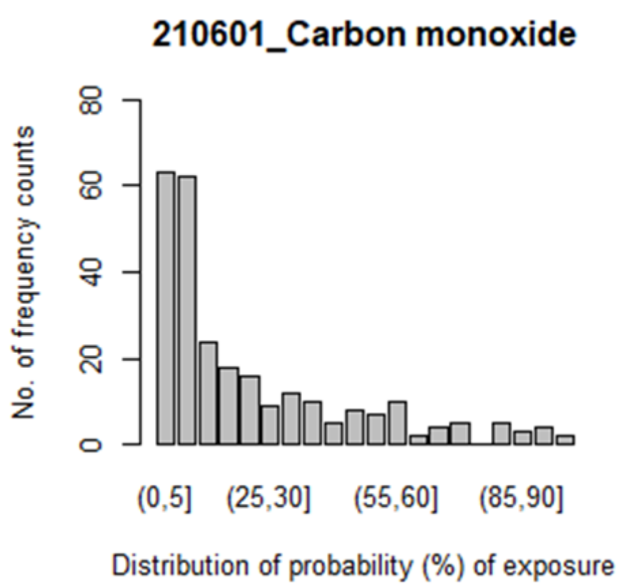


Figure 5. Frequency distribution of probability of exposure to carbon monoxide in all ISCO-68 jobs with a probability >0% assigned by CANJEM.



Some exposure misclassification is unavoidable when using any JEM including CANJEM, regardless of the choice of probability of exposure cutpoint. If we were to assume that the expert assessed exposure status is the gold standard, a selection of a lower probability of exposure cutpoint in CANJEM would entail a higher sensitivity and a lower specificity; indicating that we would likely be able to correctly identify all the "truly exposed" but there may be many "false positives" that would be put in the exposed group. While vice-versa, a selection of a higher probability threshold would entail a lower sensitivity and a higher specificity; indicating that we are more likely to miss some of the "truly exposed" but there would not be as many "false positives" included in the exposed group. There is no consensus on where the probability cutpoint should be. However, the two most frequently used cut-off thresholds of probability of exposure in epidemiological studies are ≥25% or ≥50% to define ever exposure to an agent (111-113). Consider that the proportion of exposed jobs to an agent in all jobs in CANJEM is relatively very small; the specificity values tend to be quite high and much less influenced by a given probability threshold than do the sensitivity values. To minimize exposure misclassification, we would want to maximize the number of truly unexposed jobs below the threshold and also to maximize the number of truly exposed jobs above the threshold. To address the 2nd sub-objective of this thesis, which is to compare exposure assignment concordance between CANJEM and expert assessment for agent exposures in jobs held by women, we used two probability cutpoints: 25% probability of exposure (CANJEM-25%) and 50% probability of exposure (CANJEM-50%), to determine whether jobs are exposed or unexposed to an agent based on CANJEM. Jobs with a probability of exposure above or equal to the chosen cutpoint were classified as "exposed", while jobs below the cutpoint were classified as "unexposed".

4.2.5 Analyses

This study uses expert assessment as the reference "gold standard" for retrospective exposure assessment, and CANJEM as a proxy to mimic expert assessment.

We calculated CANJEM's sensitivity and specificity for each agent (using CANJEM-25% and CANJEM-50%, separately) to examine how well it replicated expert-assigned exposures. In addition, Cohen's Kappa was calculated to evaluate the exposure concordance not due to chance (84), and Kendall's τ coefficients were calculated to measure the correlation between expert-assessed prevalence of exposure to that assessed by CANJEM-25% and CANJEM-50%. We also plotted the empirical cumulative distribution functions (ECDF) of the prevalence for each exposure assessment method. The units of observation used to estimate prevalence and concordance of each agent were the 3403 jobs. In addition, we calculated and compared the loss in statistical power associated with the use of a proxy (CANJEM) instead of the "gold standard" expert assessment. We made the supposition that there would be a hypothetical case-control study with 1000 cases and 1000 controls (assuming one job per participant), and that there would be a fixed two-fold association between occupational agent exposure and lung cancer. To demonstrate the impact of statistical power loss to detect an association under varying exposure misclassification scenarios, we selected combinations of three levels of the following parameters: sensitivity (0.30, 0.50, and 0.80), specificity (0.90, 0.95, and 0.99), and true prevalence of exposure (2%, 5%, and 15%). As a result of using an exposure assessment method with imperfect sensitivity and specificity, the estimated prevalence of exposure and the OR can be described as follows

(114, 115):

$$P^* = PU + (1 - P)(1 - V)$$

$$R^* = \frac{[URP + (1 - V)(1 - P)][(1 - U)P + V(1 - P)]}{[UP + (1 - V)(1 - P)][(1 - U)RP + V(1 - P)]}$$

Where:

P = Prevalence of exposure estimated by the gold standard

U = Sensitivity

V = Specificity

R = Hypothetical true risk (OR) of 2.0

P^*= Misclassified prevalence of exposure

R^*= Misclassified OR after exposure misclassification

We first calculated the power of the "gold standard" expert assessment to detect a two-fold risk under varying levels of exposure prevalence. We then calculated the power under varying scenarios of exposure misclassification by replacing *P* with P^*, and *R* with R^*. The proportional difference between the two calculated powers is the loss in statistical power due to exposure misclassification.

All statistical analyses were conducted using R.

4.3 Sub-objective 3 —Occupational exposures assessed by CANJEM and lung cancer risk in women in a combined dataset of international case-control studies of lung cancer

The third specific sub-objective of this thesis was to examine the associations between CANJEM-assessed prevalent occupational exposures and risk of lung cancer among women, using data combined from ten lung cancer case-control studies from Europe, Canada, and New Zealand.

4.3.1 Data source

An International combined study of lung cancer among women

This analysis included data on 7,227 female participants combined from ten case-control studies of lung cancer from Europe, Canada, and New Zealand; all participating study centers originally collected lifetime working and smoking histories, and information on socio-demographic factors such as education (116-124). Seven of the included studies were from Europe (France (120), Germany (117, 119), Italy (118, 124), Poland (121), and the United Kingdom (121)), two were from Canada (116, 123), and one from New Zealand (122). Data collection periods for these studies ranged from 1988 to 2008. The occupational data were coded or recoded from national classifications in each original study center into the International Standard Classification of Occupations (ISCO-68). Lifetime occupational and smoking information was mainly collected using face-to-face interviews (approximately 80%), the remainder by telephone interviews. In total, the 7,227 women in the combined study population had held 25,679 jobs for at least one year.

4.3.2 Design and study population

This study used data from ten population- or hospital-based case-control studies of lung cancer. Analyses were restricted to women in each of the study centers who had ever worked for at least a year, leading to a total of 3,040 lung cancer cases and 4,187 controls. Cases were incident lung cancer cases confirmed by histology or cytology, ascertained from local hospitals, clinics, or cancer registries. Controls were frequency-matched (approximately 96%) or individually-matched to cases by age and were recruited from the local general population. Two study centers recruited additional hospital controls (121, 123). Participation rates in the different study centers ranged from 53% to 89% among cases and 41% to 87% among controls.

4.3.3 Outcome measure

The outcome of interest in this study was incident, histologically confirmed, primary lung cancer.

4.3.4 Exposure assessment method

CANJEM assessment of occupational exposures

We estimated occupational exposures of women from ten lung cancer studies CANJEM. Details of CANJEM, including the background of this general-population-based JEM, its construction and application have been discussed in Section 4.2.2.2. As CANJEM continues to evolve, additional occupational source data of over 4000 jobs held between 1951 and 2011 among 1,297 women participating in a second Canadian breast cancer study conducted from 2009 to 2011 in Montreal, Canada (105) was added to CANJEM following the publication of manuscript two of this thesis. This newly added breast cancer study was conducted by the same group of researchers involved in the first breast cancer study (125) already included as source data during the initial construction of CANJEM, and used the same protocol for obtaining participant's occupational histories, and the same expert-assessment method, used in their first breast cancer study. Therefore, all five studies that formed the basis of the current CANJEM used the same expert assessment method to assess occupational exposures. Given the data collection periods of the ten lung cancer studies of female participants examined in this manuscript, we customized CANJEM to produce occupational exposure estimates using source data covering 1950 to 2011, which best reflected the period during which the combined study population was economically active.

A total of 25,679 jobs held by the combined study population from the combined dataset were linked to CANJEM using the ISCO-68 occupation code. Using the linkage strategy developed for the 2nd sub-objective of the thesis, which has been presented in manuscript two, we first attempted to link the jobs to the highest resolution (5-digit) of ISCO-68; if unlinked at the highest resolution, we then linked them at the second highest resolution (3-digit) of ISCO-68. We were able to link CANJEM and provide estimates for 96.5% of all jobs using this strategy; 83.6% of jobs were linked at the 5-digit resolution and 12.9% were linked at the 3-digit resolution. Following the job linkage, exposure to each of 258 occupational agents available in CANJEM was derived.

4.3.5 Selection of occupational agents and exposure variables of interest

It would have been untenable to present in this manuscript lung cancer risk estimates in relation to exposure to each of 258 agents generated by linkage with CANJEM. Some types of criteria were needed to reduce the number of agents for analysis. We reduced the number of agents using three criteria based on the prevalence of a given agent; the reliability of CANJEM in assigning exposure to an agent when compared to the expert assessment; and the hierarchical redundancy among the agents. Specifically, our operational decision was to only include agents with a lifetime ever exposure prevalence of 5% or higher in either cases or controls. Lifetime ever exposure to an agent was assigned if a participant was exposed in at least one of her jobs. This prevalence criterion led to the elimination of 232 agents from the investigation list. Manuscript two of this thesis was conducted to inform the current analysis regarding the selection of suitable agents to examine in this combined dataset. Agents that exhibited poor exposure assignment concordance between CANJEM and expert assessment for jobs held by women in the Montreal lung cancer study were eliminated from the list of agents to examine using CANJEM in manuscript three. Based on concordance results provided in manuscript two, we further eliminated five agents with kappa values less than 0.30. There were four agents that were not assessed in manuscript two because they were not prevalent in the Montreal study and those were also excluded from the current analysis. Finally, regarding the hierarchical redundancy among the agents, two agents that hierarchically overlapped with other more specific agents were further excluded (i.e., the agent "fabric dusts" was excluded because it overlapped with two other more specific agents "cotton dust" and "synthetic fibers"; the agent "aliphatic aldehydes" was excluded because it overlapped with the more specific "formaldehyde"). As a result of these eliminations, 15 agents satisfied all the inclusion criteria. These agents are: cleaning agents, biocides, cotton dust, synthetic fibers, formaldehyde, cooking fumes, organic solvents, cellulose, polycyclic aromatic hydrocarbons (PAHs) from petroleum, ammonia, metallic dust, alkanes C18+ (e.g., petroleum jelly), iron compounds, isopropanol, and calcium carbonate.

Three exposure variables of interest (ever exposure, duration of exposure, and cumulative exposure (CE)) were created for each of the 15 selected agents. In order to categorize an agent's exposure status (exposed or unexposed), we selected within CANJEM a probability cutpoint of 50% (CANJEM-50%) for establishing the "ever exposure" variable. For each combination of ISCO code and agent, when the probability of exposure was at least 50%, the job was considered as exposed, when the probability of exposure was less than 10%, the job was considered as unexposed, and when the probability was between 10% and 50%, the job was considered as "uncertainly exposed". A participant was classified as "ever exposed" to an agent if any of her jobs exposed her to that agent. The "duration of exposure" variable was calculated as the sum of self-reported duration of each job in which the participant was classified as ever exposed to an agent; it was then categorized into three groups (unexposed, 1–10 years, >10 years). Finally, the cumulative exposure (CE) was calculated as: $CE = \sum_{i=1}^{d} \frac{I_i}{25} \times \frac{F_i}{40}$, where *i* represents the *i*th year, *d* represents the total number of years that a participant was classified as ever exposed, *Ii* represents the intensity of exposure in year *i*, and *Fi* represents the number of hours exposed per week in year *i* (detailed justification for using this formula was discussed in Section 4.1.5). This CE variable was then categorized into three groups (unexposed, ≤ median CE, > median CE) based on agent-specific median values among ever exposed controls. Participants with uncertain exposure were excluded from the duration or CE analyses.

4.3.6 Analyses

We first conducted unconditional logistic regression models to estimate odds ratios (ORs) and 95% confidence intervals (CIs) of lung cancer associated with each agent's various exposure metrics estimated using CANJEM-50% in each of the ten case-control studies, separately. The reference unexposed category for each analysis contained women who were never exposed to the particular agent under scrutiny. Each of the logistic regression models was adjusted for the following covariates: age (log-transformed), cigarette pack-years (log [pack-years +1]; the cigarette pack-year was calculated as follows: ∑ duration (years) X average cigarette smoking intensity per day/20), years since quitting smoking cigarettes, ever employed in a blue-collar job (defined as jobs with an ISCO-68 first digit of 7, 8, or 9), and socio-economic status (SES). In all study centers except New Zealand, education (no formal education, some primary, primary/some secondary, secondary/some college, and university) was used as the proxy for SES covariate adjustment. The New Zealand Socio-Economic Index (NZSEI) was used as the proxy for SES adjustment for the New Zealand study (126). Additional adjustment for ever employed in a list of occupations known to be associated with lung cancer (127, 128) and additional smoking adjustment for never-, former-, or current-smoking status did not meaningfully change the ORs for agent estimates and therefore these two covariates were not retained in the final model. Analyses were conducted for each agent among all participants using CANJEM-50%.

The main analyses were conducted to assess lung cancer risks associated with exposure to each agent in all participating women, by smoking stratum, and by lung cancer histological subtypes in each of the ten participating study centers. Smoking stratified analyses were conducted among never-, light-, and heavy-smokers, separately. Light- and heavy-smokers were categorized based on the median value of pack-years among ever-smokers. lung subtype analyses were conducted for the three most prevalent lung cancer histological subtypes: adenocarcinoma, squamous cell carcinoma, and small cell lung carcinoma. Once the ORs and 95% CIs for each agent from each separate study center were calculated, we then agglomerated the individual study results using random-effects meta-analysis and assessed heterogeneity among studies using I^2 statistics. The I^2 is interpreted as the percentage of the total variability in a set of effect sizes due to between-study variability; and is calculated as $\frac{Q-(k-1)}{Q} * 100\%$ for when the Q statistic is greater than its degrees of freedom (k-1) where k being the number of studies, and I^2 is truncated to 0% when the Q statistics is smaller or equal to its degrees of freedom (129).

As sensitivity analyses, meta-analyzed lung cancer risks associated with exposure to each selected agent estimated using CANJEM-25%, instead of CANJEM-50%, were estimated for all women, and by smoking stratum. When categorizing agent exposure status using CANJEM-25%, any exposure with a probability ≥25% would be considered as "ever exposed", between 0% to <10% considered as "unexposed", and between 10% to <25% considered as "uncertainly exposed". Sensitivity analyses replacing the meta-analytic approach with the pooled approach were also conducted. Pooled analyses on the association between exposure to each agent and lung cancer in women from all ten participating study centers were performed. The pooled analysis for each agent was adjusted for the same set of covariates included in the meta-analysis. Because education data were unavailable in the New Zealand study, a category called “unavailable data” was assigned to all observations from this study center for the “education” covariate.

Over the past decades, epidemiologists have debated on whether occupational epidemiologic studies should be adjusted for SES (130). This is largely because occupations and their related occupational exposures are often associated with SES. There are two sides to this argument. On one hand, if SES is also a predictor for the outcome, then it would be reasonable to adjust for SES. On the other hand, if some occupational exposures are strongly tied to a socioeconomic level (e.g., asbestos exposure and working as a blue-collar worker), then conditioning on SES could cause an over-adjustment of the model, and thus may lead to a reduced statistical precision and/or an attenuated risk estimate of the main occupational exposure of interest. To explore whether there is a potential of over-adjustment in our modeling due to the adjustment of SES, we performed sensitivity analyses adjusting for a reduced set of covariates (only age and smoking) for the meta-analysis of associations between ever exposure to each agent and lung cancer risk among all women.

All statistical analyses were performed with R, version 4.3.0. Meta-analyses were performed with the “meta” package in R (131).

Chapter 5: Organization and complementary information for Chapters 6 to 8

This PhD project was conceptualized to explore occupational risk factors of lung cancer in women. As most previous occupational studies of cancer have been heavily focused on men, aside from studies focusing on female-specific cancer sites, women's occupational exposure to carcinogens has been long overlooked. We believe that it is equally important to study occupational exposures in women since women and men may have different carcinogenic responses to the same chemicals. Further, due to the difference in occupational/industrial profiles in women and men, some occupational exposures may only be found or found with sufficient prevalence in women, and these would not have been evaluated for possible carcinogenicity in past male-based studies. With this main objective in mind, three separate research projects were carried out and have subsequently led to three manuscripts.

The first study used in this thesis is a population-based Montreal case-control study of lung cancer conducted in the late 1990s to early 2000s by our own research team, where a team of experts have evaluated each subject's reported lifetime job history and assigned exposure to a list of close to 300 occupational agents to each job. Using this expert assessment approach, which is often considered as the best available approach for retrospective occupational exposure assessment, we examined the association between lung cancer risk and the 22 most prevalent agents that the women in this study population were exposed to. The results of this research are presented in Chapter 6 (manuscript 1).

Although the expert assessment approach is perhaps the best available method to retrospectively assess lifetime occupational exposure given that there is no real gold standard method for this; it is extremely costly and time-consuming. Therefore, our team has since then been working on developing a general-population JEM called CANJEM, which uses input from expert-assessed data, to provide a useful and cost-effective tool for studies that wish to retrospectively assess occupational exposures to potentially a large list of agents. An initial publication regarding the availability of CANJEM for epidemiologic and occupational medicine purposes was published in 2018 (107). CANJEM was built using expert-assessed exposure data from both male and female workers. Approximately 35% of jobs coded in those studies pertained to female workers. Some occupations were predominantly held by males and some by females, and some by both. Because we were interested to know whether CANJEM would be suitable to examine occupational exposures among a large sample of women from ten case-control studies where expert-assessed exposures were unavailable, there was a need to first validate CANJEM's performance in replicating expert-assessed occupational exposures among jobs held by women. Since both expert-assessed and CANJEM-assessed data are available for female occupational exposures in the Montreal lung cancer study, we took this unique opportunity and conducted a methodological investigation comparing concordance measures between CANJEM and the expert assessment for a large list of agents. This research is presented in Chapter 7 (manuscript 2).

Our findings from manuscript 2 indicate that CANJEM's ability to reproduce expert assessment is highly agent-dependent. Therefore, in Chapter 8 (manuscript 3), we selected 15 agents that are suitable to examine using CANJEM and are prevalent in jobs held by the combined sample of women from ten case-control studies of lung cancer, and examined potential lung cancer risk associated with exposure to each of these agents.

Chapter 6: Manuscript 1

The published manuscript can be found online at:

doi: http://dx.doi.org/10.1136/oemed-2020-106470

Workplace

Original research

Role of occupational exposures in lung cancer risk among women

Mengting Xu, [1,2] Vikki Ho, [1,2] Jack Siemiatycki [1,2]

Role of occupational exposures in lung cancer risk among women

Mengting Xu[1,2], Vikki Ho[1,2], Jack Siemiatycki[1,2]

1. Department of Social and Preventive Medicine, University of Montreal, Montreal, Canada
2. Carrefour de l'innovation, Université de Montréal Hospital Research Centre (CRCHUM), Montreal, Canada

Contributors:

MX designed the study's analytic strategy, conducted the analysis, and drafted the final manuscript. JS designed the original Montreal case–control study of lung cancer from which the data were derived. JS, and VH provided advice in analytic design.

Correspondence to:

Dr. Jack Siemiatycki

Université de Montréal Hospital Research Centre (CRCHUM),

850 Saint-Denis Street, 3rd Floor, S03-428

Montreal, Quebec H2X 0A9, Canada

E-mail: j.siemiatycki@umontreal.ca

What this paper adds

1. **What is already known about this subject?**
   - Lung cancer is the most common cause of cancer death among women.
   - Occupational risk factors for lung cancer have been extensively studied among male workers; however hardly any research has been conducted to study possible occupational risk factors for lung cancer among female workers.

2. What are the new findings?

Among the 22 occupational agents evaluated among working women, most exhibited no evidence of lung carcinogenicity. However for four of the agents, there were suggestive indications of possible excess risk. Because of limited numbers of exposed women in this study and very little prior research on working women, none of the results are conclusive.

3. How might this impact on policy or clinical practice in the foreseeable future?

These results have no immediate impact on policy, but they suggest that further research is needed on occupational risk factors among women.

Abstract

Objectives: To explore possible associations between selected occupational agents and lung cancer risk among women.

Methods: A population-based case-control study on lung cancer was conducted from 1996 to 2001 in Montreal, Canada. Cases were individuals diagnosed with incident lung cancer and population controls were randomly selected from electoral lists and frequency-matched to age and sex distributions of cases. Questionnaires on lifetime occupational history, smoking and demographic characteristics were collected during in-person interviews. As part of a comprehensive exposure assessment protocol, experts reviewed each subject's work history and assessed exposure to many agents. The current analysis, restricted to working women in the study, includes 361 cases and 521 controls. We examined the association between lung cancer and each of 22 occupational exposures, chosen because of their relatively high prevalences among these women. Each exposure was analyzed in a separate multivariate logistic regression model, adjusted for smoking and other selected covariates.

Results: There were few elevated OR estimates between lung cancer and any of the agents, and none were statistically significant, although the limited numbers of exposed women engendered wide confidence intervals.

Conclusions: There was little evidence to suggest that women in this population had experienced excess risks of lung cancer as a result of their work exposures. However, the wide confidence intervals preclude any strong inferences in this regard.

Keywords: Women; Exposure assessment; retrospective exposure assessment; Cancer; Epidemiology

Introduction

Worldwide, lung cancer is the most diagnosed malignant cancer and has been the principal cause of cancer death in women since 1987, surpassing mortality associated with breast cancer and colon cancer combined, which are the second and third leading cause of cancer death in women, respectively (21, 132). Although smoking is the primary risk factor for lung cancer among women, approximately 20% of women who develop lung cancer have never smoked (21). Among men, the most fruitful area for research on lung cancer risk factors apart from smoking has been the occupational environment. However, there has hardly been any evidence on occupational risk factors for lung cancer among women. This may be due to the historic roles of men and women and/or to the choices of researchers (85). Those few studies that have been conducted among women have suffered from inferior exposure assessment and statistical analysis methods compared to analyses of male data (85-87).

It is not sufficient to simply assume that knowledge gained through research on men will automatically apply to women. It is possible that women may have different carcinogenic responses to the same occupational exposures that have been investigated in men; this may be due to different genetic, hormonal or environmental co-factors, or it might be due to differences in the ways men and women are exposed. In addition, there are occupations that, in the past, were held primarily by women, and the exposures in these occupations would not have been evaluated for possible carcinogenicity in past male-based studies. It is thus important to devote attention to possible occupational risk factors for lung cancer among women. The aim of this present analysis is to explore the potential associations between selected occupational agents and occupations, and lung cancer risk among women using data from a case-control study of lung cancer conducted in Montreal, Canada.

Methods

Study design and subjects

A population-based lung cancer case-control study was conducted from 1996 to 2001, with the primary objective of studying the associations between a wide range of occupational, environmental and lifestyle factors and lung cancer etiology. A detailed description of the study has been published previously (95). Briefly, incident lung cancer cases were ascertained from all major hospitals in the Montreal area during the study period. Cases (n=1,203; 781 men and 422 women) included subjects aged 35-75 years, with histologically-confirmed incident primary lung cancer, who were Montreal residents and Canadian citizens. Controls (n=1,513; 936 men and 577 women) were also Montreal residents and Canadian citizens, and were randomly selected from the population-based electoral lists and were frequency matched to cases by age and sex. Response rates were 82% and 69% for female cases and controls, respectively. Ethics approval was obtained from the Institutional Review Boards of the Institut Armand-Frappier, McGill University, the Université de Montréal and all participating hospitals; all study participants provided informed consent.

This present analysis was limited to female subjects who had ever worked outside the home between the ages of 18 and 65 and within 35 years to 5 years prior to study recruitment; 361 cases and 521 controls satisfied these criteria.

Data collection and occupational exposure assessment

Data collection for cases and controls involved in-person interviews with trained interviewers. Next-of-kin provided responses for 34% of female cases and 5% of female controls that were too ill or had died. Information on subject's socio-demographic characteristics, lifestyle, medical history, family history of cancer, and detailed lifetime smoking history was collected using a structured questionnaire. In addition, a semi-structured questionnaire was administered to obtain subject's lifetime job history; namely, a detailed description for each job ever held by the subject, including information about the company, its products or main activities, the nature of the work site, subject's primary and subsidiary tasks and any additional information (i.e. use of protective equipment, equipment maintenance, activities of co-workers). Supplementary specialized questionnaires with detailed technical probing were also collected for many specific occupations (e.g. sewing machine operators, cooks, welders).

The detailed lifetime job history provided the starting point for the exposure assessment. A specially trained and uniquely experienced team of experts, comprised of chemists and occupational hygienists, and blinded to subjects' case control status, reviewed each job in each subject's lifetime job history. They coded the occupation according to the International Standard Classification of Occupations, Rev. 1968 (ISCO-68) (133). Further, based on a methodology developed by our team (97, 98), they translated each job of each subject into a list of potential exposures using a checklist of 294 agents, including broad chemical families, mixtures and specific chemicals. Exposure to each agent on our checklist was assessed by our three experts through a consensus process; one expert initially assessed exposure and one or two of the other experts reviewed the coding and discussed any points of disagreement. While the three experts had some pre-existing areas of expertise, they acquired sufficient expertise in all areas to opine about exposures in all areas. A previous inter-rater reliability study involving the same team of experts revealed high agreement between experts, with a kappa around 0.8. More information about the exposure assessment methods have been previously presented (98, 99). A worker was considered occupationally exposed to an agent if this exposure was thought to be above the background level of exposure in the general public (that is, in non-occupational settings). If the experts believed the worker was exposed to an agent, they indicated their degree of confidence in that belief by indicating if exposure was possible, probable, and definite. They also assigned semi-quantitative measures of concentration and frequency of exposure. Concentration was assigned on a low, medium, high scale; with the scale tailored to each specific agent. Frequency referred to the number of hours (up to 40 hours) a worker was exposed to an agent per week. It is noteworthy that the exposure assignment to an agent was based not only on subjects' job titles, but also on the unique characteristics of the workplace and the tasks reported. In addition to those indices of exposure, the reported work history gave information about the duration of each job, and thus the duration of exposures in each job.

Selection and parameterization of exposure variables for analyses

Although the study database contained information on possible exposure of each subject to 294 agents, for most of these agents, the numbers exposed in our study sample were so low that it would be almost hopeless to conduct analyses and report mainly uninformative results. In order to maintain a reasonable level of power to detect elevated risks and to provide meaningful confidence intervals on risk estimates, we restricted our analysis to a subset of the agents in our database, based on the prevalence and degree of exposure in our study population. The criterion we chose required that the agent should satisfy two conditions among the cases in our study sample: at least 30 cases exposed, where exposure was rated as "probable" or "definite", and at least 10% of those exposed cases had medium or high concentration of exposure and over 2 hours of exposure per week, and over 5 years duration. Using this criterion led to the earmarking of 22 agents for analysis. (The list of agents is shown in tables described below.)

For each agent we defined a cumulative exposure variable as follows:

$$\text{Cumulative Exposure} = \sum_{i=1}^{d} \frac{C_i}{25} \times \frac{F_i}{40} \qquad (1)$$

where *i* represents the *i*th year, *d* represents the total number of years exposed, *Ci* represents the concentration of exposure in year *i*, and *Fi* represents the number of hours exposed per week in year *i*. The values of *Ci* were transformed from (low, medium, high) to (1, 5, 25) as these were the approximate ratios of concentration that the experts had in mind when coding concentration. The formula for cumulative exposure assigns equal weights to the concentration and frequency of exposure through dividing each measure by their highest value.

The association between lung cancer and each selected occupational agent was analyzed using two metrics of exposure: a) any exposure and b) cumulative exposure above the agent-specific median value.

While the main purpose was to analyze possible associations with specific occupational exposures that can cut across different occupations and industries, we also took advantage of the opportunity to analyze possible associations between job titles and lung cancer. But given the rather small number of subjects who appear in each occupation category at the 3-digit level of the ISCO-68 classification, only six occupations satisfied the criterion of minimum 30 exposed cases. (The selected occupations are shown below.)

Statistical analysis

Separate unconditional logistic regression models were fitted to estimate the odds ratios (ORs) and 95% confidence intervals (CIs) between exposure to an occupational agent, or an occupation, and lung cancer risk among women. Each set of analytical models included the following covariates: age (continuous), ethnicity (French Canadian, Others) and smoking. Smoking was measured with the comprehensive smoking index (CSI), which is a continuous measure that represents subject's lifetime smoking history, incorporating smoking status, duration, time since cessation and intensity. The CSI has previously been validated for use in this Montreal lung cancer case-control study,(106) and as recommended by Leffondre,(106) we computed CSI discounting the subject's smoking history during two years prior to cancer diagnosis for cases and interview for controls, in order to reduce the chance of reverse causality bias. Because of the rather small numbers of cases and controls, we were reluctant to include too many covariates in the models. We conducted sensitivity analyses to explore the impact of including two other socioeconomic covariates (education and family income), as well as replacing the CSI covariate with three other smoking covariates (smoking status, pack-year, time since quitting smoking) in each regression model. We intentionally avoided adjustment for exposure to recognized occupational lung carcinogens as these exposures have not been empirically demonstrated to be female lung carcinogens and because the likely correlation in error of exposure assessment among different occupational exposures would lead to over-adjustment in multi-exposure models. The exclusion of other occupational exposures from the models is a conservative strategy; if an association is demonstrated, it indicates that there is an occupational risk factor, though it may require further research to ascertain whether it is the putative one we have analyzed, or another with which it is strongly correlated. It is not a strategy that induces an ostensible occupational association where there is none. All statistical analyses were conducted using R (Version 3.5.3) (134).

Results

Table 1 presents selected socio-demographic and lifestyle characteristics of 361 female lung cancer cases and 521 female controls in our Montreal population-based study. Median age was 60 years old for cases and 61 for controls. Compared to controls, lung cancer cases tended to have lower income and years of education, to use more often proxy respondents during interviews, to more likely be French Canadian and a current smoker, and to smoke more. Lifetime total working duration tended to be shorter in cases than in controls.

Occupational agents and lung cancer risk

Table 2 shows the 22 agents that satisfied our inclusion criteria, and it shows the main occupations in which each of these agents were assigned in our study population. In this table, the denominator for percentages exposed to each agent is not the total number of subjects (n=882 women); rather it is the total number of exposed jobs to a particular agent that were held by all subjects. Some occupational agents were widely spread across different occupations while others were concentrated in certain specific occupations. For example, exposure to cellulose, ammonia, toluene, synthetic adhesives, organic solvents, alkanes C5-17, aliphatic alcohols and many other agents can be described as widespread agents; while exposure to various textile dusts were mainly found in women working as sewers and embroiderers.

Table 3 presents the association between each selected occupational agent and lung cancer among women. Sensitivity analyses adjusting for additional socioeconomic covariates (education and family income) or replacing the CSI smoking covariate with three other smoking covariates (smoking status, pack-year, time since quitting smoking) in each regression model had no meaningful impact on our occupational exposure and lung cancer estimates; therefore results presented in the rest of the paper were adjusted for age, ethnicity and CSI only. For each occupational agent, two metrics of exposure were analyzed in relation to cancer risk: "*Any exposure*" and "*Cumulative exposure above the agent-specific median value*". For each agent the reference unexposed group comprised those subjects unexposed to that particular agent. While there were some point estimates of OR that departed from 1.0, none were significantly above 1.0. Using as a criterion that the point estimate should be 1.4 or greater and the lower confidence limit should be 0.8 or greater, the most suggestive associations were with formaldehyde (OR=1.4), cooking fumes (OR=1.5), toluene (OR=1.6) and MAHs (OR=1.9) among women with above median cumulative exposure to each of the agents. Three of the agents, cellulose, biocides and microorganisms, exhibited ORs that were below the null, particularly among women with above median cumulative exposure. The remaining agents did not exhibit remarkable OR point estimates, and the confidence limits were quite wide and the corresponding results are inconclusive.

To provide some perspective on these results among women, we conducted analogous analyses adjusting for the same set of covariates on the 730 male cases and 898 male controls who were part of the same Montreal study and who underwent the same exposure assessment protocol. The purpose was not to make a best estimate of each OR among males, but rather it was to compare the male and female results using the same models. In general, the results in men were rather similar to those found among women (male results not shown). That is, most OR estimates for the 22 agents analyzed here were close to the null. A positive association between above median cumulative exposure to synthetic adhesives and lung cancer was observed in men

(OR(95%CI) = 1.6(1.1-2.6)) but not in women (OR(95%CI) = 1.1(0.5-2.4)). Exposure to formaldehyde yielded quite different OR results by sex; for cumulative exposure above the median, the OR was 1.4 (0.8-2.4) among women and 0.5 (0.3-0.9) among men.

Occupations and lung cancer risk

There were only six occupations at the ISCO-68 3-digit level with at least 30 women exposed in our population. Table 4 presents the associations between these occupations and lung cancer risk, using as exposure metrics, ever exposure and more than 10 years of exposure, and using women who had never worked in the selected occupation as the reference category. Most of these job titles were not associated with lung cancer in women; however, a significantly increased risk was observed among women who had worked as waitresses, bartenders and related workers for over 10 years (OR = 2.7 (1.2-6.5).

Discussion

Most previous publications on occupational risk factors of lung cancer focused on male workers. Among those that included female workers and that reported results separately for women, most were analyses of risk by job titles. Among the very few studies that attempted to examine associations between occupational agents or chemicals and lung cancer,(49, 58, 60, 63, 66, 68-73, 135-137) most included only one or very few agents and some used rather simplistic and crude exposure assessment methods (i.e., self-reported exposure) and crude exposure metrics (ever/never exposure). The most commonly assessed agent was occupational exposure to environmental tobacco smoke (69-73).

In our study, there were no significantly elevated risks of lung cancer among female workers who were exposed to any of the selected occupational agents. We did, however, observe some suggestive elevated point estimates for cooking fumes, toluene, MAHs and formaldehyde. Cooking fumes can be both an occupational and domestic exposure. While there is no direct evidence that women exposed to cooking fumes at work are at risk of lung cancer, there is evidence, synthesized in a meta-analysis, that non-smoking Chinese women exposed to cooking fumes at home were at increased lung cancer risk (OR=1.7) (138). The relevance of these findings in Chinese women for the risks in Canadian women exposed to cooking fumes at work is not clear, since the circumstances of exposure (types of oils or fuels used, cooking temperatures, styles of cooking, types of substances cooked, kitchen design and ventilation) may be very different between the two populations (139). Toluene has been classified by the International Agency for Research on Cancer as a Group 3 agent (agents that cannot be classified as to its carcinogenicity to humans) (140); the evidence base regarding lung cancer consisted of solely one American study that demonstrated an elevated mortality in male shoe factory workers and a suggestively elevated mortality in female workers from the same cohort (141). From our own study in the Montreal population among the male subjects, there were suggestive excess lung cancer risks in men exposed to toluene and some other specific MAHs (142). Regarding formaldehyde exposure and risk of lung cancer there is an extensive body of evidence regarding risks among exposed male workers, including an analysis from our Montreal study (66), and the consensus is that there is no association (143). Although our motivation was to uncover evidence of carcinogenic agents among female workers, we found some agents that exhibited inverse associations. Exposure to cellulose, which is associated with handling of paper products and certain textiles, showed ORs below the null, as did the two broad and heterogeneous families that we call biocides and microorganisms. There may be some overlap between exposure to these agents and exposure to endotoxins, a class of pathogens that are found in some textile products and which have been hypothesized, inconclusively, of having protective actions against lung cancer (58, 135, 136).

Among the few occupation titles we were able to assess, the only remarkable result was a significant increased risk of lung cancer among women who had worked in waitressing for more than 10 years. This finding of an increased lung cancer risk among women working in waitressing jobs is consistent with findings from previous studies. Two European studies also found a more than twofold risk of lung cancer among waitresses (52, 144). It is likely that this increased risk is partially explained by waitresses' exposure to passive smoking at work. Before the ban of indoor smoking in public spaces in Montreal in 2016, it was common for servers to be routinely exposed to environmental tobacco smoke in restaurants and bars. We did not have data on each person's occupational exposure to environmental tobacco smoking. It is also possible that excess risks among waitresses may be related to exposure to cooking fumes.

Around 18% of women in our study had never worked outside the home during their adult lives and, compared to women who had worked, non-working women tended to be less educated, to have a lower household income, and to smoke more. Our strategy of excluding such women from these analyses may be questioned. But we believe that there may be such differences between women who have been in the workforce and those that have never worked that may confound the estimates of risk associated with all occupational exposures, and that may not be ascertained among the conventional covariates that we have collected. We believe that our strategy of restricting to working women allows us to make “fair” estimates of the ORs for occupational exposures. In fact we noted that women in our population who never worked had paradoxically higher lung cancer risks than women who had worked. Risk factors among those women will be explored in a subsequent paper.

The major strength of our study lies in the expert assessment of detailed lifetime occupational exposure for each participating woman. This allowed us to go beyond conventional job title analysis and provided a rare opportunity to explore the association between a large number of prevalent occupational agents and lung cancer risk in women. The use of expert exposure assessment method is generally considered as the best possible method for retrospective assessment of lifetime occupational exposures, when compared to other retrospective exposure assessment methods such as job title analysis or subject’s self-reported exposure or job-exposure matrices, typically derived in different populations (145). Further, the data collected on degree of exposure allowed us to create a metric of exposure above the median of cumulative exposure. Additional strengths include the collection of a wide variety of personal covariates, including notably a lifetime smoking history that allowed us to implement detailed control for potential confounders.

Some exposure misclassification is unavoidable in any occupational study including ours. Exposure misclassification could arise from subject’s self-reported job history, and/or during expert assessment of occupational exposures. These misclassifications were likely to be non-differential since it is unlikely that the validity of subjects’ self-reported job histories would differ based on both their exposures and lung cancer status and since experts were blinded to subject’s case control status when coding exposures.

Although our study is to date one of the largest studies that went beyond job titles to examine occupational agents in relation to lung cancer in women, the power to detect risks was still limited by sample size, by prevalence of exposure to each agent and by relatively low numbers of highly exposed women. Nonetheless, we believe these results contribute to the sparse empirical epidemiological body of evidence on potential occupational lung carcinogens in women. We intend to conduct analogous studies using larger samples of women workers.

Conclusion

There was little evidence to suggest that women in our population had experienced excess risks of lung cancer as a result of their work exposures. Additional evidence is required to discover occupational risk factors among women or to provide reassurance about lack of risk; such studies should be based on larger sample sizes and detailed exposure assessment methods.

Acknowledgments

Lesley Richardson and Marie-Elise Parent were involved in supervising fieldwork. Exposure assessment methods were expertly developed and implemented by Michel Gerin, Louise Nadon, Ramzan Lakhani, Denis Begin, and Benoit Latreille. A large number of research assistants and interviewers ensured the careful collection of data. The authors wish to express their gratitude for the collaboration of all Montreal area hospitals and all participants.

Funding

Funding for the original data collection was provided by the Canadian Institutes for Health Research (Grant Number: MOP-14704). Dr Jack Siemiatycki's research team was supported in part by the Canada Research Chairs programme and the Guzzo-SRC Chair in Environment and Cancer. Dr. Vikki Ho holds a Sex and Gender Science Chair in Cancer Research from the Canadian Institutes for Health Research. She is currently supported by the Cancer Research Society, Fonds de la recherche en santé du Québec and Ministere de l'Economie, de la Science et de l'Innovation du Quebec. Mengting Xu is currently supported by the Fonds de la recherche en santé du Québec and Institut de recherche Robert-Sauvé en santé et en sécurité du travail.

Competing Interests

None declared.

Data Sharing/Data Availability

Please email the corresponding author to request the relevant data.

Tables

Table 1. Characteristics of female workers in the Montreal population-based case-control study of lung cancer

Selected characteristics	**Cases (n=361)**	**Controls (n=521)**
Median age (years)	60	61
Median income (Cdn$)	30,340	36,343
Median years of education	9	12
Use of proxy respondents	31.6%	2.9%
Ethnicity		
French Canadians	77.6%	70.4%
Others	22.4%	29.6%
Smoking status		
Never smoker	6.4%	47.6%
Ex-smoker	19.4%	29.7%
Current smoker	74.2%	22.7%
Pack-years (mean among smokers)	52.5	37.1
Working duration		
1 to 10 years	22.7%	14.6%
11 to 20 years	31.0%	28.8%
21 to 30 years	31.3%	39.0%
More than 30 years	15.0%	17.7%

Table 2. Number of jobs in our female study population in which each of the 22 selected occupational agents was considered exposed, and the occupations (3-digit ISCO-68 job titles) in which they most frequently occurred

Agent [a]	Total No. of exposed jobs [b]	Prevalent occupations with exposure to selected agent (% of exposed jobs that were in top-ranked occupations) [c]
Cotton dust	388	Sewers and Embroiderers (47.9%); Launderers, Dry-Cleaners and Pressers (9.0%)
Wool fibers	116	Sewers and Embroiderers (53.4%)
Synthetic fibers	314	Sewers and Embroiderers (53.8%)
Polyester fibers	203	Sewers and Embroiderers (56.2%)
Treated textile fibers	465	Sewers and Embroiderers (47.3%); Launderers, Dry-Cleaners and Pressers (9.2%)
Cellulose	214	Dockers and Freight Handlers (18.2%); Bookkeepers and Cashiers (8.4%); Bookbinders and Related Workers (7.9%); Stock Clerks (6.1%); Salesmen, Shop Assistants and demonstrators (5.6%)
Ammonia	339	Waiters, Bartenders and Related Workers (20.9%); Maids and Related Housekeeping Service Workers (13.9%); Bookkeepers and Cashiers (6.5%); Charworkers, Cleaners and Related Workers (6.5%); Salesmen, Shop Assistants and demonstrators (5.9%)
Formaldehyde	374	Sewers and Embroiderers (28.1%); Waiters, Bartenders and Related Workers (17.6%); Cooks (7.2%)
Cooking fumes	345	Waiters, Bartenders and Related Workers (41.4%); Cooks (19.1%)
Isopropanol	321	Professional Nurses (16.5%); Maids and Related Housekeeping Service Workers (15.6%); Other Service Workers (11.8%); Waiters, Bartenders and Related Workers (8.1%)
Toluene	188	Waiters, Bartenders and Related Workers (25.0%); Shoe Cutters, Lasters, Sewers and Related Workers (11.2%); Bookkeepers and Cashiers (4.8%); Stenographers, Typists and Teletypists (4.3%); Sewers and Embroiderers (3.7%)

Agent [a]	Total No. of exposed jobs [b]	Prevalent occupations with exposure to selected agent (% of exposed women's jobs that occurred in these occupations) [c]
Synthetic adhesives	133	Shoe Cutters, Lasters, Sewers and Related Workers (15.8%); Bookbinders and Related Workers (10.5%); Primary Education Teachers (6.8%); Sewers and Embroiderers (6.0%); Dockers and Freight Handlers (6.0%)
Organic solvents	575	Maids and Related Housekeeping Service Workers (8.7%); Bookkeepers and Cashiers (8.2%); Stenographers, Typists and Teletypists (6.3%); Primary Education Teachers (4.9%); Waiters, Bartenders and Related Workers (4.7%)
Volatile Organic Liquids	756	Professional Nurses (7.5%); Maids and Related Housekeeping Service Workers (7.5%); Bookkeepers and Cashiers (6.9%); Stenographers, Typists and Teletypists (5.3%); Other Service Workers (5.3%)
Alkanes C5-C17	160	Maids and Related Housekeeping Service Workers (13.1%); Shoe Cutters, Lasters, Sewers and Related Workers (12.5%); Bookkeepers and Cashiers (6.3%); Dockers and Freight Handlers (5.6%); Sewers and Embroiderers (5.0%)
Aliphatic alcohols	445	Professional Nurses (11.9%); Maids and Related Housekeeping Service Workers (11.2%); Other Service Workers (8.8%); Primary Education Teachers (6.5%); Waiters, Bartenders and Related Workers (5.8%)
Aliphatic aldehydes	485	Sewers and Embroiderers (21.9%); Waiters, Bartenders and Related Workers (21.0%); Cooks (5.8%); Salesmen, Shop Assistants and demonstrators (4.9%)
PAHs	310	Sewers and Embroiderers (30.0%); Waiters, Bartenders and Related Workers (15.8%); Stenographers, Typists and Teletypists (3.9%)
Mononuclear aromatic hydrocarbons (MAHs)	102	Shoe Cutters, Lasters, Sewers and Related Workers (20.6%); Sewers and Embroiderers (5.9%); Printing Pressperson (4.9%); Bookkeepers and Cashiers (3.9%); Leather Goods Makers (3.9%)

Agent [a]	Total No. of exposed jobs [b]	Prevalent occupations with exposure to selected agent (% of exposed women's jobs that occurred in these occupations) [c]
Cleaning agents	769	Waiters, Bartenders and Related Worker (18.2%); Maids and Related Housekeeping Service Workers (14.2%); Other Service Workers (9.8%); Professional Nurses (7.7%)
Biocides	513	Maids and Related Housekeeping Service Workers (15.8%); Professional Nurses (13.5%); Other Service Workers (13.5%); Charworkers, Cleaners and Related Workers (8.6%)
Microorganisms	480	Professional Nurses (14.2%); Other Service Workers (12.7%); Primary Education Teachers (11.9%); Maids and Related Housekeeping Service Workers (8.1%); Secondary Education Teachers (7.3%)

[a] The ordering of agents in this and subsequent tables follows an ordering in our internal checklist which is intended to group agents on the basis of chemical and/or usage characteristics.

[b] This column presents the number of exposed jobs (any exposure) for each agent; there were a total of 3384 jobs in the sample of 882 subjects.

[c] Top exposed occupations to each agent are presented in a descending order based on the number of exposed jobs to that agent. Due to limited space, we only presented the top exposed occupations that, when aggregated, represented at least half of the total number of exposed jobs to each agent. For example, among 388 jobs exposed to cotton dust, 47.9% were sewers and embroiderers, and 9.0% were launderers, dry-cleaners and pressers.

Table 3. Occupational exposure to agents and lung cancer risk in women

Agent	Any exposure [a] (Ref: Unexposed)					Cumulative exposure above median [a,b] (Ref: Unexposed)				
	No. exposed cases	No. exposed controls	OR	LCL[c]	UCL[c]	No. exposed cases	No. exposed controls	OR	LCL[c]	UCL[c]
Cotton dust	79	108	1.1	0.7	1.7	38	54	1.3	0.7	2.3
Wool fibers	30	43	1.1	0.6	2.1	23	23	1.6	0.7	3.5
Synthetic fibers	64	109	0.9	0.6	1.5	30	55	1.0	0.5	1.8
Polyester fibers	45	77	0.9	0.6	1.6	27	39	1.4	0.7	2.7
Treated textile fibers	87	132	1.0	0.7	1.5	39	69	1.1	0.6	1.8
Cellulose	40	81	0.4	0.3	0.7	23	43	0.5	0.3	1.0
Ammonia	89	93	1.1	0.7	1.6	41	47	0.9	0.5	1.5
Formaldehyde	92	107	1.1	0.7	1.6	48	54	1.4	0.8	2.4
Cooking fumes	89	72	1.3	0.8	1.9	48	36	1.5	0.8	2.6
Isopropanol	78	93	1.0	0.7	1.6	23	48	0.6	0.3	1.1
Toluene	56	56	1.2	0.7	2.0	36	28	1.6	0.8	3.0
Synthetic adhesives	32	45	1.1	0.6	2.0	17	23	1.1	0.5	2.4
Organic solvents	131	168	1.1	0.8	1.6	70	86	1.2	0.7	1.8
Volatile Organic Liquids	147	215	0.9	0.7	1.3	71	108	0.9	0.6	1.4
Alkanes C5-C17	42	54	1.2	0.7	2.1	24	27	1.4	0.7	2.9
Aliphatic alcohols	102	137	1.0	0.7	1.5	36	69	0.7	0.4	1.2
Aliphatic aldehydes	116	138	1.0	0.7	1.5	66	69	1.2	0.7	1.9
Polycyclic aromatic hydrocarbons (PAHs)	64	97	0.8	0.5	1.2	47	49	1.0	0.6	1.7
Mononuclear aromatic hydrocarbons (MAHs)	32	30	1.6	0.8	3.1	15	15	1.9	0.8	4.6
Cleaning agents	151	181	1.0	0.7	1.4	85	92	1.1	0.7	1.7
Biocides	105	138	1.0	0.7	1.5	32	69	0.6	0.3	1.0
Microorganisms	80	156	0.9	0.6	1.3	26	78	0.6	0.3	1.0

[a] All models were adjusted for: age (continuous), ethnicity (French Canadian, Others) and comprehensive smoking index (CSI).

[b] Cumulative exposure defined in Formula (1).

[c] LCL: Lower 95% confidence limit. UCL: Upper 95% confidence limit.

Table 4. Exposure to 3-digit ISCO-68 job titles [a] and lung cancer risk among women

Occupations	Any exposure [b] (Ref: Unexposed)					> 10 years of exposure [b] (Ref: Unexposed)				
3-digit ISCO-68 job titles	**No. exposed cases**	**No. exposed controls**	**OR**	**LCI [c]**	**UCI [c]**	**No. exposed cases**	**No. exposed controls**	**OR**	**LCI [c]**	**UCI [c]**
3.21_Stenographers, typists and teletypists	62	92	1.1	0.7	1.7	31	51	1.1	0.6	2.0
3.31_Bookkeepers and cashiers	59	98	0.8	0.5	1.2	25	43	0.8	0.4	1.5
3.93_Correspondence and reporting clerks	33	30	1.4	0.7	2.6	14	9	1.6	0.6	4.4
5.32_ Waitresses, bartenders and related workers	56	33	1.4	0.8	2.5	32	12	2.7	1.2	6.5
5.40_Maids and related housekeeping service workers not elsewhere classified	34	39	1.1	0.6	2.1	12	11	1.3	0.5	4.0
7.95_Sewers and embroiderers	43	65	1.2	0.7	2.1	14	34	0.9	0.4	2.0

[a] These occupations were selected because they were relatively prevalent in our study sample.

[b] All models were adjusted for: age (continuous), ethnicity (French Canadian, Others) and comprehensive smoking index (CSI).

[c] LCL: Lower 95% confidence limit. UCL: Upper 95% confidence limit.

Supplementary material

In addition to the logistic regression models that were fitted to estimate the ORs and 95% CIs between exposure to an occupational agent and lung cancer risk among women in the published manuscript, we also conducted additional principal component analyses to identify major patterns of women's occupational exposure profile in this study population. We then explored lung cancer risks associated with each identified major pattern of exposure. The following texts were not retained in the final published manuscript due to word limit restriction.

Methods and results of the principal component analysis

Methods

We conducted principal component analysis (PCA) to identify major patterns of women's occupational exposure profiles (146, 147). PCA is suitable for data with a large number of potentially correlated exposure variables. It seeks to identify exposure patterns which explain a majority of the total variance in the exposure data. Each pattern is represented by a principal component (PC). In this analysis, we selected major patterns of women's exposure profile based on Kaiser's criterion (148), and explored these patterns and their associations with lung cancer using a logistic regression model; each included PCs were mutually adjusted for each other in the model and in addition, we also adjusted for the same set of covariates as in the individual agent analysis.

Results

The first 10 PCs were selected as each PC had an eigenvalue greater than 1 and altogether they contributed to explaining 70.4% of the total variance in our exposure data. We observed that a PC representing a pattern with heavy presence of cellulose, was associated with lung cancer with an OR (95%CI) of 0.8 (0.7-0.9). In addition, another PC representing a pattern with heavy presence of synthetic adhesives and mononuclear aromatic hydrocarbons was borderline associated with lung cancer (OR (95%CI) = 1.2 (1.0-1.4)).

Chapter 7: Manuscript 2

The published manuscript can be found online at:

doi: 10.1093/annweh/wxac008

ANNALS OF WORK EXPOSURES AND HEALTH

FORMERLY KNOWN AS The Annals of Occupational Hygiene

Ann Work Expo Health. 2022 Jul; 66(6): 728–740. PMCID: PMC9250288

Published online 2022 Mar 8. doi: 10.1093/annweh/wxac008 PMID: 35258522

Concordance of Occupational Exposure Assessment between the Canadian Job-Exposure Matrix (CANJEM) and Expert Assessment of Jobs Held by Women

Mengting Xu, Vikki Ho, Jerome Lavoue, Lesley Richardson, and Jack Siemiatycki

Concordance of Occupational Exposure Assessment between the Canadian Job-Exposure Matrix (CANJEM) and Expert Assessment of Jobs Held by Women

Mengting Xu[1,2], Vikki Ho[1,2], Jerome Lavoue[1,3], Lesley Richardson[1], Jack Siemiatycki[1,2]

1. CRCHUM (Centre de recherche du CHUM), University of Montreal, Montreal, Quebec, Canada
2. Department of Social and Preventive Medicine, University of Montreal, Montreal, Quebec, Canada
3. Department of Environmental and Occupational Health, University of Montreal, Quebec, Canada

Contributors:

<u>MX</u> designed the study's analytic strategy, conducted the analysis, and drafted the final manuscript. JS and JL are the PIs of the CANJEM database and provided the CANJEM database for use in this project. JS designed the original Montreal case–control study of lung cancer from which the data were derived. JS, VH, JL, and LR provided advice in analytic design.

Correspondence to:

Dr. Jack Siemiatycki

Université de Montréal Hospital Research Centre (CRCHUM),

850 Saint-Denis Street, 3rd Floor, S03-428

Montreal, Quebec H2X 0A9, Canada

E-mail: j.siemiatycki@umontreal.ca

What's important about this paper

Job-exposure matrix (JEM) is an attractive tool to use in population-based case-control studies of occupational risk factors, but the validity of exposure estimates derived from a JEM remains uncertain. We compared the exposure assessments regarding 69 occupational agents generated by using the Canadian job-exposure matrix (CANJEM) with exposure assessments generated by experts, for 3403 jobs held by Canadian women. Among female workers, concordance between CANJEM and expert assessment varied greatly by agent; our results indicate which agents provide exposure assessments that mimic those obtained with expert assessment. Potential users of CANJEM can benefit from knowing for which of the many occupational agents CANJEM performs well and for which ones it performs poorly. Creators of other JEMs may gain insights into some decisions that they can make in determining the characteristics of their JEM.

Abstract

Objectives: To compare the exposure data generated by using the Canadian job-exposure matrix (CANJEM) with data generated by expert assessment, for jobs held by women.

Methods: We selected 69 occupational agents that had been assessed by experts for each of 3403 jobs held by 998 women in a population-based case-control study of lung cancer. We then assessed the same agents among the same jobs by linking their occupation codes to CANJEM and thereby derived probability of exposure to each of the agents in each job. To create binary exposure variables, we dichotomized probability of exposure using two cutpoints: 25% and 50% (referred to as CANJEM-25% and CANJEM-50%). Using jobs as units of observation, we estimated the prevalence of exposure to each selected agent using CANJEM-25% and CANJEM-50%, and using expert assessment. Further, using expert assessment as the gold standard, for each agent, we estimated CANJEM's sensitivity, specificity, and kappa.

Results: CANJEM-based prevalence estimates correlated well with the prevalences assessed by the experts. When comparing CANJEM-based exposure estimates with expert-based exposure estimates, sensitivity, specificity, and kappa varied greatly among agents, and between CANJEM-25% and CANJEM-50% probability of exposure. With CANJEM-25%, the median sensitivity, specificity, and kappa values were 0.49, 0.99, and 0.46, respectively. Analogously, with CANJEM-50%, the corresponding values were 0.26, 1.00, and 0.35, respectively. For the following agents, we observed high concordance between CANJEM-based and expert-based assessments (sensitivity ≥0.70 and specificity ≥0.99): fabric dust, cotton dust, synthetic fibers, cooking fumes, soldering fumes, calcium carbonate, and tin compounds. We present concordance estimates for each of 69 agents.

Conclusions: Concordance between CANJEM and expert assessment varied greatly by agents. Our results indicate which agents provide data that mimic best those obtained with expert assessment.

Keywords: retrospective exposure assessment; expert assessment; exposure assessment; JEM; CANJEM; women; exposure misclassification; concordance

Introduction

Retrospective assessment of occupational exposures is a major challenge in occupational epidemiology. In the particular context of community-based case-control studies of diseases with long latency, several methods have been used to assess past occupational exposure to specific agents. These include: expert assessment of occupational exposures based on self-reported job histories, and job-exposure matrices (JEMs) (76, 98, 149). Each method has advantages and disadvantages, with trade-offs between validity and feasibility (80, 83). Well-trained and experienced experts would have access to similar sources of information as creators of a JEM, but they are able to take into account the idiosyncracies of each evaluated job. Expert assessment of occupational exposures, based on worker's description of the job, generally would provide more valid estimates of exposure than a JEM. However, expert assessment is a very costly and time-consuming approach (76, 150). Using a well-founded JEM is attractive since it is usually inexpensive to implement, but the validity of exposure estimates derived from a JEM and assigned to a particular worker is uncertain, for at least two distinct reasons. First, the validity of the summary data provided for an occupation depends on the quality of the expertise and the data that went into the creation of the JEM. Second, a JEM provides summary data for workers who share a given occupation title, and thus is unable to account for within-occupation exposure heterogeneity.

Although occupational epidemiologic studies have contributed greatly to identifying workplace risk factors, these findings were mostly identified in male workers (86, 87). In addition to under-representation of women in past studies, exposure assessment of women's jobs also tended to be less detailed than that of men's jobs (85-87). Due to the gender difference in occupational profiles, some occupational exposures may only be prevalent in women's jobs and hence less likely to have been evaluated in past occupational studies (94, 151-153). Thus, it is important to evaluate women's occupational exposures with quality exposure assessment methods.

A team of experts assessed occupational exposures, derived from lifetime occupational histories of women and men, reported within a population-based case-control study of lung cancer in Montreal, Canada (97, 98). The expert-assessed exposure data in this study, along with three other studies that shared a common expert-assessment protocol, was later used to develop the Canadian job-exposure matrix (CANJEM). To evaluate the validity of CANJEM for women, we compared exposure assessment of a selected list of 69 agents, using CANJEM and our expert assessments in jobs held by women in our

Montreal case-control study of lung cancer. Specifically, we used the following measures of concordance: similarity of the estimated prevalence of exposure, and the sensitivity, specificity, and kappa of the CANJEM-derived estimates vis-à-vis expert-derived estimates.

Methods

Data sources

<u>The Montreal case-control study</u>

A population-based case-control study was conducted from 1996 to 2001 in Montreal, Canada to study occupational, environmental and lifestyle factors and lung cancer risk (154). Eligible participants were Montreal residents aged 35 to 75 years; incident lung cancer cases ascertained across all major Montreal area hospitals and randomly-sampled population controls, frequency-matched to cases by age group and sex. Trained interviewers conducted in-depth interviews with the participants and elicited detailed job histories and information on potential confounding factors. Lifetime job history was collected using a semi-structured questionnaire concerning each job including the company, its products and main activities, the nature of the worksite, primary and subsidiary tasks, the use of protective equipment, machine maintenance, and activities of coworkers. For certain specific occupations (e.g., sewing machine operators, cooks, nurses), a supplementary questionnaire with detailed technical questions was used (97, 98, 154). The Montreal case-control study included 430 female cases and 568 female controls who held at least one job; resulting in a total of 3403 jobs. Ethics approval was obtained from Institutional Review Boards of Institut Armand-Frappier, McGill University, Université de Montréal and all participating hospitals; all study participants provided informed consent.

<u>Expert assessment data on occupational exposure</u>

A team of specially trained and experienced experts including chemists and occupational hygienists, blinded to participants' case-control status, reviewed each job. Each occupation was coded to the International Standard Classification of Occupations, Rev. 1968 (ISCO-68) (133), and the 1971 Canadian Classification and Dictionary of Occupations (CCDO-71) (155). Further, through a consensus process, the experts assessed potential exposures in each job using a checklist of nearly 300 agents. Potential exposure was categorized as possible, probable, or definite. Semi-quantitative measures of intensity and frequency of exposure were assigned. The assessment was based not only on participant's occupation, but also on the unique characteristics of the workplace and the tasks reported (98, 156). This expert assessment was the “gold standard” with which we compared the assessments provided by CANJEM. Although expert assessment does not necessarily represent the absolute truth of participants' past occupational exposure, we believe it is closer to the truth than any JEM can be.

CANJEM data on occupational exposure

CANJEM is a general population JEM built from expert assessment data of 31,673 jobs held by 8,760 participants in four Montreal case-control studies that used the same expert assessment approach (157, 158), including the lung cancer case-control study described above. CANJEM was created to provide other researchers a cost-effective alternative to the expert approach.

CANJEM is comprised of three axes: occupation code, time period and agent. CANJEM offers multiple occupation classifications (157, 158). For the present project, we used an international classification system, ISCO-68, that is used in many countries and that is included in CANJEM. ISCO-68 has a 5-digit resolution which can also be meaningfully analyzed at 2-digit or 3-digit resolutions. The time axis allows customized time periods ranging from the year 1930 to 2005. The agent axis includes 258 agents. Each unique combination of those three axes defines a cell in CANJEM. Each cell (i.e. for a specific occupation-time-agent combination) includes information on the probability of exposure; when the probability is above 0%, it also provides estimates of confidence (possible, probable, definite), intensity (low, medium, high), and frequency (number of hours per week).

Linking the Montreal case-control study to CANJEM

We linked each of the 3403 jobs held by our female study participants to CANJEM. We chose the 1950-2005 time period in CANJEM, which corresponded to the period of jobs held by most of our study participants. We first linked the jobs at the highest resolution (5-digit) of ISCO-68 occupation codes. For jobs that could not be linked at the 5-digit resolution, we used the 3-digit resolution. Using this strategy, we linked 99.95% of all jobs in our study population to CANJEM. The remaining un-linkable jobs were excluded from this analysis.

Selection of agents and selection of CANJEM probability of exposure cutpoint

Of the 258 agents in CANJEM, we selected for the present analysis those agents that had been listed as probably or definitely exposed in at least 30 of the 3403 jobs held by women in our study population. Sixty-nine agents satisfied this criterion.

We compared the estimated prevalence of exposure status to each agent and evaluated the concordance between the CANJEM-derived assessment and the expert-derived assessment. The expert consensus assessment method assigned a binary exposure status (exposed / unexposed) to an agent within a given job. CANJEM provides the user with a probability of exposure to an agent, and users can select a cutpoint of probability of exposure to dichotomize the exposure status. For both methods, in order to reduce the influence of trivial exposures, ever exposure to an agent was restricted to jobs with a probable or definite exposure. Jobs with a possible exposure were considered as unexposed. Exposed jobs were required to have a minimum exposure frequency of 0.5 h/week, and a frequency-weighted intensity corresponding to at least 2h/week at low intensity.

The use of any JEM, including CANJEM, would inevitably involve some degree of exposure misclassification, regardless of the choice of probability of exposure cutpoint. It is difficult to choose a cutpoint, partly due to the lack of a real gold standard, and partly because the most suitable cutpoint would likely differ among agents. We used two probability cutpoints to distinguish exposed from unexposed jobs, 25% probability of exposure (CANJEM-25%) and 50% probability of exposure (CANJEM-50%). These two cutpoints have often been used in previous studies using JEMs (111-113). Jobs with a probability of exposure above or equal to the chosen cutpoint were classified as "exposed" and jobs below the cutpoint were classified as "unexposed". Hereafter, the methods being compared in this paper will be referred to as "expert assessment", "CANJEM-25%" and "CANJEM-50%".

Statistical analyses

To evaluate the performance of CANJEM in replicating expert-assigned exposure, sensitivity and specificity were calculated for each agent. In our study, the expert assessment is the reference "gold standard" method and CANJEM is the proxy that aims to mimic expert assessment. We present information on agents' sensitivity and specificity values in various formats, sometimes by citing the numerical estimates of those parameters and sometimes by referring to the estimates as "high" or "low". There is no objective criterion or convention to categorize sensitivity or specificity as "high" or "low"; the two concordance measures should be interpreted with each other into account. Cohen's kappa was also calculated for each agent to evaluate the exposure concordance not due to chance (84). Kendall's τ coefficients were calculated to measure the correlation between expert-assessed prevalence of exposure to that assessed by CANJEM-25% and CANJEM-50%. Further, we plotted the empirical cumulative distribution functions (ECDF) of the prevalence for each approach. For semantic simplicity, we refer to percent of all jobs in our study that were considered exposed to the agent as "prevalence of the agent". The units of observation used to estimate prevalence and concordance of each agent were the 3403 jobs.

We also examined the loss of statistical power to detect a risk due to exposure misclassification when using a proxy for a gold standard method. We assumed a hypothetical case-control study of 1000 cases and 1000 controls, where the association between exposure to an occupational agent and lung cancer was fixed to be two-fold (odds ratio (OR)=2.0). We simplified the hypothetical situation to posit one job per participant. To demonstrate the impact under varying exposure misclassification scenarios, we selected combinations of three levels of the following parameters: true prevalence of exposure (2%, 5%, and 15%), sensitivity (0.30, 0.50, and 0.80), and specificity (0.90, 0.95, and 0.99). The impact of using an exposure assessment method with imperfect sensitivity and specificity on estimated prevalence of exposure and on the OR can be described in the following formulas (114, 115):

$$P^* = PU + (1-P)(1-V) \qquad (1)$$

$$R^* = \frac{[URP + (1-V)(1-P)][(1-U)P + V(1-P)]}{[UP + (1-V)(1-P)][(1-U)RP + V(1-P)]} \qquad (2)$$

Where:

P = Prevalence of exposure estimated by the gold standard

U = Sensitivity

V = Specificity

R = Hypothetical true OR

P^*= Misclassified prevalence of exposure after exposure misclassification

R^*= Misclassified OR after exposure misclassification

We first calculated the power of the "gold standard" expert assessment to detect a two-fold risk (OR = 2.0) under varying levels of exposure prevalence (159, 160). We then calculated the power under varying scenarios of exposure misclassification by replacing *P* with P^*, and *R* with R^*. The proportional difference between the two calculated powers is what we call the loss in statistical power due to exposure misclassification.

Results

Comparison of estimated prevalence of exposure between CANJEM vs. expert-assessment

In Table 1, our selected 69 agents are grouped by their chemical nature and by use, and sorted in descending order based on the prevalence of exposure within each grouping of agents. Estimated prevalence of exposure to these agents ranged from 0.9% to 21.5% based on expert assessment; from 0.0% to 23.2% based on CANJEM-25%, and from 0.0% to 20.8% based on CANJEM-50%. "Cleaning agents" was the most prevalent agent across all exposure assessment methods. For most agents, the prevalence estimated, using either cutpoint of CANJEM, correlated quite well with that estimated by the experts. The Kendall's τ coefficients, between expert-assessed prevalence and prevalence assessed by CANJEM-25% and CANJEM-50%, were 0.62 and 0.53 respectively. Overall, CANJEM-25% tended to produce similar or slightly higher than expert-assessed prevalence and CANJEM-50% tended to produce lower than expert-assessed prevalence (Appendix 1). For some agents, estimated prevalence was quite similar between the two versions of CANJEM (e.g., fabric dust, cooking fumes, cleaning agents) while some other agents dropped sharply in estimated prevalence when switching from CANJEM-25% to CANJEM-50% (e.g., organic solvents, ashes, polyester fibres).

Concordance between CANJEM vs. expert-assessment – sensitivity, specificity, and kappa

Sensitivity, specificity, and kappa values are presented in Table 1 for each agent, separately for CANJEM-25% and CANJEM-50%. Tables 2 and 3 present two complementary summarizations of the agent-by-agent results shown in Table 1.

Across all agents shown in Table 1, with CANJEM-25%, the median sensitivity, specificity, and kappa values were 0.49, 0.99, and 0.46, respectively. Analogously, with CANJEM-50%, the median sensitivity, specificity, and kappa values were 0.26, 1.00, and 0.35, respectively. As we move from CANJEM-25% to CANJEM-50%, we observed a decrease in sensitivity and an increase in specificity for most agents. This was particularly pronounced for some agents (polyester fibres, nylon fibres, wool fibres, ashes, cosmetic talc, and anaesthetic gases). The kappa values tended to be greater with CANJEM-25% than with CANJEM-50%, but not always. Particularly low kappa values (<0.10) were observed using both versions of CANJEM for silk fibres, acrylic fibres, ozone, propane combustion products and methanol. Using either CANJEM cutpoint, there was good concordance in sensitivity and specificity between CANJEM and expert assessment for several agents (e.g., fabric dust, cotton dust, synthetic fibres, cooking fumes, soldering fumes, calcium carbonate, tin compounds, etc.). Concordance was poor using both CANJEM cutpoints for around a quarter of our selected agents. For the other agents, the degree of concordance differed depending on which CANJEM cutpoint was used. Figure 1 provides a visual demonstration of sensitivity and specificity for each agent when comparing CANJEM-25% to expert assessment. Given the generally low prevalences of exposure that we experience in this study, the specificity of CANJEM-based exposure estimates vs Expert assessment was rarely lower than 0.90. By contrast, the sensitivity was rarely higher than 0.90. It is not meaningful to compare absolute level of sensitivity with absolute level of specificity.

Table 2 shows the joint categorical distribution of sensitivity and specificity values among the 69 agents, with each of the two CANJEM assessments vs. the expert assessment. Most agents had relatively low sensitivity and quite high specificity using both versions of CANJEM. Three of 69 agents (fabric dust, calcium carbonate, and soldering fumes) had both high sensitivity (0.70-0.89) and high specificity (≥0.99) in both CANJEM versions; tin compounds had high sensitivity and specificity in CANJEM-25%, while cooking fumes had high sensitivity and specificity in CANJEM-50%. A much larger proportion of agents had high sensitivity (0.70-0.89) but lower specificity (0.90-0.98) in CANJEM-25% than in CANJEM-50% (17 percent vs. 4 percent); while a much smaller proportion of agents had lower sensitivity (<0.70) but high specificity (≥0.99) in CANJEM-25% than in CANJEM-50% (58 percent vs. 84 percent).

Table 3 shows selected summary concordance statistics between CANJEM-based assessment and expert assessment of exposure across the agents. Although agent-specific statistics vary, in general, CANJEM-25% tended to mimic expert-assessed prevalence of exposure more closely than did CANJEM-50%. Specificity values were generally over 0.90 using both CANJEM cutpoints; with CANJEM-50% reaching a specificity of 1.00 for around half of the agents. CANJEM-25% tended to produce higher sensitivity values and higher kappa values on average than CANJEM-50%.

Hypothetical scenarios of exposure misclassification and their impact on power loss in epidemiological studies

As one measure of the impact of measurement error involved in the use of CANJEM instead of the “gold standard” expert assessment, we calculated and compared the loss in statistical power associated with the use of CANJEM. Table 4 presents various levels of loss of power to detect a two-fold risk under different scenarios of exposure misclassification in a hypothetical case-control study of 1000 cases and 1000 controls. The table also shows the impact of misclassification on estimates of prevalence and OR, and illustrates the known phenomenon that non-differential misclassification of binary exposure status will lead to attenuated ORs. We chose exposure misclassification scenarios that mostly represent patterns of misclassification observed in our own study; three levels of prevalence that are common for occupational exposures; and a hypothetical true OR of 2.0 which is plausible in occupational exposure and cancer research. Departure from a sensitivity of 1.00 will have the effect of artificially reducing the estimated prevalence; departure from a specificity of 1.00 will artificially increase the estimated prevalence. When there is imperfect sensitivity and imperfect specificity, there are countervailing effects: the net effect may be to increase or decrease the estimated prevalence. The extent of the impact depends on the true prevalence. Table 4 shows how these countervailing pressures play out under various plausible scenarios. At lower true prevalence a smaller gain in specificity, compared to a larger gain in sensitivity, tended to yield better power and to produce an OR estimate closer to the true OR. As true prevalence increases, an improvement in sensitivity would have more influence on improving the OR estimates and power. Within the same level of exposure misclassification, there was less dilution of the ORs toward the null and less power loss when the true prevalence was relatively high (i.e., 15% vs. 2%).

Discussion

On the basis of our CANJEM vs expert assessment comparison, it is not straightforward to determine which agents assessed by CANJEM have sufficient validity to be included in epidemiologic analyses of risk. Ideally, we would hope that the CANJEM-derived estimates of exposure are perfectly correlated with the expert assessments (i.e., sensitivity and specificity and kappa all equal to 1.0), but this was never expected and it does not occur for any agent. The further the values of these parameters drift from the ideal of 1.0, the less reliable is the CANJEM-derived estimate as a proxy for the expert assessment. If the values of concordance parameters exceed 0.0, as they do for all the agents in our study, there is some opportunity for the CANJEM-derived estimates to absorb sufficient information to reflect a real risk, should there be one. We have not drawn any arbitrary cutpoints on the scales of sensitivity, specificity, or kappa to designate agents with high concordance, but we do list those agents that had the highest concordance values, and that consequently have the best chance that the CANJEM-derived estimates would replicate the expert assessments.

We compared the prevalence and concordance measures for occupational exposure to 69 agents in jobs held by women using CANJEM with those derived using an expert assessment approach. For most agents, the prevalence estimated using either CANJEM-25% or CANJEM-50% cutpoint correlated well with the prevalence based on the expert assessments. Using expert assessment as the gold standard, CANJEM performed well under both probability of exposure cutpoints in terms of sensitivity, specificity, and kappa for several agents, including fabric dust, cotton dust, synthetic fibers, cooking fumes, soldering fumes, calcium carbonate, and tin compounds; for many other agents, the CANJEM-based assessments did not perform well in one or both versions of CANJEM.

As we reported elsewhere (161), the most prevalent occupations in our sample of Montreal women include various office jobs, service industry jobs such as waitresses, maids and housekeepers, and jobs in the textile industry such as sewers and embroiderers. Many of the agents with good agreement between expert assessment and CANJEM were also prevalent exposures in those occupations (e.g., fabric dust, cotton dust, synthetic fibers, and cooking fumes). A few previous studies have compared occupational exposure assessments between JEMs and expert assessments (162-164). The occupational agents examined included asbestos (163, 164), diesel motor emissions (164), crystalline silica (164), polycyclic aromatic hydrocarbons (163), welding fumes (163), aromatics (162), chlorinated aliphatics (162), lead (162), formaldehyde (162) and insecticides (162). The kappa values for a JEM-based exposure assessment compared to expert assessment varied somewhat across these comparisons, but they were almost all below 0.50. It is not clear whether these observations reflected characteristics of the agent under investigation, or the specific characteristics of the JEM and experts being compared. Similarly, we observed kappas below 0.50 for crystalline silica and for polycyclic aromatic hydrocarbons; formaldehyde and lead had a slightly higher kappas (around 0.50) in our analysis than in previous studies. A majority of our selected 69 agents had kappas below 0.50.

The impact of the observed misclassification would be a reduction in power to detect real relative risks. The hypothetical exercise that we undertook demonstrated that the power loss could be quite substantial under a variety of hypothetical scenarios of exposure misclassification, exposure prevalence, and relative risk. This exercise highlighted the importance of choosing agents for analysis that manifest relatively high values of sensitivity and specificity. Flegal et al. (115) proposed that if the prevalence of exposure is less than $1/(1+\sqrt{R})$, which translates to 41% prevalence given an OR of 2.0, specificity values would tend to have a larger impact than sensitivity on biasing the risk estimates. The choice of occupational agents to examine using CANJEM should prioritize agents with high specificity; since prevalence of exposure to many agents tends to be low, accompanied by a low to moderate effect size.

In contrast with previous attempts to evaluate performance of JEMs, we focused on female workers, a neglected but growing part of the workforce. In order to ascertain whether the results we observed among women were relevant only to women, we also conducted a supplementary analysis among male workers that was analogous to the one reported in this paper, though less extensive. This analysis among men was restricted to the following five agents that were also part of the present analysis among women: fabric dust, cotton dust, cleaning agents, cooking fumes, and organic solvents. We compared kappas of CANJEM against expert assessment for those five agents between women and men. Using either CANJEM cutpoint, kappas calculated using men's data are over 0.10 lower than those calculated using women's data for fabric dust and cotton dust, close to 0.20 lower for cleaning agents, over 0.10 higher for organic solvents, and about the same for cooking fumes.

We presented CANJEM's ability to replicate expert assessment using our 2-step approach to linkage of occupation titles, two different cutpoints in the probability of exposure in CANJEM, and our definition of exposure status to an agent, but results may vary if these settings differ. Although the agents we assessed were the most prevalent ones in our sample of women's jobs, a majority of our participants were unexposed to most of the studied agents. Inclusion of a large number of unexposed jobs leads to an overall high specificity. CANJEM generally performs better for more broadly defined exposures (e.g., fabric dust) than more specific exposures (e.g., silk dust), although there were some exceptions to this generalization. We presented in this paper the concordance between CANJEM vs. our expert assessment in assessment of the binary exposure status variable (exposed / unexposed); it is likely that the concordance between the two methods would be lower in assessing more refined exposure parameters (e.g., intensity of exposure) as there would be more opportunity for exposure misclassification to occur.

We consider that our comparison of CANJEM-derived estimates with expert assessment method represents a "best-case scenario" for CANJEM-derived estimates. Namely, the CANJEM-derived estimates were not entirely independent of the expert assessments, as CANJEM was built from the database of expert assessments conducted as part of four case-control studies in Montreal, one of which was the same lung cancer case-control study linked to CANJEM in this paper. But we are unable to estimate how much lower the sensitivity and specificity would be if we applied CANJEM to a different population for which expert assessments of exposure were available. The generalizability of our findings to studies that might use CANJEM in other female populations would depend on how similar occupational exposure profiles are in those populations to that of the Canadian female population, conditional on occupation. Notwithstanding the theoretical, financial and feasibility advantages of using a JEM instead of expert assessment, it might not be a suitable exposure assessment tool for some agents (162-164). Like any JEM, CANJEM cannot assign distinct exposure information based on idiosyncratic features of a particular job or worker that are not captured by distinct occupation codes. Agents with low probability of exposure across many occupations would generally have poorer JEM performance compared to agents with high probability of exposure in a few occupations. For example, although cooking fumes and organic solvents displayed overall similar levels of prevalence in our study, the sensitivity, specificity, and kappa values were much higher for cooking fumes than they were for organic solvents. Exposure to cooking fumes was predominantly present with high probability of exposure in a few occupations including waitresses/bartenders and cooks, whereas exposure to organic solvents was spread with lower probabilities of exposure among many occupations. Users of CANJEM need to carefully evaluate whether the agents are widespread or concentrated in a limited number of occupations and how CANJEM performed in their study.

Conclusion

In conclusion, among female workers, concordance between CANJEM and expert assessment varied greatly by agent. Our results indicate which agents provide data that mimic best the exposure assessments that would be obtained with expert assessment.

Funding

Funding for the Montreal case-control study of lung cancer data collection was provided by the Canadian Institutes for Health Research (Grant No: MOP-14704). This work was funded in part by the GRePEC program, a joint initiative of the Cancer Research Society, the Quebec Ministry of Economy, Science and Innovation and the Fonds de Recherche du Québec – Santé (Grant No: 16264). JS's research team was supported in part by the Canada Research Chairs programme and the Guzzo-SRC Chair in Environment and Cancer. VH holds a Sex and Gender Science Chair in Cancer Research from the Canadian Institutes for Health Research (Grant No: GS3-171364) and is the recipient of a Research Scholar J2 Award from the Fonds de recherche du Québec – Santé (FRQS). She was supported by the Cancer Research Society's GRePEC salary award in collaboration with the FRQS and Ministere de l'Economie, de la Science et de l'Innovation du Quebec during the conduct of this work. MX received doctoral training scholarships from the Fonds de la recherche en santé du Québec and Institut de recherche Robert-Sauvé en santé et en sécurité du travail. The funders had no role in study design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

The authors declare no conflicts of interest.

Data availability

The data underlying this article can be shared on reasonable request to the corresponding author.

Tables

Table 1. Concordance between CANJEM and Expert assessments of exposure for each of 69 occupational agents[a] in each of 3403 jobs[b] held by women in the Montreal lung cancer case-control study.

Agents[a]	Expert assessment	CANJEM-25%[c]				CANJEM-50%[c]			
	Percent of jobs exposed	Percent of jobs exposed	Sensitivity[d]	Specificity[d]	Kappa	Percent of jobs exposed	Sensitivity[d]	Specificity[d]	Kappa
Organic solids									
Fabric dust	14.8%	12.5%	0.77	0.99	0.81	11.0%	0.71	0.99	0.78
Cotton dust	12.0%	11.7%	0.79	0.98	0.78	11.2%	0.77	0.98	0.77
Synthetic fibres	9.2%	11.0%	0.79	0.96	0.69	9.4%	0.71	0.97	0.68
Cellulose (paper fibre)	6.5%	6.3%	0.60	0.97	0.58	2.4%	0.30	1.00	0.42
Polyester fibres	5.9%	9.6%	0.79	0.95	0.57	1.3%	0.13	0.99	0.19
Nylon fibres	4.1%	8.6%	0.76	0.94	0.46	0.6%	0.04	1.00	0.05
Wool fibres	3.3%	9.2%	0.84	0.93	0.42	1.5%	0.20	0.99	0.26
Rayon fibres	2.0%	0.9%	0.15	0.99	0.19	0.0%	0.00	1.00	0.00
Organic dyes and pigments	1.8%	1.9%	0.49	0.99	0.47	0.9%	0.30	1.00	0.39
Silk fibres	1.8%	0.6%	0.07	1.00	0.09	0.0%	0.00	1.00	0.00
Flour dust	1.7%	2.5%	0.67	0.99	0.52	0.7%	0.32	1.00	0.43
Sugar dust	1.5%	0.8%	0.37	1.00	0.48	0.6%	0.35	1.00	0.48
Starch dust	1.5%	1.8%	0.42	0.99	0.38	0.3%	0.19	1.00	0.32
Leather dust	1.4%	1.3%	0.63	1.00	0.63	1.3%	0.63	1.00	0.63
Plastic dusts	1.1%	1.2%	0.36	0.99	0.33	0.3%	0.19	1.00	0.30
Flax fibres	1.0%	0.5%	0.11	1.00	0.15	0.0%	0.00	1.00	0.00
Acrylic fibres	1.0%	0.3%	0.06	1.00	0.09	0.0%	0.00	1.00	0.00
Carbon black	0.9%	0.9%	0.56	1.00	0.56	0.3%	0.22	1.00	0.32
Inorganic solids									
Ashes	5.3%	6.5%	0.77	0.98	0.68	0.0%	0.00	1.00	0.00
Calcium carbonate	4.2%	5.0%	0.89	0.99	0.80	4.0%	0.79	0.99	0.80
Cosmetic talc	2.6%	5.6%	0.80	0.96	0.49	0.1%	0.02	1.00	0.04
Abrasive dust	2.1%	1.8%	0.21	0.99	0.21	0.8%	0.10	0.99	0.13
Metallic dust	2.0%	2.4%	0.49	0.99	0.43	1.1%	0.31	1.00	0.40
Cristalline silica	1.9%	1.0%	0.14	0.99	0.17	0.1%	0.03	1.00	0.06
Inorganic pigments	1.8%	1.5%	0.48	0.99	0.52	0.9%	0.34	1.00	0.44

Agents[a]	Expert assessment	CANJEM-25%[c]				CANJEM-50%[c]			
	Percent of jobs exposed	Percent of jobs exposed	Sensitivity[d]	Specificity[d]	Kappa	Percent of jobs exposed	Sensitivity[d]	Specificity[d]	Kappa
Mild steel dust	1.0%	1.2%	0.55	0.99	0.49	0.5%	0.30	1.00	0.40
Organic gases									
Formaldehyde	11.0%	18.4%	0.75	0.89	0.49	10.7%	0.47	0.94	0.41
Propellant gases	1.6%	0.6%	0.31	1.00	0.44	0.6%	0.31	1.00	0.45
Anaesthetic gases	1.0%	2.8%	0.89	0.98	0.47	0.0%	0.00	1.00	0.00
Inorganic gases									
Ozone	6.6%	0.4%	0.01	1.00	0.01	0.0%	0.00	1.00	0.00
Ammonia	5.2%	4.6%	0.48	0.98	0.48	1.6%	0.26	1.00	0.37
Hydrogen chloride	1.1%	1.2%	0.61	0.99	0.57	0.7%	0.34	1.00	0.41
Fumes and smokes									
Cooking fumes	10.8%	12.7%	0.84	0.96	0.74	9.0%	0.71	0.99	0.75
Engine emissions	1.8%	3.2%	0.45	0.98	0.31	1.1%	0.24	0.99	0.29
Natural gas combustion products	1.7%	2.1%	0.26	0.98	0.22	0.0%	0.00	1.00	0.00
Plastics pyrolysis fumes	1.5%	0.3%	0.06	1.00	0.10	0.2%	0.06	1.00	0.10
Soldering fumes	1.1%	1.3%	0.87	1.00	0.79	1.0%	0.76	1.00	0.80
Propane combustion products	1.1%	0.0%	0.00	1.00	0.00	0.0%	0.00	1.00	0.00
Metal oxide fumes	0.9%	1.4%	0.33	0.99	0.26	0.5%	0.07	1.00	0.08
Organic liquids and vapours									
Organic solvents	13.0%	13.8%	0.51	0.92	0.42	5.4%	0.29	0.98	0.35
Isopropanol	6.7%	7.5%	0.52	0.96	0.45	2.2%	0.22	0.99	0.31
Synthetic adhesives	3.7%	3.1%	0.50	0.99	0.53	2.3%	0.46	0.99	0.55
Toluene	2.3%	1.9%	0.46	0.99	0.49	1.4%	0.43	1.00	0.53
Inks	1.7%	1.6%	0.47	0.99	0.47	0.4%	0.17	1.00	0.27
Methanol	1.5%	0.1%	0.00	1.00	0.00	0.1%	0.00	1.00	0.00
Acetone	1.1%	0.9%	0.28	1.00	0.30	0.0%	0.00	1.00	0.00
Benzene	1.1%	2.1%	0.61	0.99	0.41	0.6%	0.34	1.00	0.44
Xylene	1.1%	1.5%	0.58	0.99	0.49	0.7%	0.34	1.00	0.41
Acetic acid	1.1%	0.4%	0.18	1.00	0.27	0.2%	0.13	1.00	0.22
Waxes, polishes	1.0%	0.4%	0.09	1.00	0.12	0.1%	0.03	1.00	0.05
Ethanol	0.9%	0.6%	0.47	1.00	0.55	0.6%	0.47	1.00	0.57

Agents[a]	Expert assessment	CANJEM-25%[c]				CANJEM-50%[c]			
	Percent of jobs exposed	Percent of jobs exposed	Sensitivity[d]	Specificity[d]	Kappa	Percent of jobs exposed	Sensitivity[d]	Specificity[d]	Kappa
Inorganic liquids and vapours									
Inorganic acid solutions	1.6%	2.2%	0.69	0.99	0.56	0.5%	0.26	1.00	0.39
Caustic alkali solutions	1.1%	0.5%	0.11	1.00	0.14	0.1%	0.08	1.00	0.14
Chemical families									
Aliphatic aldehydes	11.5%	19.6%	0.76	0.88	0.49	12.0%	0.51	0.93	0.43
Aliphatic alcohols	9.1%	8.3%	0.45	0.95	0.43	3.2%	0.21	0.99	0.28
Hypochlorites	4.1%	1.4%	0.20	0.99	0.28	0.8%	0.14	1.00	0.23
PAHs from any source	3.1%	5.7%	0.44	0.96	0.28	1.6%	0.23	0.99	0.29
Mononuclear aromatic hydrocarbons	2.9%	3.7%	0.56	0.98	0.47	1.9%	0.44	0.99	0.53
Alkanes C5-C17	2.6%	3.4%	0.42	0.98	0.34	1.4%	0.28	0.99	0.36
PAHs from petroleum	2.2%	5.0%	0.55	0.96	0.31	1.6%	0.28	0.99	0.31
Alkanes C18+	2.0%	2.5%	0.50	0.99	0.43	0.8%	0.24	1.00	0.33
Lead compounds	1.4%	3.1%	0.81	0.98	0.49	1.4%	0.64	1.00	0.63
Aromatic amines	1.4%	1.3%	0.46	0.99	0.46	0.6%	0.35	1.00	0.48
Iron compounds	1.3%	1.4%	0.52	0.99	0.49	0.7%	0.39	1.00	0.50
Aliphatic ketones	1.2%	1.1%	0.28	0.99	0.27	0.1%	0.00	1.00	0.00
Tin compounds	1.0%	1.5%	0.89	0.99	0.72	0.8%	0.60	1.00	0.69
General categories									
Cleaning agents	21.5%	23.2%	0.84	0.94	0.75	20.8%	0.77	0.94	0.72
Biocides	13.8%	14.6%	0.74	0.95	0.67	11.3%	0.65	0.97	0.67
Bleaches	1.4%	1.2%	0.43	0.99	0.45	0.7%	0.34	1.00	0.45

a. We selected all agents with at least 30 exposed jobs based on the expert assessment of jobs held by our study population; this led to the selection of 69 agents among all 258 agents available in CANJEM. In this table the agents are ordered by two features. First, they are grouped into chemical-physical-use categories. Within each category they are ordered by percent of exposure in our sample of jobs.

b. The 3,403 jobs were derived from 998 women. The unit of observation is jobs held by women. There were no adjustments for multiple jobs per woman.

c. Two different cutpoints were used to define ever exposure in CANJEM: 1) CANJEM-25%: jobs with probability of exposure of at least 25% were categorized as exposed to a selected agent and jobs with probability of exposure below 25% were categorized as unexposed, and 2) CANJEM-50%: jobs with probability of exposure of at least 50% were categorized as exposed to a selected agent and jobs with probability of exposure below 50% were categorized as unexposed.

d. Sensitivity and specificity were computed considering expert assessment as the gold standard.

Table 2. Joint categorical distribution of sensitivity and specificity among the 69 agents, with each of the two CANJEM assessments (25% or 50%) vs. the expert assessment[a].

Sensitivity	Specificity	CANJEM-25% Number (%) of agents	CANJEM-50% Number (%) of agents
0.00 – 0.29	0.88 – 0.89	0 (0.0%)	0 (0.0 %)
0.00 – 0.29	0.90 – 0.98	1 (1.4 %)	1 (1.4 %)
0.00 – 0.29	0.99 – 1.00	16 (23.2 %)	37 (53.6 %)
0.30 – 0.49	0.88 – 0.89	0 (0.0 %)	0 (0.0 %)
0.30 – 0.49	0.90 – 0.98	5 (7.2 %)	1 (1.4 %)
0.30 – 0.49	0.99 – 1.00	13 (18.8 %)	18 (26.1 %)
0.50 – 0.69	0.88 – 0.89	0 (0.0 %)	0 (0.0 %)
0.50 – 0.69	0.90 – 0.98	5 (7.2 %)	2 (2.9 %)
0.50 – 0.69	0.99 – 1.00	11 (15.9 %)	3 (4.3 %)
0.70 – 0.89	0.88 – 0.89	2 (2.9 %)	0 (0.0 %)
0.70 – 0.89	0.90 – 0.98	12 (17.4 %)	3 (4.3 %)
0.70 – 0.89	0.99 – 1.00	4 (5.8 %)	4 (5.8 %)
Total number of agents (%)		69 (100%)	69 (100%)

a. We have sliced the range of sensitivity estimates into four categories (0.00 – 0.29), (0.30 – 0.49), (0.50 – 0.69), (0.70 – 0.89), and we sliced the range of specificity estimates into three categories (0.88 – 0.89), (0.90 – 0.98), (0.99 – 1.00). This was based on our "eye-ball" perception of the empirical distributions of these parameters in our dataset, rather than on any *a priori* notion of the significance of these cutpoints.

Table 3. Frequency distribution of exposure prevalence, sensitivity, specificity, and kappa across 69 agents reported in Table 1, for CANJEM-25% and CANJEM-50% compared with Expert assessment

Exposure assessment method	Comparison parameters	Min[a]	25%-ile[a]	Median[a]	75%-ile[a]	Max[a]
Expert - reference method						
	% jobs exposed	0.9%	1.1%	1.8%	4.1%	21.5%
CANJEM-25%[b]						
	% jobs exposed	0.0%	1.1%	1.9%	5.6%	23.2%
	Sensitivity[c]	0.00	0.31	0.49	0.74	0.89
	Specificity[c]	0.88	0.98	0.99	0.99	1.00
	Kappa[d]	0.00	0.28	0.46	0.53	0.81
CANJEM-50%[b]						
	% jobs exposed	0.0%	0.2%	0.7%	1.6%	20.8%
	Sensitivity[c]	0.00	0.07	0.26	0.39	0.79
	Specificity[c]	0.93	0.99	1.00	1.00	1.00
	Kappa[d]	0.00	0.10	0.35	0.48	0.80

a. Median, minimum value, 25th percentile, 75th percentile, and maximum value among 69 agents in Table 1.

b. CANJEM probability of exposure transformed to binary exposure status with cutpoint at 25% or 50% probability of exposure.

c. Sensitivity and specificity computed with reference to expert assessment.

d. Kappa between CANJEM assessment and expert assessment.

Table 4. Loss of statistical power to detect an association under various scenarios of exposure misclassification in a hypothetical case-control study.

Assumptions: 1000 cases and 1000 controls; the reference ("gold standard") expert assessment entails an OR of 2.0; prevalence of exposure as indicated in each panel.

Sensitivity[a]	**Specificity**[b]	**OR_{miss}**[c]	**$Prevalence_{miss}$**[d]	**$Power_{miss}$**[e]	**Relative Loss of power**[f]
Panel A assumption Exposure prevalence = 2%					
1.00	1.00	2.0	2%	0.72	n.a.
0.30	0.90	1.1	10%	0.06	91.7%
0.30	0.95	1.1	6%	0.08	88.9%
0.30	0.99	1.4	2%	0.16	77.8%
0.50	0.90	1.1	11%	0.09	87.5%
0.50	0.95	1.2	6%	0.13	81.9%
0.50	0.99	1.5	2%	0.29	59.7%
0.80	0.90	1.1	11%	0.16	77.8%
0.80	0.95	1.2	7%	0.25	65.3%
0.80	0.99	1.6	3%	0.48	33.3%
Panel B assumption Exposure prevalence = 5%					
1.00	1.00	2.0	5%	0.97	n.a.
0.30	0.90	1.1	11%	0.15	84.5%
0.30	0.95	1.2	6%	0.23	76.3%
0.30	0.99	1.6	2%	0.45	53.6%
0.50	0.90	1.2	12%	0.30	69.1%
0.50	0.95	1.3	7%	0.45	53.6%
0.50	0.99	1.7	3%	0.71	26.8%
0.80	0.90	1.3	14%	0.55	43.3%
0.80	0.95	1.5	9%	0.73	24.7%
0.80	0.99	1.8	5%	0.90	7.2%
Panel C assumption Exposure prevalence = 15%					
1.00	1.00	2.0	15%	1.00	n.a.
0.30	0.90	1.3	13%	0.65	35.0%
0.30	0.95	1.5	9%	0.81	19.0%
0.30	0.99	1.8	5%	0.94	6.0%
0.50	0.90	1.5	16%	0.92	8.0%
0.50	0.95	1.6	12%	0.97	3.0%
0.50	0.99	1.9	8%	0.99	1.0%
0.80	0.90	1.6	21%	0.99	1.0%
0.80	0.95	1.7	16%	1.00	0.0%
0.80	0.99	1.9	13%	1.00	0.0%

a. Hypothesized sensitivity vs. expert assessment.

b. Hypothesized specificity vs. expert assessment.

c. Expected OR after misclassification, as per Formula 2 in the text.

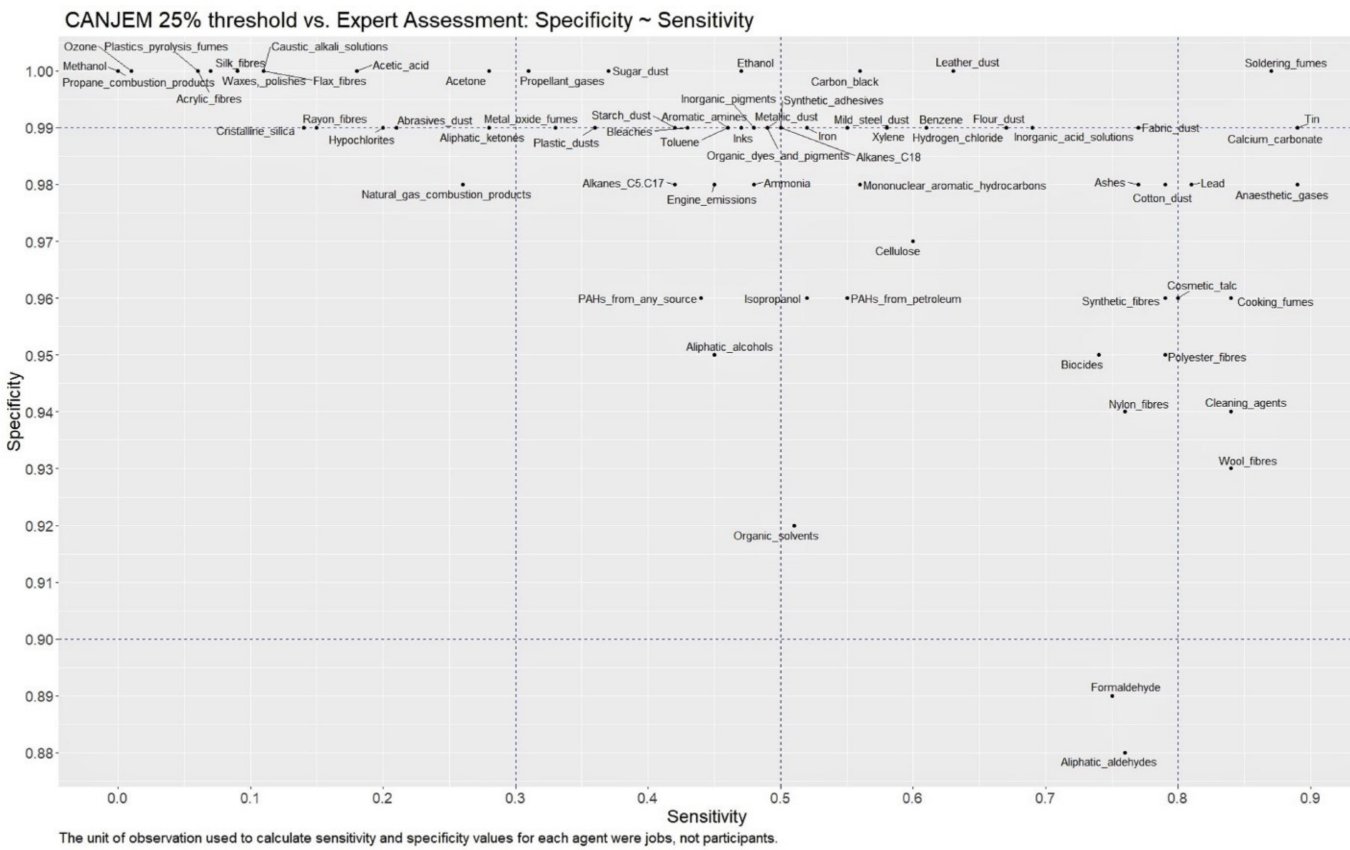
d. Expected prevalence of exposure after misclassification, as per Formula 1 in the text.

e. Expected power under the conditions of an unmatched case-control study with 1000 cases and 1000 controls to detect an association with $OR=OR_{miss}$ and $Prevalence=Prevalence_{miss}$ and sample sizes (2x1000) and alpha of 0.05.

f. Relative loss of power = $(Power_{orig} - Power_{miss}) / Power_{orig}$

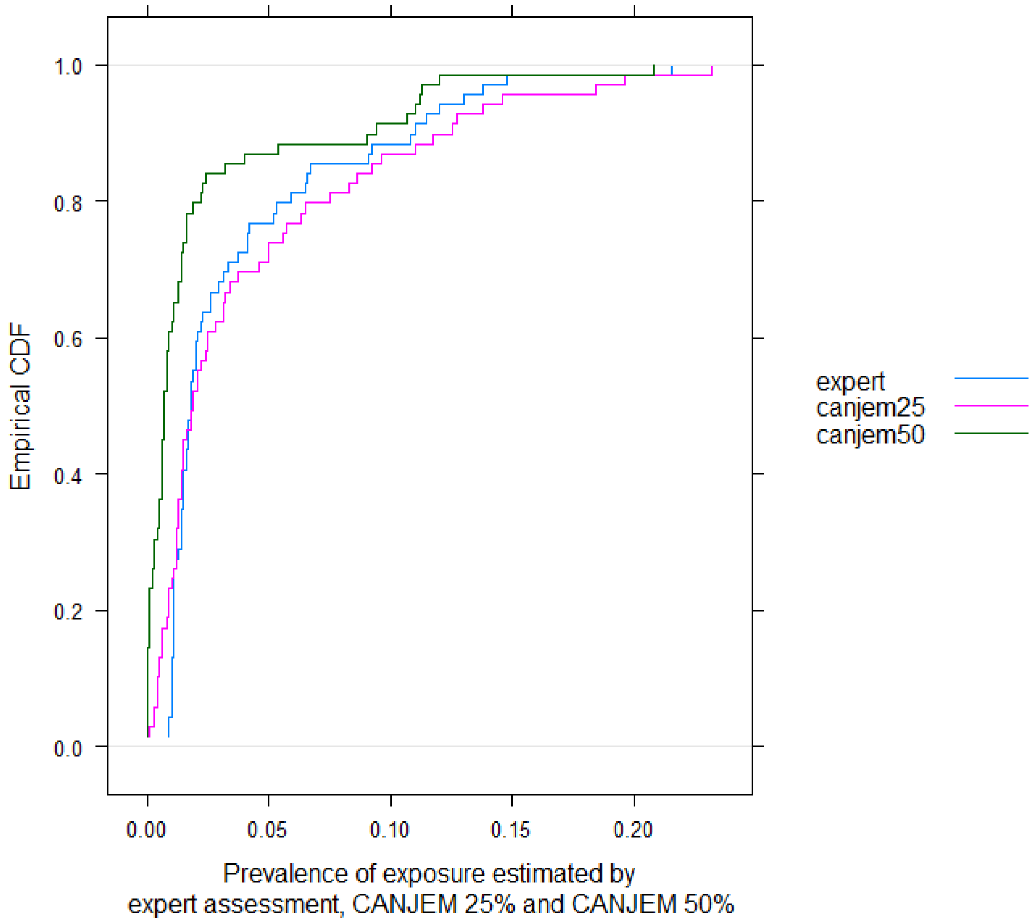
Figure

Figure 1. Sensitivity and specificity values (CANJEM-25% versus expert assessment) for 69 agents with at least 30 exposed jobs in our study population.



Supplementary material

Appendix 1. Empirical cumulative distribution function (ECDF) of estimated prevalence of exposure to our selected 69 agents; using expert assessment, CANJEM-25%, and CANJEM-50%.



Chapter 8: Manuscript 3

Status: Ready for submission

Prevalent occupational exposures and risk of lung cancer among women: Results from the application of the Canadian Job-Exposure Matrix (CANJEM) to a combined set of ten case-control studies

Mengting Xu[1], Vikki Ho[1], Jérôme Lavoué[1], Ann Olsson[2], Joachim Schüz[2], Lesley Richardson[1], Marie-Elise Parent[3], John R. McLaughlin[4], Paul A. Demers[5], Pascal Guénel[6], Loredana Radoi[6], Heinz-Erich Wichmann[7], Wolfgang Ahrens[8], Karl-Heinz Jöckel[9], Dario Consonni[10], Maria Teresa Landi[11], Lorenzo Richiardi[12], Lorenzo Simonato[13], Andrea 't Mannetje[14], Beata Świątkowska[15], John K. Field[16], Neil Pearce[17], Jack Siemiatycki[1]

1. University of Montreal Hospital Research Center (CRCHUM), Montreal, Canada
2. International Agency for Research on Cancer (IARC/WHO), Lyon, France
3. Epidemiology and Biostatistics Unit, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique, Laval, Quebec, Canada
4. Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
5. Occupational Cancer Research Centre, Ontario Health, Toronto, Canada
6. Center for Research in Epidemiology and Population Health (CESP), Team Exposome and Heredity, U1018 Inserm, University Paris-Saclay, Institut Gustave Roussy, Villejuif, France
7. Institute of Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuherberg, Germany
8. Leibniz Institute for Prevention Research and Epidemiology—BIPS, Bremen, Germany
9. Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany
10. Epidemiology Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

11. National Cancer Institute, Bethesda, Maryland, USA

12. Department of Medical Sciences, Cancer Epidemiology Unit, University of Turin, Turin, Italy

13. Department of Cardiovascular Sciences and Public Health, University of Padova, Padova, Italy

14. Research Centre for Hauora and Health, Massey University, Wellington, New Zealand

15. The Nofer Institute of Occupational Medicine, Lodz, Poland

16. Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, The University of Liverpool, Liverpool, UK

17. Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

Contributors:

MX designed the study's analytic strategy, conducted the analysis, and drafted the final manuscript. JS and JL are the PIs of the CANJEM database and provided the CANJEM database for use in this project. AO, JS, MEP, JRM, PAD, PG, LR, HEW, WA, KHJ, DC, MTL, LR, LS, AM, BS, JKF, NP, and JS are the PIs of the participating study centers and provided lung cancer databases for use in this project. All co-authors provided critical review of the manuscript.

Correspondence to:

Dr. Jack Siemiatycki

University of Montreal Hospital Research Center (CRCHUM)

Montreal, Quebec, Canada

E-mail: j.siemiatycki@umontreal.ca

Novelty and Impact

There is little research on occupational risk factors for lung cancer among women. This study examined the associations between 15 prevalent occupational exposures and lung cancer risk among women. This is one of the largest epidemiological studies on occupational causes of lung cancer in women, combining evidence from ten case-control studies in seven countries. None of the agents assessed showed consistent and compelling associations with lung cancer. In some of the main or subgroup analyses, the following agents indicated possible elevated ORs: metallic dust, iron compounds, isopropanol, and organic solvents.

List of abbreviations

IARC, International Agency for Research on Cancer; ISCO-68, International Standard Classification of Occupations, Revised Edition 1968; CANJEM, Canadian job-exposure matrix; JEM, job-exposure matrix; PAHs, polycyclic aromatic hydrocarbons; CE, cumulative exposure; ORs, odds ratios; CIs, confidence intervals; SES, socio-economic status

Abstract

Worldwide, lung cancer is the second leading cause of cancer death in women. The present study explored the associations between occupational exposures that are prevalent among women, and lung cancer. Data from ten case-control studies of lung cancer from Europe, Canada, and New Zealand conducted between 1988 and 2008 were combined. Lifetime occupational history and information on non-occupational factors including smoking were available for 3040 incident lung cancer cases and 4187 controls. We linked each reported job to the Canadian Job-Exposure Matrix (CANJEM), which provided estimates of probability, intensity, and frequency of exposure to each selected agent in each job. For this analysis, we selected 15 agents (cleaning agents, biocides, cotton dust, synthetic fibers, formaldehyde, cooking fumes, organic solvents, cellulose, polycyclic aromatic hydrocarbons from petroleum, ammonia, metallic dust, alkanes C18+, iron compounds, isopropanol, and calcium carbonate) that had lifetime exposure prevalence of at least 5% in the combined study population. For each agent, we estimated lung cancer risk in each study center for ever exposure, by duration of exposure, and by cumulative exposure, using separate logistic regression models adjusted for smoking and other covariates. We then estimated the meta-odds ratios using random-effects meta-analysis. None of the agents assessed showed consistent and compelling associations with lung cancer among women. The following agents showed elevated ORs in some analyses: metallic dust, iron compounds, isopropanol, and organic solvents. Future research into occupational lung cancer risk factors among women should prioritize these agents.

Key words: Lung cancer; Women; Occupational exposures; Job-exposure matrix; Metals

Introduction

Worldwide, lung cancer is the third most diagnosed malignant cancer and the second leading cause of cancer death in women(12). Tobacco smoking is the leading risk factor for lung cancer in women, as well as in men. However, in Western countries, around 20% of women diagnosed with lung cancer had never smoked(21). Numerous occupational exposures have been identified as risk factors for lung cancer(165). Among all cancers attributed to exposure to an occupational agent by the International Agency for Research on Cancer (IARC), lung cancer was the most commonly associated cancer site(166). As early as the 1970s, Doll and Peto estimated that approximately 5% of lung cancer mortality in US women was attributable to occupational factors(40). Similar results were reported for occupationally-attributable lung cancer risk among female workers in Germany in the 1990s(41). A 2017 study concluded that a set of ten recognized occupational carcinogens accounted for 2% of all incident lung cancer cases among French women(43). These estimates likely underestimate the real burden of occupational risk factors for lung cancer in women, since risks were only estimated for a limited number of known carcinogens.

Despite the progress in identifying occupational lung carcinogens over the past decades, epidemiologic evidence of possible carcinogenicity is still sparse or entirely lacking for many occupational exposures. Moreover, much of past occupational cancer research focused on industrial workforces in male-dominated occupations; consequently, there has been little empirical evidence on occupational exposures incurred by women and the associated cancer risks, and published studies concerning women workers tended to be small and rather underpowered(167). In addition, it is potentially misleading to assume that women and men exposed to the same occupational agent would have the same level of risk for cancer, given the biological sex-differences in absorption rate, metabolism, and cellular response(90, 168).

In this study, we aim to explore associations between occupational exposures prevalent in women and lung cancer risk using data of female workers from ten case-control studies of lung cancer.

Methods

Study population

The current analysis includes female participants from ten case-control studies of lung cancer from Europe, Canada, and New Zealand, which collected lifetime working and smoking histories of study participants, including males and females(116, 117, 119-124, 169). Data collection periods for these studies ranged from 1988 to 2008. Seven of the included studies were from Europe (France(120), Germany(117, 119), Italy(124, 169), Poland(121), and the United Kingdom(170)), two were from Canada(116, 123), and one from New Zealand(122). Lifetime occupational and smoking information was mainly collected using face-to-face interviews (approximately 80%), the rest was collected using telephone interviews. Cases in each study were incident lung cancer cases confirmed by histology or cytology, ascertained from local hospitals, clinics, or cancer registries. Controls were frequency-matched (approximately 96%) or individually-matched to cases by age and were recruited from the local general population. Two studies recruited additional hospital controls(121, 123). Participation proportions in the different study centers ranged from 53% to 89% among cases and 41% to 87% among controls. The current analysis included 3040 female lung cancer cases and 4187 female controls. Online Table S1 presents the number of cases and controls in each of the ten study centers and the time period during which the fieldwork was conducted. In aggregate, the 7227 female workers in the combined study population had held 25,679 jobs that lasted at least one year. The principal investigator(s) of each of the ten original studies obtained ethical approval from local institutional ethics review boards, and all participants gave informed consent.

Occupational exposure assessment

Participants' jobs were coded according to the International Standard Classification of Occupations, Revised Edition 1968 (ISCO-68)(133). Occupational exposure to specific agents was assigned by linking participant's job titles to the Canadian job-exposure matrix (CANJEM).

Detailed methodological descriptions of CANJEM (http://canjem.ca/) have been published(157, 158). Briefly, CANJEM is a general population job-exposure matrix (JEM) built from expert assessment of jobs held by participants in the time period 1950-2011 in five Montreal-based case-control studies (multi-site cancers(171), lung cancer(116), breast cancer(104, 105), and brain cancer(172)).

The same expert assessment method and the same team of experts was used in all of the studies on which CANJEM was built. When the team inferred that an agent was present in a worker's workplace, they noted the following dimensions of exposure: *confidence* that the worker really was exposed (possible, probable or definite exposure); *intensity* (on a semi-quantitative scale by agent, where "low" represented a concentration above the background environmental level, and "high" represented the highest levels of concentration to that agent encountered in the Montreal work environment), and *frequency* of exposure (number of hours per week).

The exposure indices provided by CANJEM are formed by three axes: occupational code, time period, and agent. Each cell within CANJEM presents the proportion of all workers with a given occupation code who were considered exposed to a given agent. Further, mirroring the original expert decisions about each agent in each job, each cell describes the frequency distributions of confidence, intensity, and frequency of exposure among the workers who were considered to have been exposed.

CANJEM was built based on jobs held by all participants in the five case-control studies mentioned above, including 65% of male jobs and 35% of female jobs. Some occupations were predominantly held by males and some by females, and some by both males and females. In an analysis of male-female differences in exposure assignment for job titles in which both sexes were present, it was found that for most of those job titles, there was considerable concordance in the exposure profiles between male and female workers(173). For the present analysis, to benefit from the much larger sample, the CANJEM estimates were based on all workers, males and females.

For each of the 25,679 jobs held by women combined from the ten study centers, we linked the ISCO-68 occupation code to CANJEM. We first attempted to link the jobs to the highest resolution (5-digit) of ISCO-68; if unlinkable at the highest resolution, we then linked them at the second highest resolution (3-digit) of ISCO-68. We were able to link CANJEM and provide estimates for 96.5% of all jobs using this strategy (83.6% of jobs were linked at the 5-digit resolution and 12.9% at the 3-digit resolution); the remaining jobs were excluded from the analysis. For each linked job, CANJEM provides the probability of exposure (ranging from 0% to 100%) to each of 258 occupational agents that were part of a checklist evaluated by the expert exposure assessors in the original case-control studies used to build CANJEM. This probability of exposure was calculated as the proportion of jobs with a given occupation code in the CANJEM source database that were considered by the experts to be exposed to the agent. When the probability of exposure to an agent is above 0%, CANJEM also provides estimates of confidence, intensity, and frequency of exposure based on the distributions of these parameters among workers who were considered exposed in the original Montreal studies. CANJEM allows the user to select any level of confidence as a threshold for considering the worker to have been “exposed”, and we chose for the present analysis to include as “exposed” those exposure situations noted by the experts as Probable or Definite.

In order to categorize an agent’s exposure status in a given job as exposed or unexposed, so as to be able to compute odds ratios, it was necessary to select a cutpoint on the probability of exposure scale. We chose a probability cutpoint of 50% (referred to as CANJEM-50%). For each combination of ISCO code and agent, when the probability of exposure was at least 50%, the job was considered as exposed to the agent, when the probability of exposure was less than 10%, the job was considered as unexposed to the agent, and when the probability of exposure was between 10% and 50%, the job was considered as “uncertainly exposed”. We also conducted sensitivity analyses changing the probability of exposure cutpoint to 25% (referred to as CANJEM-25%), where jobs with a probability of exposure of at least 25% were considered exposed, those with less than 10% were considered unexposed, and those between 10% and 25% were considered “uncertainly exposed”.

Selection of agents

It would have been untenable to present results in the present paper for all 258 agents present in CANJEM. It was necessary to significantly reduce the number of agents to be investigated. Three criteria were used: prevalence of the agent in the combined study sample, validity of CANJEM in assigning exposure to the agent, and redundancy among agents. To reduce statistical imprecision, we eliminated all agents that had very few exposed women in the combined dataset for the present analysis; the operational decision was to only include agents with a lifetime ever exposed prevalence of 5% or higher in either cases or controls. When combining the ten case-control studies and applying CANJEM-50%, this led to elimination of 232 agents. The validity of exposure assessment via CANJEM is difficult to ascertain, and it undoubtedly varies by agent. While we do not have data on the validity of the CANJEM-based assessments, we do have some data on the reliability of the assessments. Namely, using one of the Montreal case-control study datasets used to construct CANJEM as a testing ground, we applied CANJEM to the dataset and we compared the resulting exposure estimates with those that had been produced originally in the case-by-case assessment of exposures by a team of experts(174). Based on those results we further eliminated five agents with kappa values less than 0.30. Four agents were not assessed in our previous investigation(174) and were therefore excluded from the current analysis. Finally, we excluded two agents that hierarchically overlapped with other more specific agents (i.e., fabric dusts overlapped with cotton dust and synthetic fibers; aliphatic aldehydes overlapped with formaldehyde). Following these exclusions, we were left with the following 15 agents that form the focus of the present paper (listed in descending prevalence among cases): cleaning agents, biocides, cotton dust, synthetic fibers, formaldehyde, cooking fumes, organic solvents, cellulose, polycyclic aromatic hydrocarbons (PAHs) from petroleum, ammonia, metallic dust, alkanes C18+ (e.g., petroleum jelly), iron compounds, isopropanol, and calcium carbonate. As can be seen, this is an eclectic list that contains specific well-defined chemicals, families of chemicals and general use categories. The agents were not selected based on previous evidence of lung carcinogenicity; thus, we adopted an attitude of pure exploration, allowing the data to drive the results. For the selected agents, Kappa values calculated from our above-mentioned investigation were above 0.70 for four agents, between 0.40 and 0.69 for six agents, and between 0.30 to 0.39 for five agents(174).

Exposure variables

For each of the 15 selected agents we conducted risk analyses in relation to the following metrics of exposure: ever exposure (never, uncertain, ever); duration of exposure (never, 1–10 years, >10 years); and cumulative exposure (CE). A participant would be considered "ever exposed" to an agent if any of her jobs exposed her to that agent. Duration of exposure was calculated as the sum of self-reported duration of each job in which the participant was exposed to an agent. CE was calculated as: $CE = \sum_{i=1}^{d} \frac{I_i}{25} \times \frac{F_i}{40}$, where *i* represents the *i*th year, *d* represents the total number of years exposed, *Ii* represents the intensity of exposure in year *i*, and *Fi* represents the number of hours exposed per week in year *i*. The values of *Ii* were transformed from low, medium, high to ratios of 1, 5, 25 as these were the approximate ratios of intensity that the experts had in mind when coding intensity of exposure for most agents. The formula for cumulative exposure assigns equal weights to the intensity and frequency of exposure through dividing each measure by their highest value. We further categorized CE into three groups (never, ≤ median CE, > median CE) based on agent-specific median values among exposed controls. Participants with uncertain exposure were excluded from the duration or CE analyses.

Statistical analysis

Unconditional logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) of lung cancer associated with each agent's various exposure metrics in each of the ten case-control studies, separately. The reference unexposed category for computation of the ORs for a given agent comprised participants who were never exposed to that agent. Models were adjusted for age (log-transformed), cigarette pack-years (log [pack-years +1]; pack-year was calculated as duration (years) x average cigarette smoking intensity per day/20)(175), years since quitting smoking cigarettes, ever employed in a blue-collar job (defined as jobs with an ISCO-68 first digit of 7, 8, or 9), and socio-economic status (SES). In all study centers except New Zealand, education (no formal education, some primary, primary/some secondary, secondary/some college, and university) was used as the proxy for SES covariate adjustment, and in New Zealand, the Socio-Economic Index (NZSEI) was used as the proxy for SES(126). The main analyses were conducted to assess ORs associated with exposure to each agent and lung cancer risk in all participating women, by smoking stratum, and by lung cancer histological subtypes. Smoking stratified analyses were conducted among never-, light-, and heavy-smokers, separately. Light- and heavy-smokers were categorized based on the median value of pack-years among controls who were ever-smokers. lung subtype analyses were conducted for the three most prevalent lung cancer histological subtypes: adenocarcinoma, squamous cell carcinoma, and small cell lung carcinoma. ORs and 95% CIs for each agent from each separate study center were then agglomerated using random-effects meta-analysis, and heterogeneity among studies was assessed using I^2 statistics(129).

As sensitivity analyses, meta-analyzed lung cancer risks associated with exposure to each selected agent estimated using CANJEM-25%, instead of CANJEM-50%, were estimated for all women, and by smoking stratum. In addition, we also performed pooled analyses on the association between exposure to each agent and lung cancer, including women from all ten participating study centers. Pooled analysis for each agent was adjusted for the same set of covariates included in the main meta-analysis. Because education data were unavailable in the New Zealand study, a category called “unavailable data” was assigned to all observations from this study center for the “education” covariate. Finally, we performed sensitivity analyses adjusting for a reduced set of covariates (only age and smoking) for the meta-analysis of associations between ever exposure to each agent and lung cancer risk among all women.

Note on "statistical significance". The concept of "statistical significance" is frequently misused and misinterpreted(176). In the context of this study, the use of accurate wording that avoids the somewhat clichéd and objectionable "statistical significance" terminology would significantly burden the text. Our use of that terminology here is a convenient and widely understood shorthand for much longer and more accurate phrases to indicate that an observed OR estimate deviates from 1.0 in a way that is highly unlikely to be explained by natural statistical variability. We do not impute a causal interpretation on the use of this terminology.

Analyses were performed with R (V 4.3.0). Meta-analyses were performed with the "meta" package(131).

Results

Selected socio-demographic, smoking, and occupational characteristics of 3040 female lung cancer cases and 4187 female controls in the ten case-control studies are presented in Table 1. Both cases and controls had a median age of 61 years. Socioeconomic status represented by education was available in nine study centers and was lower in cases than in controls. In all study centers, lung cancer cases were more likely to be smokers and to smoke more than controls. The median number of jobs held was three for both cases and controls; however, the proportion of women who had ever held blue-collar jobs was higher in cases.

Selected occupational agents

Online Table S2 shows the definition of each included agent, up to five most prevalent occupations (ISCO-68 job titles) classified as ever exposed to that agent based on CANJEM-50% in our study sample of women, and the prevalence of lifetime exposure to each agent. Comparing the crude prevalence between cases and controls, we note that most of the prevalence estimates were higher among cases, and noticeably so for the following agents: alkanes C18+, iron compounds, metallic dust, organic solvents, cooking fumes, isopropanol, and PAHs from petroleum.

Occupational agents and lung cancer risk among women in ten case-control studies, overall and among never-smokers

Counting the main analyses and all the sensitivity and subgroup analyses, we derived 330 estimates of the OR between each agent and lung cancer. Most of these results are presented in the Online Tables S3 to S8. Table 2 presents six of the results for each agent that we believe are most informative in inferring whether there is evidence of an association. The selected results in Table 2 include results for ever exposure to each agent estimated using CANJEM-50% and CANJEM-25%, and ORs estimated using meta- and pooled-analysis. Meta-ORs are presented separately for all participants and for non-smokers only. Additional meta-ORs for >10 years of exposure and high CE exposure are presented for all participants.

The analysis of each agent was conducted in models with all covariates mentioned above, but without any of the other occupational agents. Thus, mutual confounding among the agents cannot be excluded. Never exposure to an agent under investigation was used as the referent category in all analyses.

In our main meta-analyses where exposure to agents were estimated using CANJEM-50%, there were no clear associations between any of the 15 agents and lung cancer in all women combined. But on the other hand, many of the OR results were compatible with some indication of an increase in risk. To flag those agents that exhibited "suggestive" evidence of a possible association, we implemented the following threshold criteria: the point estimate should be at least 1.10 and the lower 95%CI at least 0.90; and for inverse associations, the point estimate should be at most 0.90 and the upper 95%CI at most 1.10. With these criteria, the following agents exhibited suggestive elevated meta-ORs: isopropanol and organic solvents. There were no clear suggestive inverse associations. Sensitivity analysis re-defining agent exposure using CANJEM-25% instead of CANJEM-50% resulted in overall similar meta-ORs; for some agents, the results were more towards the null. Sensitivity analysis replacing meta-analysis with pooled logistic regression (with study center as a covariate) also yielded overall similar results; with the exception that exposures to calcium carbonate and cellulose became statistically significantly below the null. Sensitivity analyses adjusted only for age and smoking produced overall similar but slightly further from the null results to the meta-analyses that also included some socio-economic covariates (online Table S8).

Among never-smokers, there was an increased risk of lung cancer in women exposed to metallic dust (meta-OR, 95% CI=1.78 (1.12 - 2.81)) vs. those that were never exposed, and a below-the-null OR in women who were exposed to calcium carbonate (meta-OR, 95% CI=0.61 (0.39 – 0.98)). In addition, there was also a suggestive positive OR in never-smokers with exposure to iron compounds. In the sensitivity analysis using CANJEM-25% to categorize exposure status (online Table S7), there were statistically significant increased risks of lung cancer in never-smokers with exposure to metallic dust and iron compounds.

Occupational agents and lung cancer risks by histological subtypes

Table 3 presents the meta-ORs between ever exposure to each agent and each of the following histological subtypes of lung cancer: adenocarcinoma, squamous cell carcinoma, and small cell carcinoma. For alkanes C18+, there were elevated ORs of both squamous cell and small cell carcinomas. For isopropanol, cleaning agents, biocides and cooking fumes, there were elevated ORs of squamous cell carcinoma; and for metallic dust and iron compounds, there were elevated OR of small cell carcinoma. A below-the-null association was observed between formaldehyde and small cell carcinoma. None of the 15 examined agents exhibited suggestively increased risks with lung adenocarcinoma, the most prevalent lung cancer subtype in our study population.

Discussion

We estimated exposure to fifteen relatively prevalent occupational agents, using CANJEM, in an analysis that combined data from ten case-control studies of lung cancer in women. Despite the fact that this was one of the largest datasets ever assembled on a variety of occupational agents and cancer among women, the power to detect risks was modest and many of the OR estimates were quite imprecise.

None of the agents analyzed manifested a pattern of results that persuasively argued for a causal association with lung cancer in our study population. The following agents exhibited some suggestively increased ORs in some of the main or subgroup analyses: metallic dust, iron compounds, isopropanol, and organic solvents. None of the associations showed high heterogeneity in OR estimates among the ten participating centers.

The paucity of previous research on occupational exposures and cancer among women makes it hard to compare our results with prior knowledge; so, for some agents, we will compare our results with prior evidence of carcinogenicity among male or female workers.

Metallic dust and iron compounds: Past occupational studies have shown excess lung cancer risk among workers exposed to compounds of chromium, nickel, beryllium, cadmium, and arsenic(177, 178). However, previous evidence regarding associations between lung cancer and iron, lead, titanium and many other metallic compounds were inconclusive or lacking(177, 178). These studies did not focus on metallic dust specifically but rather on metal compounds in general, and they mostly included male workers. Our research team had previously conducted expert assessment of occupational exposure to a large list of agents including metallic dust in women and men from a Montreal-area population-based case-control study(161). The experts assigned exposure to metallic dust to jobs with exposure to any metal dusts. The specified metallic dust considered include dust from bronze, brass, stainless steel, mild steel, aluminum alloy, chrome, iron, nickel, copper, zinc, cadmium, tin, and lead. In this Montreal-based study, men who were assigned exposure to metallic dust tended to work in heavy industries with large machine tools, whereas commonly exposed jobs among women include punch press operator and sheet metal worker in light industries. The main sources of exposure for women determined by the experts for the Montreal study were dusts from mild steel, brass, and bronze; iron compounds; and occasionally arc or gas welding fumes. In the present study, we observed elevated ORs among women occupationally exposed to metallic dust and iron compounds. The elevated risk seen for metallic dust might be partially attributable to exposure to iron compounds. Occupational exposures during iron and steel founding, and welding fumes, have been classified as causes of lung cancer by IARC(166).

Organic solvents: Occupational exposure to organic solvents was associated with a weakly suggestive elevated risk of lung cancer in our study. An occupational case-control study conducted in France has reported a positive association between lung cancer risk and women ever exposed to perchloroethylene, a common chlorinated solvent(67).

Isopropanol: We observed a suggestive positive risk among women occupationally exposed to isopropanol, but no prior publications were identified for this agent among women.

Calcium carbonate: There were below-the null ORs associated with exposure to calcium carbonate among our study population of women. Teaching is the predominant occupation with this exposure, because of chalk use, and female teachers have been reported to have a lower lung cancer risk when compared to those in other occupations. In the large NOCCA study with 45-year follow-up data on cancer incidence by occupational category for 15 million people, the standardized incidence ratio of lung cancer among female teachers was 0.55 (95%CI, 0.53-0.58)(48). It is thought that non-occupational confounders (namely, smoking) may be responsible for low lung cancer risks among teachers(179), and this may also explain a low risk among women exposed to calcium carbonate.

Cotton dust: It has been hypothesized that the presence of endotoxin in cotton textile manufacturing and agriculture industries could be protective for lung cancer(180). In our analysis, the association between cotton dust and lung cancer was rather null. Most women exposed to cotton dust in our study population were sewers, tailors and dressmakers, and hence had only worked with finished products of chemically treated cotton textiles; whereas endotoxin is mostly found at earlier stages of textile manufacturing where workers are exposed to raw cotton.

In our lung cancer histological subtype analyses, there were statistically significant positive associations between exposure to several agents (isopropanol, alkanes C18+, cleaning agents, and biocides) and risk for squamous cell carcinoma or small cell lung carcinoma, but not for adenocarcinoma. Since adenocarcinoma is less strongly associated with smoking compared to the other two examined subtypes(19), it is possible that the increased risks observed for different agents and squamous cell carcinoma or small cell lung carcinoma could be partially attributed to residual confounding due to smoking or to the particular susceptibility of these cell types of tumours to chemical carcinogenesis.

We chose to use random-effects meta-analysis instead of pooled logistic regression as the main analysis to examine lung cancer association with each agent. The choice between the two modeling approaches represents a trade-off between bias and precision. Compared to pooled analysis, meta-analysis provides a better control for confounding since it allows the effect of confounders to differ by study center, and therefore reduces bias at the cost of increasing variance(181). The meta-analysis approach also allowed the use of all available information for model adjustment, including different SES proxies in participating centers. Given that we have a relatively large sample of women in most of the study centers, we were able to carry out separate logistic regression analysis in each center and derive informative ORs for the meta-analyses. However, depending on the agent, for some subgroup meta-analyses, small numbers led to imprecise OR estimates. We also performed sensitivity analysis examining agent-lung cancer associations using the pooled logistic regression approach, which yielded similar results to those observed in the meta-analyses.

For the present study, separate models were conducted for our meta-analyses of lung cancer risk associated with each agent. Since models were not mutually adjusted for the presence of other occupational agents under investigation, there may be some mutual confounding if there are true risk factors among the selected agents.

We estimated women's occupational exposure to each agent using three exposure metrics: ever exposure, duration of exposure, and cumulative exposure. In the present analysis, we did not examine the effects associated with lagged exposure to each agent, nor did we examine peak exposure, which might also be a factor for lung carcinogenesis for some agents. We reasoned that there were already a huge number of analyses presented in this paper and that the further proliferation of models and results, at a cost of numbing the reader's attention, would do little to clarify possible causal associations.

Using CANJEM, we were able to assess women's lifetime occupational exposures to various agents for a large-scale analysis of ten case-control studies of lung cancer. Such an endeavour of assigning agent-specific exposure in a large study population with lifetime occupational histories would not have been feasible using case-by-case expert assessment due to cost and time constraints. CANJEM, like other JEMs, represents a reproducible and efficient tool which offers a transparent and systematic way to translate job titles into specific exposures, guaranteeing a standardized exposure assessment within and between different studies(182-184). But there are certainly limitations to the validity of deriving exposure data from any JEM such as CANJEM.

The construction of a JEM can be accomplished in many ways with different degrees of expertise and data-based evidence. The validity of JEM entries depends on these factors, and they are difficult to discern, as JEM builders themselves are usually unable to objectively estimate the validity of the data in the JEM given a lack of available true gold standard measurement of past exposures. CANJEM was built from a large database of exposure assessments by "experts" in a series of case-control studies conducted in Montreal and involving job histories spanning about 50 years. The team of experts used a variety of information sources to derive their exposure estimates. Still, like any other JEM, CANJEM is premised on the notion that workers with the same job title, as encapsulated in a given occupational classification system, share similar occupational exposures. This may be true for some occupations and agents, but not all. There is exposure variability among workers who have the same occupation, and this is not normally captured in a JEM. To partially remediate the magnitude of exposure misclassification in our study, we classified exposures with a probability below but relatively close to the chosen cutpoint as "uncertain exposures" and removed them from the reference category in all regression analyses. In addition, we carried out sensitivity analyses replacing CANJEM-50% with a lower probability cutpoint (CANJEM-25%) to define exposure and obtained similar results.

The use of a JEM built in a particular population to estimate exposure in different populations is another potential source of error. CANJEM was built from information about the Montreal working population during the second half of the 20th century. In the present analysis, we applied CANJEM to ten different populations in Europe, Canada, and New Zealand. We chose participating centers in countries that underwent industrialization in similar time periods, in hope that the workplace exposure profiles in given occupations would be roughly similar.

A final source of potential error in the exposure assessment is that since CANJEM was built using source exposure data from both male and female workers, the exposure estimate output of CANJEM would not be able to distinguish any potential exposure differences, if it exists, between a female or a male worker with the same job title.

All of these sources of error would create exposure misclassification, which is expected to be non-differential by disease status since both cases and controls would be assigned the same exposure for a given job title, and therefore it would likely lead to attenuated estimates of ORs.

Conclusion

None of the agents assessed here manifested consistently increased lung cancer risks in women. However, the following agents showed elevated ORs in some of the main or subgroup analyses: metallic dust, iron compounds, organic solvents, and isopropanol. Future research into occupational lung cancer risk factors among women should prioritize these agents.

Author Contributions

MX designed the study's analytic strategy, conducted the analysis, and drafted the final manuscript, under supervision of JS, VH, and LR. JS and JL are the PIs of the CANJEM database and provided the CANJEM database for use in this project. AO, JS, MEP, JRM, PAD, PG, LR, HEW, WA, KHJ, DC, MTL, LR, LS, AM, BS, JKF, NP, and JS are the PIs of the participating study centers and provided lung cancer databases for use in this project. All co-authors provided critical review of the manuscript. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

Conflict of Interest

None declared.

Data Availability Statement

The data that support the findings of our study can be requested from the corresponding author who will in turn make the request to each of the participating center Principal Investigators, who are located in several different countries and operate under their respective national data protection regulations.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer /World Health Organization.

Funding

Mengting Xu was supported by the Fonds de la recherche en santé du Québec and Institut de recherche Robert-Sauvé en santé et en sécurité du travail. The participating case-control studies were funded by: the French National Research Agency (ANR); the French National Institute of Cancer (INCA); the French Agency for Food, Environmental and Occupational Health and Safety (ANSES); the National Agency of Public Health (Santé Publique France); the Foundation for Medical Research (FRM); the Foundation of France; the Association for Research on Cancer (ARC); the French Ministry of Labour (Direction Générale du Travail); the French Ministry of work, solidarity and public function (Direction

Générale de la Santé); the German Federal Ministry of Education, Science, Research, and Technology (grant no. 01 HK 173/0); the Intramural Research Program of NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics and the Center for Cancer Research; the Canadian Cancer Society; the Fonds de recherche en santé du Québec; and the Canadian Institutes for Health Research (MOP-14704).

Ethics Statement

The principal investigator(s) of each of the ten original studies that contributed data to this analysis obtained ethical approval from local institutional ethics review boards, and all participants gave informed consent.

Tables

Table 1. Characteristics of women included in the ten case-control studies of lung cancer; frequency distributions among cases and among controls.

Selected characteristics	**Cases (n=3040)**	**Controls (n=4187)**
Study centers		
Canada-Montreal	14.1%	13.6%
Canada-Toronto	6.4%	11.8%
France (10 departments)	20.0%	18.0%
Germany-Munich plus selected regions	16.8%	12.9%
Germany-Bremen	5.4%	3.9%
Italy-Lombardy	11.8%	10.9%
Italy-Turin and Veneto	4.9%	6.0%
New Zealand	7.5%	8.5%
Poland-Lodz and Warsaw	7.8%	6.2%
United Kingdom-Liverpool	5.2%	8.2%
Age (Median in years [25% –75% percentile])	61 [53 – 68]	61 [53 – 69]
Education[a]		
University	11.6%	18.3%
Secondary / Some college (10-13 yrs)	20.5%	23.6%
Primary/ Some secondary (6-9 yrs)	42.3%	33.2%
Some primary (<6 yrs)	14.6%	13.7%
No formal education	0.8%	0.9%
Not available	10.1%	10.3%
Ever held blue-collar job(s)	48.6%	38.5%
Number of jobs held for at least a year (median)	3	3
Smoking status		

Never smoker	23.8%	58.9%
Former smoker	22.8%	22.0%
Current smoker	52.7%	18.4%
Missing	0.7%	0.7%
Pack-years (median among smokers)	31.5	14.7

a. Information on education was available for all study centers except for New Zealand. For New Zealand, as a proxy socioeconomic status variable, we used a variable derived from the occupational class of the longest held occupation of the participant. Occupational class was determined using a classification of New Zealand occupations based on average levels of income and education in national census data. Values ranged from 10 (lowest class) to 90 (highest class). The median values of this variable among the New Zealand study participants were: 38.3 among cases and 34.0 among controls.

Table 2. Odds ratio between exposure to each of 15 selected agents, estimated using CANJEM-50% and CANJEM-25%, and lung cancer risk among All women and Never-smoker women, combined analysis of ten studies.

Agent	Population	Exposure metric	CANJEM version	Statistical approach [a]	N exposed cases [b]	N exposed controls [b]	N never-exposed cases [b]	N never-exposed controls [b]	OR	95% CI
Metallic dusts	All	Ever exposed	CANJEM-50%	Meta-analysis	214	190	2421	3569	1.08	0.74 - 1.58
	All	Ever exposed	CANJEM-50%	Pooled analysis	214	190	2421	3569	1.13	0.89 - 1.45
	All	Ever exposed	CANJEM-25%	Meta-analysis	310	277	2421	3569	1.09	0.81 - 1.48
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	46	92	608	2161	**1.78**	**1.12 - 2.81**
	All	> 10 years	CANJEM-50%	Meta-analysis	68	60	2421	3569	1.17	0.63 - 2.18
	All	High CE	CANJEM-50%	Meta-analysis	116	96	2421	3569	1.26	0.87 - 1.81
Calcium carbonate	All	Ever exposed	CANJEM-50%	Meta-analysis	109	330	2757	3625	0.77	0.44 - 1.34
	All	Ever exposed	CANJEM-50%	Pooled analysis	109	330	2757	3625	**0.62**	**0.47 - 0.80**
	All	Ever exposed	CANJEM-25%	Meta-analysis	183	460	2757	3625	0.82	0.53 - 1.26
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	35	209	658	2113	**0.61**	**0.39 - 0.98**
	All	> 10 years	CANJEM-50%	Meta-analysis	75	214	2757	3625	0.89	0.48 - 1.65
	All	High CE	CANJEM-50%	Meta-analysis	55	168	2757	3625	0.81	0.42 - 1.56
Cotton dust	All	Ever exposed	CANJEM-50%	Meta-analysis	606	721	2062	3005	0.92	0.73 - 1.17
	All	Ever exposed	CANJEM-50%	Pooled analysis	606	721	2062	3005	0.87	0.73 - 1.03
	All	Ever exposed	CANJEM-25%	Meta-analysis	672	823	2062	3005	0.91	0.72 - 1.14
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	133	448	510	1775	0.87	0.58 - 1.30
	All	> 10 years	CANJEM-50%	Meta-analysis	234	327	2062	3005	0.87	0.68 - 1.12
	All	High CE	CANJEM-50%	Meta-analysis	288	365	2062	3005	0.93	0.71 - 1.22
Synthetic fibers	All	Ever exposed	CANJEM-50%	Meta-analysis	521	655	2149	3137	0.91	0.75 - 1.10
	All	Ever exposed	CANJEM-50%	Pooled analysis	521	655	2149	3137	0.88	0.74 - 1.06
	All	Ever exposed	CANJEM-25%	Meta-analysis	609	734	2149	3137	0.95	0.77 - 1.18
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	123	415	534	1845	0.84	0.58 - 1.23
	All	> 10 years	CANJEM-50%	Meta-analysis	208	305	2149	3137	0.87	0.68 - 1.10
	All	High CE	CANJEM-50%	Meta-analysis	231	328	2149	3137	0.89	0.71 - 1.12
Cellulose	All	Ever exposed	CANJEM-50%	Meta-analysis	296	399	2383	3390	0.82	0.61 - 1.11
	All	Ever exposed	CANJEM-50%	Pooled analysis	296	399	2383	3390	**0.73**	**0.60 - 0.89**
	All	Ever exposed	CANJEM-25%	Meta-analysis	446	584	2383	3390	0.84	0.66 - 1.07

Agent	Population	Exposure metric	CANJEM version	Statistical approach [a]	N exposed cases [b]	N exposed controls [b]	N never-exposed cases [b]	N never-exposed controls [b]	OR	95% CI
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	47	198	612	2063	0.99	0.65 - 1.50
	All	> 10 years	CANJEM-50%	Meta-analysis	90	132	2383	3390	0.93	0.61 - 1.40
	All	High CE	CANJEM-50%	Meta-analysis	153	200	2383	3390	0.90	0.63 - 1.31
Ammonia	All	Ever exposed	CANJEM-50%	Meta-analysis	272	293	1930	2954	1.09	0.88 - 1.37
	All	Ever exposed	CANJEM-50%	Pooled analysis	272	293	1930	2954	1.11	0.90 - 1.37
	All	Ever exposed	CANJEM-25%	Meta-analysis	893	952	1930	2954	1.06	0.90 - 1.25
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	74	188	461	1721	1.09	0.78 - 1.52
	All	> 10 years	CANJEM-50%	Meta-analysis	92	119	1930	2954	0.99	0.71 - 1.39
	All	High CE	CANJEM-50%	Meta-analysis	129	147	1930	2954	1.08	0.80 - 1.45
Formaldehyde	All	Ever exposed	CANJEM-50%	Meta-analysis	514	645	1662	2515	0.92	0.77 - 1.09
	All	Ever exposed	CANJEM-50%	Pooled analysis	514	645	1662	2515	0.88	0.75 - 1.05
	All	Ever exposed	CANJEM-25%	Meta-analysis	915	1011	1662	2515	1.01	0.87 - 1.17
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	107	389	443	1548	0.91	0.68 - 1.21
	All	> 10 years	CANJEM-50%	Meta-analysis	199	280	1662	2515	0.95	0.72 - 1.24
	All	High CE	CANJEM-50%	Meta-analysis	239	328	1662	2515	0.93	0.74 - 1.18
Cooking fumes	All	Ever exposed	CANJEM-50%	Meta-analysis	485	498	2135	3164	1.03	0.86 - 1.24
	All	Ever exposed	CANJEM-50%	Pooled analysis	485	498	2135	3164	1.03	0.88 - 1.21
	All	Ever exposed	CANJEM-25%	Meta-analysis	784	869	2135	3164	1.00	0.85 - 1.16
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	77	270	540	1884	0.95	0.70 - 1.28
	All	> 10 years	CANJEM-50%	Meta-analysis	209	188	2135	3164	1.08	0.75 - 1.56
	All	High CE	CANJEM-50%	Meta-analysis	247	252	2135	3164	1.08	0.85 - 1.37
Isopropanol	All	Ever exposed	CANJEM-50%	Meta-analysis	159	163	2011	2970	1.19	0.90 - 1.57
	All	Ever exposed	CANJEM-50%	Pooled analysis	159	163	2011	2970	1.16	0.89 - 1.51
	All	Ever exposed	CANJEM-25%	Meta-analysis	666	796	2011	2970	1.00	0.87 - 1.15
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	30	79	503	1784	1.46	0.89 - 2.42
	All	> 10 years	CANJEM-50%	Meta-analysis	69	78	2011	2970	1.14	0.67 - 1.95
	All	High CE	CANJEM-50%	Meta-analysis	84	82	2011	2970	1.33	0.81 - 2.18
Organic solvents	All	Ever exposed	CANJEM-50%	Meta-analysis	449	435	1197	1881	1.07	0.88 - 1.31
	All	Ever exposed	CANJEM-50%	Pooled analysis	449	435	1197	1881	1.01	0.84 - 1.22
	All	Ever exposed	CANJEM-25%	Meta-analysis	1240	1334	1197	1881	0.98	0.84 - 1.15

Agent	Population	Exposure metric	CANJEM version	Statistical approach [a]	N exposed cases [b]	N exposed controls [b]	N never-exposed cases [b]	N never-exposed controls [b]	OR	95% CI
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	74	213	335	1130	0.98	0.70 - 1.39
	All	> 10 years	CANJEM-50%	Meta-analysis	157	161	1197	1881	1.18	0.88 - 1.58
	All	High CE	CANJEM-50%	Meta-analysis	222	219	1197	1881	1.16	0.92 - 1.47
Iron compounds	All	Ever exposed	CANJEM-50%	Meta-analysis	160	140	2584	3733	1.10	0.75 - 1.61
	All	Ever exposed	CANJEM-50%	Pooled analysis	160	140	2584	3733	1.09	0.82 - 1.43
	All	Ever exposed	CANJEM-25%	Meta-analysis	239	215	2584	3733	1.09	0.79 - 1.51
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	34	64	638	2240	1.59	0.94 - 2.70
	All	> 10 years	CANJEM-50%	Meta-analysis	50	47	2584	3733	1.15	0.54 - 2.45
	All	High CE	CANJEM-50%	Meta-analysis	80	73	2584	3733	1.13	0.75 - 1.71
Alkanes C18+	All	Ever exposed	CANJEM-50%	Meta-analysis	183	159	2288	3350	1.14	0.86 - 1.51
	All	Ever exposed	CANJEM-50%	Pooled analysis	183	159	2288	3350	1.10	0.84 - 1.43
	All	Ever exposed	CANJEM-25%	Meta-analysis	340	349	2288	3350	0.98	0.80 - 1.21
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	36	82	564	1980	1.30	0.81 - 2.07
	All	> 10 years	CANJEM-50%	Meta-analysis	63	50	2288	3350	1.37	0.85 - 2.19
	All	High CE	CANJEM-50%	Meta-analysis	98	85	2288	3350	1.19	0.82 - 1.72
PAHs from petroleum	All	Ever exposed	CANJEM-50%	Meta-analysis	279	295	1964	2868	0.92	0.72 - 1.17
	All	Ever exposed	CANJEM-50%	Pooled analysis	279	295	1964	2868	0.89	0.72 - 1.11
	All	Ever exposed	CANJEM-25%	Meta-analysis	558	644	1964	2868	0.87	0.73 - 1.02
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	54	156	491	1709	1.05	0.72 - 1.53
	All	> 10 years	CANJEM-50%	Meta-analysis	96	105	1964	2868	1.04	0.68 - 1.59
	All	High CE	CANJEM-50%	Meta-analysis	164	155	1964	2868	1.06	0.78 - 1.45
Cleaning agents	All	Ever exposed	CANJEM-50%	Meta-analysis	1288	1508	1146	1779	0.98	0.85 - 1.12
	All	Ever exposed	CANJEM-50%	Pooled analysis	1288	1508	1146	1779	0.96	0.84 - 1.09
	All	Ever exposed	CANJEM-25%	Meta-analysis	1428	1721	1146	1779	0.98	0.85 - 1.12
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	259	855	305	1056	0.83	0.67 - 1.04
	All	> 10 years	CANJEM-50%	Meta-analysis	747	828	1146	1779	1.06	0.91 - 1.22
	All	High CE	CANJEM-50%	Meta-analysis	755	783	1146	1779	1.09	0.92 - 1.29
Biocides	All	Ever exposed	CANJEM-50%	Meta-analysis	966	1096	1544	2383	1.03	0.89 - 1.18
	All	Ever exposed	CANJEM-50%	Pooled analysis	966	1096	1544	2383	1.01	0.89 - 1.16
	All	Ever exposed	CANJEM-25%	Meta-analysis	1126	1334	1544	2383	1.03	0.90 - 1.17

Agent	Population	Exposure metric	CANJEM version	Statistical approach [a]	N exposed cases [b]	N exposed controls [b]	N never-exposed cases [b]	N never-exposed controls [b]	OR	95% CI
	Never-smokers	Ever exposed	CANJEM-50%	Meta-analysis	206	610	401	1413	0.96	0.77 - 1.20
	All	> 10 years	CANJEM-50%	Meta-analysis	518	589	1544	2383	1.06	0.90 - 1.25
	All	High CE	CANJEM-50%	Meta-analysis	489	550	1544	2383	1.07	0.91 - 1.26

a. The final model for each study center was adjusted for age (log-transformed), cigarette smoking (log [lifetime pack-years +1], and years since quitting), ever employed in a blue-collar job (yes/no), education or NZSEI (in OCANZ study center). The Meta-OR and 95%CI for each agent-lung cancer association was calculated using random-effects meta-analysis. The pooled OR was calculated including study center as an additional covariate. Results are not shown for 1–10 years of exposure, or low CE to each agent (available on request).

b. The number of never exposed and ever exposed women to an agent does not add up to the total number of participants, as there were also women with uncertain exposure, which are excluded here.

Table 3. Odds ratio between ever exposure to each of 15 selected agents, estimated using CANJEM-50%, and lung cancer in women by histological subtypes, meta-analysis of ten studies.

Agent	Exposure Metrics	Lung cancer histological subtypes								
		Adenocarcinoma			Squamous cell carcinoma			Small cell lung carcinoma		
		No. cases	Meta-OR	Meta-OR 95% CI	No. cases	Meta-OR	Meta-OR 95% CI	No. cases	Meta-OR	Meta-OR 95% CI
Metallic dust	Never exposed (Ref) [a]	1076	Ref	-	465	Ref	-	363	Ref	-
	Ever exposed	69	1.08	0.74 - 1.56	55	1.45	0.80 - 2.65	45	1.65	0.88 - 3.13
Calcium carbonate	Never exposed (Ref) [a]	1192	Ref	-	545	Ref	-	442	Ref	-
	Ever exposed	55	0.70	0.42 - 1.17	19	0.82	0.33 - 2.02	15	0.81	0.37 - 1.75
Cotton dust	Never exposed (Ref) [a]	911	Ref	-	388	Ref	-	321	Ref	-
	Ever exposed	248	0.92	0.71 - 1.18	141	0.98	0.63 - 1.51	100	0.74	0.51 - 1.07
Synthetic fibers	Never exposed (Ref) [a]	943	Ref	-	406	Ref	-	336	Ref	-
	Ever exposed	222	0.99	0.78 - 1.26	114	0.93	0.61 - 1.42	82	0.78	0.52 - 1.16
Cellulose	Never exposed (Ref) [a]	1052	Ref	-	467	Ref	-	358	Ref	-
	Ever exposed	120	0.85	0.65 - 1.12	63	0.93	0.62 - 1.38	60	1.06	0.55 - 2.01
Ammonia	Never exposed (Ref) [a]	888	Ref	-	347	Ref	-	295	Ref	-
	Ever exposed	97	0.90	0.68 - 1.20	59	1.42	0.94 - 2.13	62	1.28	0.83 - 1.98
Formaldehyde	Never exposed (Ref) [a]	732	Ref	-	307	Ref	-	251	Ref	-
	Ever exposed	218	0.97	0.77 - 1.21	126	1.21	0.82 - 1.80	78	**0.68**	**0.47 - 0.99**
Cooking fumes	Never exposed (Ref) [a]	948	Ref	-	405	Ref	-	320	Ref	-
	Ever exposed	194	1.02	0.82 - 1.28	106	1.27	0.92 - 1.76	93	1.20	0.84 - 1.73
Isopropanol	Never exposed (Ref) [a]	893	Ref	-	374	Ref	-	314	Ref	-
	Ever exposed	64	1.16	0.82 - 1.65	36	**1.99**	**1.04 - 3.80**	30	1.40	0.80 - 2.45
Organic solvents	Never exposed (Ref) [a]	565	Ref	-	210	Ref	-	174	Ref	-
	Ever exposed	174	0.97	0.75 - 1.24	86	1.06	0.64 - 1.75	80	0.97	0.65 - 1.47
Iron compounds	Never exposed (Ref) [a]	1138	Ref	-	499	Ref	-	393	Ref	-
	Ever exposed	50	1.09	0.73 - 1.63	42	1.33	0.76 - 2.33	38	1.99	0.94 - 4.25
Alkanes C18+	Never exposed (Ref) [a]	1037	Ref	-	424	Ref	-	344	Ref	-
	Ever exposed	53	0.95	0.64 - 1.41	48	1.49	0.93 - 2.40	42	**1.90**	**1.13 - 3.19**
PAHs from petroleum	Never exposed (Ref) [a]	911	Ref	-	368	Ref	-	280	Ref	-
	Ever exposed	92	0.75	0.52 - 1.08	70	1.14	0.77 - 1.69	58	1.28	0.82 - 2.00

Agent	Exposure Metrics	Lung cancer histological subtypes								
		Adenocarcinoma			Squamous cell carcinoma			Small cell lung carcinoma		
		No. cases	**Meta-OR**	**Meta-OR 95% CI**	**No. cases**	**Meta-OR**	**Meta-OR 95% CI**	**No. cases**	**Meta-OR**	**Meta-OR 95% CI**
Cleaning agents	Never exposed (Ref) [a]	541	Ref	-	190	Ref	-	166	Ref	-
	Ever exposed	513	0.85	0.72 - 1.01	283	**1.42**	**1.08 - 1.86**	225	0.88	0.63 - 1.22
Biocides	Never exposed (Ref) [a]	713	Ref	-	271	Ref	-	234	Ref	-
	Ever exposed	389	0.95	0.80 - 1.13	217	**1.38**	**1.01 - 1.88**	170	0.96	0.72 - 1.29

a. The number of never exposed and ever exposed cases to an agent does not add up to the number of cases for each lung cancer subtype, as there were also cases with uncertain exposure, which are excluded here.

Online Supplementary Tables

Table S1. Number of cases and controls in each of ten study centers and fieldwork period.

Select characteristics	Canada-Montreal(116)	Canada-Toronto(123)	France(120)	Germany-Munich(117)	Germany-Bremen(119) [a]	Italy-Lombardy (169)	Italy-Turin and Veneto(124)	New Zealand (122)	Poland(121)	United Kingdom(170)
Fieldwork period	1996–2001	1997–2002	2001–2007	1990–1996	1988–1993	2002–2005	1990–1992	2007–2008	1998–2002	1998–2002
No. Cases	430	194	608	510	165	360	149	227	238	159
Squamous Cell Carcinoma	77	28	95	125	32	40	74	34	50	46
Small Cell Carcinoma	73	13	78	129	45	36	9	34	59	16
Adenocarcinoma	208	74	337	169	60	193	42	99	65	62
Other cancer sub-types	72	79	98	87	28	91	24	60	64	35
No. controls	568	496	752	540	164	457	251	357	259	343

a. The Germany-Bremen sample included a small number of participants from Frankfort.

Table S2. Definition of each of the 15 selected agents and top occupations considered exposed to each agent in 7227 women from ten case-control studies of lung cancer.

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
Inorganic solids						
Metallic dust	Any metal dusts generated, regardless of the specific metals involved or whether they are known or unknown. Most metals will have undergone a certain amount of surface oxidation but exposure to specific metal oxides (e.g., lead oxides; iron oxides) was coded only when the main exposure was to the oxide itself and not to the metal dust.	Machinery Fitter, Machine Assembler and Precision-Instrument Maker; Metal-Press Operator; Machine-Tool Operator; Compositor and Type-Setter; Mechanical Products Inspector and Tester	7.0%	4.5%	Canada-Toronto (1.2%)	United Kingdom (13.1%)
Calcium carbonate	A mineral occurring naturally in a great variety of calcite rocks which are collectively known as limestone. It has been used as a flux in the melting of iron, as a filler in asphalt, putty, crayons, paints, rubber, plastics and linoleum, for writing on blackboards and as a mild abrasive in polishes.	Primary Education Teacher; Secondary Education Teacher	3.6%	7.8%	Germany-Bremen (1.2%)	New Zealand (15.7%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
Organic solids						
Cotton dust	Dust generated during carding, spinning, weaving, cutting, sewing or handling of cotton or cotton-containing textiles. Cotton is a natural fiber obtained from the Gossypium plant; chemically it is about 90% cellulose and 6% moisture, the remainder being impurities. The textile may have been treated with starches, dyes, inks, sizing or other finishing materials, which may have been coded separately.	Sewing-Machine Operator; Hand and Machine Sewer; Tailor and Dressmaker; Chambermaid	19.8%	17.1%	Canada-Toronto (4.6%)	Italy-Turin and Veneto (27.5%)
Synthetic fibers	Dust generated during the manufacturing, spinning, weaving, cutting sewing or handling of artificial or truly synthetic fibers or of textiles containing artificial or synthetic fibers. Artificial fibers are those in which the fiber-forming material is of natural origin (eg., viscose rayon which is	Sewing-Machine Operator; Hand and Machine Sewer; Tailor and Dressmaker; Knitter	17.0%	15.5%	Canada-Toronto (4.8%)	Italy-Lombardy (63.1%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
	regenerated cellulose and celluose acetate fibers) and the true synthetic fibers are those in which the fiber-forming material is derived from petrochemicals or coal chemicals. They are often treated with starches, dyes, inks, sizing or other finishing materials, some of which were coded separately.					
Cellulose	The main constituent of the cell walls of plants. Industrial cellulose is made from wood or cotton pulp. It is used for paper making but also as a starting material for cellulose acetate and cellulose nitrate. Exposure has been mainly coded to workers exposed to paper fibres.	Hand Packer; Librarian; Mail Sorting Clerk; Mailperson; Library Clerk	9.0%	9.5%	Italy-Lombardy (2.0%)	United Kingdom (29.7%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
Inorganic gases						
Ammonia	A by-product of coal distillation and is also produced by passing nitrogen, hydrogen and a catalyst through an electric arc. It is an important source of various nitrogen containing compounds. An enormous quantity of ammonia is used in the production of fertilizers. As a gas it has been used in refrigeration and in nitriding, bright annealing, and for sintering metals. As an aqueous solution (NH4OH), it has been used in the textile and pharmaceutical industries, in medicine, in trade sale paints, in fire extinguishers, and in consumer cleaning products.	Farm Helper; Women's Hairdresser; Chambermaid; Farm Worker	8.9%	6.9%	Italy-Lombardy (1.5%)	Germany-Bremen (17.1%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
Organic gases						
Formaldehyde	A colorless gas obtained by the oxidation of methyl alcohol, it is marketed as a 37% solution by weight under the name of formalin. Formaldehyde has been mainly used for plastics and resin manufacture (see urea-formaldehyde, melamine-formaldehyde, and phenol-formaldehyde), as a disinfectant and fumigant, and as a preservative and hardener of tissues in embalming fluids. Exposure to formaldehyde in the workplace can result from the use of formaldehyde gas or formaldehyde solutions, from outgassing or thermal decomposition of formaldehyde resins or from thermal decomposition of other resins, plastics or organic materials.	Sewing-Machine Operator; Women's Hairdresser; Cook; Tailor and Dressmaker	16.8%	15.3%	Canada-Toronto (8.5%)	Canada-Montreal (22.0%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
Fumes and smokes						
Cooking fumes	A mixture of volatile substances of variable composition resulting from the thermal degradation of fats and other food constituents. Significant quantities of aliphatic aldehydes (formaldehyde and acrolein) have been measured. The temperature and method used for cooking (deep-frying, roasting, charcoal broiling), the type of fat involved, and the number of times it has previously been heated can influence the level of contaminants present in the resulting fumes.	Waitress; Cook; Other Waitress, Bartender and Related Worker; Working Proprietor (Restaurant)	15.9%	11.8%	Italy-Lombardy (5.5%)	New Zealand (21.0%)
Organic liquids and vapours						
Isopropanol	A colorless, flammable, mobile liquid, produced by the hydration of propylene from cracked gases. It has been used mainly in the manufacture of acetone, but is also used in extraction processes, as a solvent (chiefly for oils, perfumes and synthetic resins), in liniments, skin lotions, cosmetics and pharmaceuticals.	Women's Hairdresser; Nurse; Chambermaid; Hairdresser, Barber, Beautician and Related Worker; Offset Pressperson	5.2%	3.9%	Italy-Lombardy (1.3%)	Canada-Montreal (6.0%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
	It has been used in rubbing alcohols and as an antistalling agent in winter grade motor fuels.					
Organic solvents	Organic liquids used as paint thinners, spot removers, dry cleaning agents, diluents, degreasers, chemical reagents, liquid extraction agents, and for many other purposes. Among the first organic liquids used for this purpose were turpentine, benzene, gasoline and naphtha. More recently, non-flameable chlorinated hydrocarbons came into wider use.	Women's Hairdresser; Medical Science Technician; Chambermaid; Solderer; Leather Goods Maker	14.7%	10.3%	Poland (6.9%)	United Kingdom (14.3%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
Chemical families						
Iron compounds	Comprises iron (Fe) dust, iron oxides and iron fumes (all of which were also coded separately), dust from iron-containing alloys (mild and stainless steel were also coded separately), iron-containing ores and all other iron-containing substances. Iron is the most common of the commercial metals and forms a large group of materials known as ferroalloys. Several iron compounds have been used as paint pigments, polishing compounds, and coatings for magnetic tapes while the soluble salts have been used as dyeing mordants, catalysts, fertilizers, in sewage treatments, and in feeds.	Machinery Fitter, Machine Assemblers and Precision-Instrument Maker; Metal-Press Operator; Machine-Tool Operator; Welder and flame-Cutter	5.2%	3.3%	Canada-Toronto (0.8%)	United Kingdom (12.0%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
Alkanes C18+	Includes all saturated hydrocarbons having more than 18 carbon atoms, with the general formula CnH2n+2. They are all solids at standard conditions. One mixture of these long-chained hydrocarbons, known as petroleum jelly, is widely used in lubricating oils and greases and for compounding in rubber and resins. Highly refined, it is used in the pharmaceutical industry. Paraffin waxes, which were also coded separately, also fall into this category.	Spinner and Winder; Metal-Press Operator; Cloth Weaver; Machine-Tool Operator	6.0%	3.8%	Canada-Toronto (0.8%)	United Kingdom (9.6%)
PAHs from petroleum	Polycyclic aromatic hydrocarbons are a group of chemicals made up of three or more benzene rings interlinked in various arrangements. They are naturally present in fossil fuels or can be formed by thermal decomposition of any organic material containing carbon and hydrogen. Crude oil, certain petroleum-derived substances (e.g., heavy fuel oil, asphalt, etc.) and their combustion products contain PAHs, albeit in smaller quantities than similar coal-derived products. Furthermore,	Nursery Worker and Gardener; Salesperson, Shop Assistant and Demonstrator; Spinner and Winder; Metal-Press Operator	9.1%	7.0%	Canada-Montreal (2.8%)	United Kingdom (13.7%)

Agent	Agent definition [a]	Most prevalent occupations (ISCO-68 job titles) ever exposed to selected agent [b]	Lifetime prevalence of exposure among cases [c]	Lifetime prevalence of exposure among controls [c]	Study center with the lowest prevalence (among controls)	Study center with the highest prevalence (among controls)
	concentrations of PAHs may increase in some of these products during use (e.g., used motor oils).					
General categories						
Cleaning agents	Materials which have cleansing action such as soap. Their main function is to aid water in the cleaning process. They may be simple sulphonated fatty acids or complex synthetic materials. Organic solvents were excluded here and have been coded separately.	Charworker; Housemaid; Nurse; Waitress	42.1%	35.7%	Italy-Lombardy (21.2%)	Germany-Bremen (48.2%)
Biocides	Includes all products used to disinfect, deodorize, sterilize and sanitize. This implies the capability of killing micro-organisms (algae, bacteria, viruses, etc.). This group therefore includes bactericides, algicides, fungicides, germicides and preservatives. Agricultural pesticides were coded separately.	Charworker; Housemaid; Nurse; Women's Hairdresser	31.6%	26.0%	Poland (15.8%)	Germany-Bremen (39.6%)

a. Definition for each agent can be found at: http://canjem.ca/.

b. Up to five most prevalent ISCO-68 job titles assigned as exposed to each agent using CANJEM-50%, among our study population of 7227 women.
c. We refer to percent of all women in our study population that were considered exposed to the agent as lifetime prevalence of exposure to the agent.

Table S3. Odds ratio between exposure to each of 15 selected agents, estimated using CANJEM-50%, and lung cancer in women, meta-analysis of ten studies, using three exposure metrics.

Agent	Exposure Metrics	No. of cases	No. of controls	Meta-OR [a]	95% CI [a]	I^2
Metallic dust	Never exposed (Ref) [b]	2421	3569	Ref	-	
	Ever exposed	214	190	1.08	0.74 - 1.58	47.8%
	Duration of exposure: >10 years	68	60	1.17	0.63 - 2.18	30.1%
	Cumulative exposure: high	116	96	1.26	0.87 - 1.81	6.3%
Calcium carbonate	Never exposed (Ref) [b]	2757	3625	Ref	-	
	Ever exposed	109	330	0.77	0.44 - 1.34	64.8%
	Duration of exposure: >10 years	75	214	0.89	0.48 - 1.65	61.5%
	Cumulative exposure: high	55	168	0.81	0.42 - 1.56	60.6%
Cotton dust	Never exposed (Ref) [b]	2062	3005	Ref	-	
	Ever exposed	606	721	0.92	0.73 - 1.17	39.0%
	Duration of exposure: >10 years	234	327	0.87	0.68 - 1.12	12.2%
	Cumulative exposure: high	288	365	0.93	0.71 - 1.22	28.7%
Synthetic fibers	Never exposed (Ref) [b]	2149	3137	Ref	-	
	Ever exposed	521	655	0.91	0.75 - 1.10	0.2%
	Duration of exposure: >10 years	208	305	0.87	0.68 - 1.10	0.0%
	Cumulative exposure: high	231	328	0.89	0.71 - 1.12	0.0%
Cellulose	Never exposed (Ref) [b]	2383	3390	Ref	-	
	Ever exposed	296	399	0.82	0.61 - 1.11	45.1%
	Duration of exposure: >10 years	90	132	0.93	0.61 - 1.40	28.1%
	Cumulative exposure: high	153	200	0.90	0.63 - 1.31	40.0%
Ammonia	Never exposed (Ref) [b]	1930	2954	Ref	-	
	Ever exposed	272	293	1.09	0.88 - 1.37	0.0%
	Duration of exposure: >10 years	92	119	0.99	0.71 - 1.39	0.0%

Agent	Exposure Metrics	No. of cases	No. of controls	Meta-OR [a]	95% CI [a]	I^2
	Cumulative exposure: high	129	147	1.08	0.80 - 1.45	0.0%
Formaldehyde	Never exposed (Ref) [b]	1662	2515	Ref	-	
	Ever exposed	514	645	0.92	0.77 - 1.09	0.0%
	Duration of exposure: >10 years	199	280	0.95	0.72 - 1.24	16.4%
	Cumulative exposure: high	239	328	0.93	0.74 - 1.18	7.3%
Cooking fumes	Never exposed (Ref) [b]	2135	3164	Ref	-	
	Ever exposed	485	498	1.03	0.86 - 1.24	11.3%
	Duration of exposure: >10 years	209	188	1.08	0.75 - 1.56	46.0%
	Cumulative exposure: high	247	252	1.08	0.85 - 1.37	9.1%
Isopropanol	Never exposed (Ref) [b]	2011	2970	Ref	-	
	Ever exposed	159	163	1.19	0.90 - 1.57	0.0%
	Duration of exposure: >10 years	69	78	1.14	0.67 - 1.95	28.5%
	Cumulative exposure: high	84	82	1.33	0.81 - 2.18	28.2%
Organic solvents	Never exposed (Ref) [b]	1197	1881	Ref	-	
	Ever exposed	449	435	1.07	0.88 - 1.31	0.0%
	Duration of exposure: >10 years	157	161	1.18	0.88 - 1.58	10.4%
	Cumulative exposure: high	222	219	1.16	0.92 - 1.47	0.0%
Iron compounds	Never exposed (Ref) [b]	2584	3733	Ref	-	
	Ever exposed	160	140	1.10	0.75 - 1.61	32.0%
	Duration of exposure: >10 years	50	47	1.15	0.54 - 2.45	38.6%
	Cumulative exposure: high	80	73	1.13	0.75 - 1.71	0.0%
Alkanes C18+	Never exposed (Ref) [b]	2288	3350	Ref	-	
	Ever exposed	183	159	1.14	0.86 - 1.51	0.0%
	Duration of exposure: >10 years	63	50	1.37	0.85 - 2.19	0.0%
	Cumulative exposure: high	98	85	1.19	0.82 - 1.72	0.0%
PAHs from petroleum	Never exposed (Ref) [b]	1964	2868	Ref	-	

Agent	Exposure Metrics	No. of cases	No. of controls	Meta-OR [a]	95% CI [a]	I^2
	Ever exposed	279	295	0.92	0.72 - 1.17	15.5%
	Duration of exposure: >10 years	96	105	1.04	0.68 - 1.59	22.0%
	Cumulative exposure: high	164	155	1.06	0.78 - 1.45	14.9%
Cleaning agents	Never exposed (Ref) [b]	1146	1779	Ref	-	
	Ever exposed	1288	1508	0.98	0.85 - 1.12	0.0%
	Duration of exposure: >10 years	747	828	1.06	0.91 - 1.22	0.0%
	Cumulative exposure: high	755	783	1.09	0.92 - 1.29	20.0%
Biocides	Never exposed (Ref) [b]	1544	2383	Ref	-	
	Ever exposed	966	1096	1.03	0.89 - 1.18	0.0%
	Duration of exposure: >10 years	518	589	1.06	0.90 - 1.25	6.1%
	Cumulative exposure: high	489	550	1.07	0.91 - 1.26	0.0%

a. The final model for each study center was adjusted for age (log-transformed), cigarette smoking (log [lifetime pack-years +1], and years since quitting), ever employed in a blue-collar job (yes/no), education or NZSEI (in OCANZ study center). The Meta-OR and 95%CI for each agent-lung cancer association was calculated using random-effects meta-analysis.

b. The number of never exposed and ever exposed women to an agent does not add up to the total number of participants, as there were also women with uncertain exposure, which are excluded here.

Table S4. Odds ratio between ever exposure to each of 15 selected agents, estimated using CANJEM-50%, and lung cancer in women by smoking category, meta-analysis of ten studies.

Agent	**Stratum of Smoking**	**No. of exposed cases**	**No. of exposed controls**	**Meta-OR [a]**	**95% CI [a]**
Metallic dust	Never smokers	46	92	**1.78**	**1.12 - 2.81**
	Light smokers	56	64	1.05	0.65 - 1.69
	Heavy smokers	110	34	0.85	0.45 - 1.59
Calcium carbonate	Never smokers	35	209	**0.61**	**0.39 - 0.98**
	Light smokers	21	97	0.46	0.21 - 1.01
	Heavy smokers	53	24	0.76	0.32 - 1.80
Cotton dust	Never smokers	133	448	0.87	0.58 - 1.30
	Light smokers	103	167	0.94	0.66 - 1.32
	Heavy smokers	321	103	0.99	0.69 - 1.42
Synthetic fibers	Never smokers	123	415	0.84	0.58 - 1.23
	Light smokers	114	146	0.94	0.64 - 1.37
	Heavy smokers	279	90	1.07	0.73 - 1.57
Cellulose	Never smokers	47	198	0.99	0.65 - 1.50
	Light smokers	78	115	0.94	0.63 - 1.41
	Heavy smokers	171	83	0.68	0.33 - 1.37
Ammonia	Never smokers	74	188	1.09	0.78 - 1.52
	Light smokers	74	72	1.19	0.70 - 2.02
	Heavy smokers	124	31	0.94	0.57 - 1.56
Formaldehyde	Never smokers	107	389	0.91	0.68 - 1.21
	Light smokers	119	157	0.93	0.65 - 1.31
	Heavy smokers	286	95	0.85	0.59 - 1.21
Cooking fumes	Never smokers	77	270	0.95	0.70 - 1.28

Agent	Stratum of Smoking	No. of exposed cases	No. of exposed controls	Meta-OR [a]	95% CI [a]
	Light smokers	126	143	1.11	0.75 - 1.64
	Heavy smokers	278	79	1.06	0.77 - 1.48
Isopropanol	Never smokers	30	79	1.46	0.89 - 2.42
	Light smokers	46	60	1.23	0.74 - 2.04
	Heavy smokers	83	22	0.99	0.55 - 1.78
Organic solvents	Never smokers	74	213	0.98	0.70 - 1.39
	Light smokers	130	147	1.30	0.91 - 1.86
	Heavy smokers	244	73	1.06	0.71 - 1.59
Iron compounds	Never smokers	34	64	1.59	0.94 - 2.70
	Light smokers	40	48	0.99	0.57 - 1.73
	Heavy smokers	85	28	0.74	0.40 - 1.37
Alkanes C18+	Never smokers	36	82	1.30	0.81 - 2.07
	Light smokers	47	47	1.12	0.64 - 1.96
	Heavy smokers	98	30	0.93	0.53 - 1.63
PAHs from petroleum	Never smokers	54	156	1.05	0.72 - 1.53
	Light smokers	75	90	0.89	0.58 - 1.39
	Heavy smokers	147	46	0.90	0.57 - 1.42
Cleaning agents	Never smokers	259	855	0.83	0.67 - 1.04
	Light smokers	180	442	1.22	0.94 - 1.57
	Heavy smokers	659	199	0.91	0.60 - 1.38
Biocides	Never smokers	206	610	0.96	0.77 - 1.20
	Light smokers	142	329	1.16	0.88 - 1.54
	Heavy smokers	473	149	1.02	0.68 - 1.52

a. The final model for each study center was adjusted for age (log-transformed), cigarette smoking (log [lifetime pack-years +1], and years since quitting), ever employed in a blue-collar job (yes/no), education or NZSEI (in OCANZ study center). Smoking covariates were not adjusted for in analyses of never-smokers. The Meta-OR and 95%CI for each agent-lung cancer association was calculated using random-effects meta-analysis.

Table S5. Odds ratio between exposure to each of 15 selected agents, estimated using CANJEM-25%, and lung cancer in women, meta-analysis of ten studies, using three exposure metrics.

Agent	Exposure Metrics	No. of cases	Meta-OR [a]	95% CI [a]	I^2 [b]
Metallic dust	Never exposed (Ref) [b]	2421	Ref	-	
	Ever exposed	310	1.09	0.81 - 1.48	44.0%
	Duration of exposure: >10 years	104	0.94	0.55 - 1.60	49.4%
	Cumulative exposure: high	172	1.19	0.89 - 1.60	5.4%
Calcium carbonate	Never exposed (Ref) [b]	2757	Ref	-	
	Ever exposed	183	0.82	0.53 - 1.26	65.1%
	Duration of exposure: >10 years	114	0.84	0.47 - 1.52	71.2%
	Cumulative exposure: high	77	0.73	0.39 - 1.35	64.6%
Cotton dust	Never exposed (Ref) [b]	2062	Ref	-	
	Ever exposed	672	0.91	0.72 - 1.14	35.7%
	Duration of exposure: >10 years	267	0.87	0.68 - 1.13	19.5%
	Cumulative exposure: high	325	0.94	0.72 - 1.22	31.0%
Synthetic fibers	Never exposed (Ref) [b]	2149	Ref	-	
	Ever exposed	609	0.95	0.77 - 1.18	19.8%
	Duration of exposure: >10 years	244	0.91	0.72 - 1.14	0.0%
	Cumulative exposure: high	272	0.97	0.74 - 1.26	25.5%
Cellulose	Never exposed (Ref) [b]	2383	Ref	-	
	Ever exposed	446	0.84	0.66 - 1.07	43.8%
	Duration of exposure: >10 years	151	0.97	0.71 - 1.32	23.8%
	Cumulative exposure: high	225	0.90	0.64 - 1.27	53.4%
Ammonia	Never exposed (Ref) [b]	1930	Ref	-	
	Ever exposed	893	1.06	0.90 - 1.25	16.4%
	Duration of exposure: >10 years	437	1.09	0.90 - 1.33	13.4%

Agent	Exposure Metrics	No. of cases	Meta-OR [a]	95% CI [a]	I^2 [b]
	Cumulative exposure: high	436	1.11	0.93 - 1.33	0.0%
Formaldehyde	Never exposed (Ref) [b]	1662	Ref	-	
	Ever exposed	915	1.01	0.87 - 1.17	0.0%
	Duration of exposure: >10 years	439	1.07	0.89 - 1.28	0.0%
	Cumulative exposure: high	450	1.04	0.87 - 1.25	0.0%
Cooking fumes	Never exposed (Ref) [b]	2135	Ref	-	
	Ever exposed	784	1.00	0.85 - 1.16	14.2%
	Duration of exposure: >10 years	326	1.14	0.93 - 1.40	0.0%
	Cumulative exposure: high	444	1.06	0.86 - 1.30	22.2%
Isopropanol	Never exposed (Ref) [b]	2011	Ref	-	
	Ever exposed	666	1.00	0.87 - 1.15	0.0%
	Duration of exposure: >10 years	314	1.02	0.82 - 1.27	21.3%
	Cumulative exposure: high	308	1.00	0.83 - 1.21	0.0%
Organic solvents	Never exposed (Ref) [b]	1197	Ref	-	
	Ever exposed	1240	0.98	0.84 - 1.15	0.0%
	Duration of exposure: >10 years	641	1.08	0.91 - 1.28	0.0%
	Cumulative exposure: high	659	1.10	0.93 - 1.31	0.0%
Iron compounds	Never exposed (Ref) [b]	2584	Ref	-	
	Ever exposed	239	1.09	0.79 - 1.51	38.1%
	Duration of exposure: >10 years	76	0.98	0.58 - 1.68	30.5%
	Cumulative exposure: high	120	1.05	0.69 - 1.59	25.7%
Alkanes C18+	Never exposed (Ref) [b]	2288	Ref	-	
	Ever exposed	340	0.98	0.80 - 1.21	0.0%
	Duration of exposure: >10 years	105	0.87	0.59 - 1.29	20.7%
	Cumulative exposure: high	189	1.01	0.71 - 1.44	36.7%
PAHs from petroleum	Never exposed (Ref) [b]	1964	Ref	-	

Agent	Exposure Metrics	No. of cases	Meta-OR [a]	95% CI [a]	I^2 [b]
	Ever exposed	558	0.87	0.73 - 1.02	0.0%
	Duration of exposure: >10 years	209	0.82	0.62 - 1.09	25.0%
	Cumulative exposure: high	281	0.86	0.70 - 1.06	0.0%
Cleaning agents	Never exposed (Ref) [b]	1146	Ref	-	
	Ever exposed	1428	0.98	0.85 - 1.12	0.0%
	Duration of exposure: >10 years	869	1.05	0.89 - 1.23	20.6%
	Cumulative exposure: high	807	1.06	0.90 - 1.24	16.0%
Biocides	Never exposed (Ref) [b]	1544	Ref	-	
	Ever exposed	1126	1.03	0.90 - 1.17	0.0%
	Duration of exposure: >10 years	628	1.07	0.92 - 1.25	0.0%
	Cumulative exposure: high	567	1.06	0.91 - 1.24	0.0%

a. The final model for each study center was adjusted for age (log-transformed), cigarette smoking (log [lifetime pack-years +1], and years since quitting), ever employed in a blue-collar job (yes/no), education or NZSEI (in OCANZ study center). The Meta-OR and 95%CI for each agent-lung cancer association was calculated using random-effects meta-analysis.

b. The number of never exposed and ever exposed cases to an agent does not add up to 3040 cases, as there were also cases with uncertain exposure, which are excluded here.

Table S6. Odds ratio between exposure to each of 15 selected agents, estimated using CANJEM-50%, and lung cancer in women, pooled-analysis of ten studies, using three exposure metrics.

Agent	**Exposure Metrics**	**No. of cases**	**Pooled OR**	**Pooled 95% CI [a]**
Metallic dust	Never exposed (Ref) [b]	2421	Ref	-
	Ever exposed	214	1.13	0.89 - 1.45
	Duration of exposure: >10 years	68	1.25	0.83 - 1.89
	Cumulative exposure: high	116	1.27	0.92 - 1.76
Calcium carbonate	Never exposed (Ref) [b]	2757	Ref	-
	Ever exposed	109	**0.62**	**0.47 - 0.80**
	Duration of exposure: >10 years	75	**0.64**	**0.47 - 0.88**
	Cumulative exposure: high	55	**0.60**	**0.42 - 0.85**
Cotton dust	Never exposed (Ref) [b]	2062	Ref	-
	Ever exposed	606	0.87	0.73 - 1.03
	Duration of exposure: >10 years	234	0.88	0.70 - 1.09
	Cumulative exposure: high	288	0.91	0.74 - 1.12
Synthetic fibers	Never exposed (Ref) [b]	2149	Ref	-
	Ever exposed	521	0.88	0.74 - 1.06
	Duration of exposure: >10 years	208	0.87	0.69 - 1.09
	Cumulative exposure: high	231	0.88	0.71 - 1.10
Cellulose	Never exposed (Ref) [b]	2383	Ref	-
	Ever exposed	296	**0.73**	**0.60 - 0.89**
	Duration of exposure: >10 years	90	0.82	0.60 - 1.12
	Cumulative exposure: high	153	0.79	0.61 - 1.02
Ammonia	Never exposed (Ref) [b]	1930	Ref	-
	Ever exposed	272	1.11	0.90 - 1.37

Agent	Exposure Metrics	No. of cases	Pooled OR	Pooled 95% CI [a]
	Duration of exposure: >10 years	92	0.98	0.71 - 1.35
	Cumulative exposure: high	129	1.07	0.81 - 1.41
Formaldehyde	Never exposed (Ref) [b]	1662	Ref	-
	Ever exposed	514	0.88	0.75 - 1.05
	Duration of exposure: >10 years	199	0.93	0.74 - 1.16
	Cumulative exposure: high	239	0.94	0.76 - 1.16
Cooking fumes	Never exposed (Ref) [b]	2135	Ref	-
	Ever exposed	485	1.03	0.88 - 1.21
	Duration of exposure: >10 years	209	1.14	0.90 - 1.44
	Cumulative exposure: high	247	1.09	0.88 - 1.34
Isopropanol	Never exposed (Ref) [b]	2011	Ref	-
	Ever exposed	159	1.16	0.89 - 1.51
	Duration of exposure: >10 years	69	1.12	0.76 - 1.64
	Cumulative exposure: high	84	1.22	0.85 - 1.74
Organic solvents	Never exposed (Ref) [b]	1197	Ref	-
	Ever exposed	449	1.01	0.84 - 1.22
	Duration of exposure: >10 years	157	1.19	0.92 - 1.55
	Cumulative exposure: high	222	1.13	0.90 - 1.42
Iron compounds	Never exposed (Ref) [b]	2584	Ref	-
	Ever exposed	160	1.09	0.82 - 1.43
	Duration of exposure: >10 years	50	1.12	0.70 - 1.79
	Cumulative exposure: high	80	1.05	0.72 - 1.53
Alkanes C18+	Never exposed (Ref) [b]	2288	Ref	-
	Ever exposed	183	1.10	0.84 - 1.43
	Duration of exposure: >10 years	63	1.32	0.86 - 2.04
	Cumulative exposure: high	98	1.23	0.87 - 1.73

Agent	Exposure Metrics	No. of cases	Pooled OR	Pooled 95% CI [a]
PAHs from petroleum	Never exposed (Ref) [b]	1964	Ref	-
	Ever exposed	279	0.89	0.72 - 1.11
	Duration of exposure: >10 years	96	0.98	0.71 - 1.37
	Cumulative exposure: high	164	1.06	0.81 - 1.38
Cleaning agents	Never exposed (Ref) [b]	1146	Ref	-
	Ever exposed	1288	0.96	0.84 - 1.09
	Duration of exposure: >10 years	747	1.05	0.91 - 1.20
	Cumulative exposure: high	755	1.05	0.91 - 1.21
Biocides	Never exposed (Ref) [b]	1544	Ref	-
	Ever exposed	966	1.01	0.89 - 1.16
	Duration of exposure: >10 years	518	1.04	0.89 - 1.21
	Cumulative exposure: high	489	1.05	0.90 - 1.23

a. The final model for each study center was adjusted for age (log-transformed), cigarette smoking (log [lifetime pack-years +1], and years since quitting), ever employed in a blue-collar job (yes/no), education or NZSEI (in OCANZ study center).

b. The number of never exposed and ever exposed cases to an agent does not add up to 3040 cases, as there were also cases with uncertain exposure, which are excluded here.

Table S7. Meta-analysis on the association between ever exposure to each selected agent estimated using CANJEM-25% and lung cancer risk in women by smoking stratum.

Agent (Ever exposure)	Stratum of Smoking	No. exposed cases	Meta-OR [a]	95% CI [a]
Metallic dust	**Never smokers**	60	**1.65**	**1.11 - 2.46**
	Light smokers	83	1.19	0.75 - 1.87
	Heavy smokers	165	0.91	0.52 - 1.60
Calcium carbonate	Never smokers	52	0.71	0.42 - 1.20
	Light smokers	45	0.65	0.38 - 1.10
	Heavy smokers	86	1.04	0.57 - 1.93
Cotton dust	Never smokers	145	0.83	0.57 - 1.21
	Light smokers	170	0.91	0.65 - 1.27
	Heavy smokers	354	1.06	0.74 - 1.52
Synthetic fibers	Never smokers	136	0.83	0.58 - 1.20
	Light smokers	147	1.04	0.72 - 1.49
	Heavy smokers	323	1.17	0.80 - 1.71
Cellulose	Never smokers	72	0.91	0.61 - 1.36
	Light smokers	119	0.96	0.68 - 1.35
	Heavy smokers	252	0.69	0.45 - 1.07
Ammonia	Never smokers	204	0.96	0.72 - 1.28
	Light smokers	242	1.23	0.88 - 1.73
	Heavy smokers	441	1.00	0.73 - 1.35
Formaldehyde	Never smokers	179	0.95	0.74 - 1.21
	Light smokers	229	1.09	0.82 - 1.44
	Heavy smokers	501	0.92	0.68 - 1.24
Cooking fumes	Never smokers	163	0.96	0.75 - 1.23

Agent (Ever exposure)	**Stratum of Smoking**	**No. exposed cases**	**Meta-OR [a]**	**95% CI [a]**
	Light smokers	211	1.08	0.74 - 1.56
	Heavy smokers	406	0.97	0.73 - 1.29
Isopropanol	Never smokers	152	1.08	0.84 - 1.40
	Light smokers	204	1.15	0.84 - 1.57
	Heavy smokers	307	0.81	0.59 - 1.11
Organic solvents	Never smokers	241	0.81	0.63 - 1.06
	Light smokers	354	**1.37**	**1.03 - 1.82**
	Heavy smokers	635	0.88	0.64 - 1.22
Iron compounds	**Never smokers**	50	**1.60**	**1.02 - 2.51**
	Light smokers	60	1.06	0.68 - 1.68
	Heavy smokers	127	0.78	0.48 - 1.25
Alkanes C18+	Never smokers	62	1.01	0.71 - 1.45
	Light smokers	93	1.05	0.69 - 1.58
	Heavy smokers	182	0.91	0.61 - 1.38
PAHs from petroleum	Never smokers	113	0.91	0.70 - 1.20
	Light smokers	156	0.85	0.63 - 1.16
	Heavy smokers	285	0.82	0.59 - 1.16
Cleaning agents	Never smokers	304	0.83	0.67 - 1.03
	Light smokers	407	1.27	0.99 - 1.63
	Heavy smokers	708	0.86	0.65 - 1.14
Biocides	Never smokers	245	0.96	0.78 - 1.19
	Light smokers	324	1.16	0.91 - 1.47
	Heavy smokers	550	0.97	0.73 - 1.28

a. The final model for each study center was adjusted for age (log-transformed), cigarette smoking (log [lifetime pack-years +1], and years since quitting), ever employed in a blue-collar job (yes/no), education or NZSEI (in OCANZ study center). Smoking covariates were not adjusted for in analyses of never-smokers. The Meta-OR and 95%CI for each agent-lung cancer association was calculated using random-effects meta-analysis.

Table S8. Comparison of odds ratios between ever exposure to each of 15 selected agents, estimated using CANJEM-50%, and lung cancer in women, using different covariate adjustments, meta-analysis of ten studies.

Agent (Ever exposure)	Meta-OR and 95%CI (age and smoking adjusted [a])	Meta-OR and 95%CI (fully adjusted [b])
Metallic dust	1.28 (0.89 - 1.83)	1.08 (0.74 - 1.58)
Calcium carbonate	0.54 (0.39 - 0.76)	0.77 (0.44 - 1.34)
Cotton dust	1.14 (0.97 - 1.33)	0.92 (0.73 - 1.17)
Synthetic fibers	1.14 (0.98 - 1.32)	0.91 (0.75 - 1.10)
Cellulose	0.97 (0.73 - 1.28)	0.82 (0.61 - 1.11)
Ammonia	1.28 (1.04 - 1.58)	1.09 (0.88 - 1.37)
Formaldehyde	1.09 (0.94 - 1.28)	0.92 (0.77 - 1.09)
Cooking fumes	1.15 (0.94 - 1.41)	1.03 (0.86 - 1.24)
Isopropanol	1.22 (0.93 - 1.60)	1.19 (0.90 - 1.57)
Organic solvents	1.27 (1.05 - 1.52)	1.07 (0.88 - 1.31)
Iron compounds	1.31 (0.91 - 1.89)	1.10 (0.75 - 1.61)
Alkanes C18+	1.38 (0.99 - 1.92)	1.14 (0.86 - 1.51)
PAHs from petroleum	1.10 (0.85 - 1.44)	0.92 (0.72 - 1.17)
Cleaning agents	1.12 (0.98 - 1.29)	0.98 (0.85 - 1.12)
Biocides	1.16 (1.02 - 1.32)	1.03 (0.89 - 1.18)

a. Models were adjusted for age (log-transformed), and cigarette smoking (log [lifetime pack-years +1], and years since quitting)

b. Additionally adjusted for ever employed in a blue-collar job (yes/no), and socio-economic status (education or NZSEI (in OCANZ study center)).

Chapter 9: General discussions and conclusion

The overarching objective of this thesis is to contribute knowledge on occupational risk factors for lung cancer in women. To achieve this, three separate projects were conducted. We first examined the associations between prevalent occupational exposures as assessed by experts and risk of lung cancer among women in a Montreal lung cancer case-control study. We then carried out a methodological investigation to compare the exposure assignment concordance between CANJEM to that of the "gold-standard" expert assessment for prevalent occupational exposures in women from the same Montreal study. Finally, in the absence of available expert-assessed exposures, we analyzed lung cancer risks among women associated with occupational exposures estimated by CANJEM, using information from a combined international dataset of ten case-control studies of lung cancer.

9.1 Summary of main findings

In our first manuscript (Chapter 6), using the expert assessment method, which is considered as the "gold-standard" approach for retrospective assessment of lifetime occupational exposures to agents in case-control studies, we explored possible associations between lung cancer and 22 agents that were prevalent in jobs held by 882 women using data from a population-based Montreal case-control study of lung cancer. Overall, most agents exhibited no evidence of lung carcinogenicity. Although we observed a few elevated ORs including cooking fumes and mononuclear aromatic hydrocarbons for lung cancer, none of those elevated ORs was statistically significant. In addition, we also examined the associations between occupations held by women and their subsequent lung cancer risk and found that women who had worked over 10 years as waitresses, bartenders and related workers experienced over two-fold risk of lung cancer compared to women working in other occupations. Because of the limited numbers of exposed women in this study and very little prior research on women's occupational risk factors for lung cancer, none of our observed results are conclusive.

Ideally, future studies of occupational risk factors for lung cancer among women should entail much larger sample sizes than what has been done and should be based on more valid retrospective exposure assessment protocols than what has been achieved in the past. However, due to financial and time restraints, it is questionable whether such studies will be conducted. To nevertheless provide some informative data for detecting lung cancer risks among women, we assembled a large combined international dataset (sub-objective 3), with close to 10 times the sample size of the Montreal study. To support that analysis, we examined in manuscript two (Chapter 7) the validity of using an alternative to expert assessment - the Canadian job exposure matrix (CANJEM) - to assess occupational exposures. In this manuscript, we compared concordance of the exposure assessment of a large list of occupational agents generated using CANJEM vs. the expert assessment, for all jobs held by the 882 Montreal women examined in our first manuscript. We found that concordance between CANJEM and the "gold-standard" expert assessment varied greatly by agent and by specific configurations of CANJEM (i.e., the choice of probability of exposure cutpoint for defining ever exposure). In general, either choice of probability cutpoint to define exposure in CANJEM (CANJEM-25% or CANJEM-50%) yielded very high specificity values for all examined agents when compared to the expert assigned exposure status, indicating that both versions of CANJEM can be confidently used to correctly identify a majority of truly unexposed jobs. CANJEM-50% performed even better than CANJEM-25% in terms of specificity, while CANJEM-25% tended to produce higher sensitivity and kappa values compared to CANJEM-50%. In addition, we carried out an exercise to demonstrate the impact of varying degrees of power loss under different hypothetical scenarios of exposure misclassification, exposure prevalence, and relative risk of disease. The findings from this exercise highlighted the key point that using agents with high specificity is important to avoid significant loss of statistical power when detecting an association, especially when the prevalence of exposure is low. This exercise also revealed that as exposure prevalence increases, higher sensitivity increases the power of a study. Based on our findings, we proposed that future users of CANJEM should prioritize agents with high specificity, since prevalence of exposure to many agents tends to be low, accompanied by a low to moderate effect size. This manuscript includes detailed lists of the sensitivity, specificity, and kappa values of all 69 agents to aid potential users of CANJEM in choosing which agents to examine in their own occupational studies.

Lastly, based on the results from manuscript two regarding the agent-specific exposure concordance between CANJEM and expert assessment, in our third manuscript (Chapter 8), we chose to use CANJEM to evaluate 15 occupational agents with minimally adequate exposure concordance to expert assessment and reasonably high prevalence of exposure in a pooled study of lung cancer involving data from ten study centers in Europe, Canada, and New Zealand. Using this combined large dataset containing 7,227 women, which had held a total of 25,679 jobs through their entire work history, we observed that most of the 15 examined agents were not associated with lung cancer. Exposure to metallic dust was significantly associated with an increased risk of lung cancer among female workers who had never smoked. In addition, some other agents, including alkanes C18+, organic solvents, isopropanol, and iron compounds, have also shown elevated point estimates of risk. By contrast, exposure to cotton dust and synthetic fibers manifested below-the-null point estimates of OR. None of the agents assessed manifested consistent and statistically significant association with lung cancer in women.

Overall, lung cancer risk analyses from our first and third manuscripts provided rather consistent findings. Most of our examined occupational agents were not associated with lung cancer risk among women. Both the Montreal study and the international pooled study found some slightly elevated ORs associated with exposure to cooking fumes and organic solvents, but none of the observations were statistically significant. Certain agents revealing elevated ORs in the pooled study, including metallic dust, iron compounds, and alkanes C18+, were unfortunately not examined in the Montreal study due to limited number of exposed women.

9.2 Contribution and originality of the thesis

Women's occupational risk factors for lung cancer have been overlooked in the past. Using detailed occupational exposure assessments, this thesis identified a large list of prevalent occupational exposures in women - many of which have never been assessed in previous studies - and conducted subsequent lung cancer risk analyses. Methodologically, this thesis contributes to exploring the reliability of using CANJEM as an alternative approach to expert assessment for assessing occupational exposures to a large list of agents among women. Concordance of exposure assignment between CANJEM and expert assessment for each agent assessed using data from a Montreal-based case-control study of lung cancer was provided as a reference for future users of CANJEM when choosing suitable agents that best mimic expert-assessed exposures to examine in their own occupational epidemiologic studies. Finally, the results of this thesis' lung cancer risk analyses suggest that occupational exposure to metallic dust, iron compounds, alkanes C18+, isopropanol, and organic solvents may increase lung cancer risk in women. In light of the scarcity of literature on women's occupational lung cancer risk factors, this thesis identified some exposures to prioritize for future research on occupational risk factors in women.

9.3 Limitations of the thesis

9.3.1 Selection bias

Sources of selection bias could arise in case-control studies from biased sampling or participation of cases and controls into the study. That is, cases and controls in the study population are un-representative of the distribution of exposure and outcome of all cases and controls in the source population of interest. This results in a systematic error that threatens the internal validity of a study and can either bias the study results towards or away from the null. This potential bias could be applicable to manuscripts 1 and 3.

In manuscript 1, we analyzed incident lung cancer cases – who were Montreal residents and Canadian citizens – ascertained from all major hospitals in the Montreal area during the study period. To ensure that both the case and control groups come from the same catchment population, controls were also restricted to Montreal residents and Canadian citizens and were randomly selected from the population-based electoral lists, and were frequency matched to cases by age and sex. Therefore, the potential for selection bias due to biased sampling is low. The response proportions in this study were rather high, reaching 82% and 69% for female cases and controls, respectively. Nonetheless, a potential for nonresponse bias, a source of selection bias, cannot be ruled out. A previous occupational case-control study of lung cancer conducted in Italy with available census information on education, marital status, and size of apartment of respondent and non-respondent cases and controls found that less educated cases and more educated controls were more likely to participate in their study (185). We do not have characteristics information on non-respondents in our study. If we assume that less educated cases and more educated controls were also more likely to participate in the Montreal lung cancer study, and that education is also associated with agent exposures, then our estimated ORs could be slightly biased away from the null.

Compared to manuscript 1, the ascertainment and participation of cases and controls in manuscript 3 were more complicated. Since manuscript 3 used data combined from ten separately conducted case-control studies across different countries, during different time periods, and using different sampling strategies, the potential for selection bias due to sampling selection varies by study center. Cases in each study center were female incident lung cancer cases confirmed by histology or cytology, who were ascertained either from local hospitals, clinics, or cancer registries. Female controls were frequency-matched (approximately 96%) or individually-matched to cases by age and were recruited from the local general population. Two study centers recruited additional hospital controls (121, 123). The response proportions in the study centers ranged from 53% to 89% among cases and 41% to 87% among controls. It is not possible to rule out the potential for nonresponse bias in this combined database; however, one cannot deduce the likelihood of selection bias solely based on response proportions (186-188). A combined set of case-control studies has the advantage over an individual study in that it is unlikely for most or all of the ten component studies to be biased in exactly the same manner, so the result of the combined dataset is less likely to be biased than any individual study.

More information contrasting the characteristics of the participating cases and controls to that of the source population is required to fully assess the magnitude of non-response bias. Unfortunately, such information is rarely available in observational case-control studies.

9.3.2 Information bias

In occupational cancer research, information bias in the form of exposure misclassification, has long been one of the greatest concerns (76). This is especially challenging for the retrospective assessment of occupational exposures in case-control studies in the general population. For chronic diseases such as lung cancer with decade-long induction periods, it is necessary that occupational exposures be reconstructed for a participant's entire work history. However, this is a very difficult endeavor as participants may have occupied different occupations across different industries throughout their working years. Except for a few agents (e.g., asbestos, crystalline silica) that have relevant and available historical exposure measurements, most occupational exposures lack quantitative exposure data. Thus, qualitative or semi-quantitative methods of assessment have been widely used in population-based case-control studies of chronic diseases (78). The two exposure assessment methods used in this thesis, the expert assessment and the JEM-based assessment are both semi-quantitative assessment methods that face the potential of misclassification bias. For both methods, there is a risk that participants would not provide an accurate recall of their job history (i.e., self-reported job titles), however, this source of misclassification was likely to be non-differential since it is unlikely that the validity of participants' self-reported job histories would differ based on both their exposures and lung cancer status. In our first manuscript, another source of misclassified exposure could arise due to the use of next of kin as proxy respondents for cases and controls who were too ill or had died at the time of interview. A higher proportion of proxy respondents was used to recall job history and other information for cases than for controls. This could potentially lead to a differential exposure misclassification; however, we initially explored restricting our analysis to self-respondents and found no meaningful differences in agent and cancer associations other than widened confidence intervals caused by a smaller sample size of women included in the analyses. Hence, both self and proxy respondents' data were included in the final results presented in manuscript 1. Exposure misclassification by the experts from reviewing participants' job histories and specific job tasks could also arise during the expert assessment of occupational exposures, but since the experts were blinded to participant's case control status when coding exposures, this source of misclassification was also considered non-differential. A previous publication on the inter-rater reliability study involving the same team of experts implicated in the Montreal lung cancer study revealed high inter-rater agreements (kappa≈0.8) between the experts for exposure assessment (98, 99). In our second manuscript, we focused on assessing the ability for CANJEM to mimic expert assessment of occupational exposures. The calculated sensitivity and specificity values for CANJEM provided in this investigation were compared against the expert assessments, not against the "real exposure" to the participants decades prior, as there is no way of obtaining such data retrospectively. In this manuscript, although the agents we assessed were the most prevalent ones in our sample of women's jobs from the Montreal lung cancer study, a majority of our participants were unexposed to most of these agents. Inclusion of a large number of unexposed jobs leads to an overall high specificity. We found that the level of exposure misclassification due to the use of CANJEM as a replacement for expert assessment differs greatly by agent. CANJEM, like any JEM, cannot assign distinct exposure information based on idiosyncratic features of a particular job or worker that are not captured by distinct occupation codes, therefore, agents with low probability of exposure across many occupations would generally have poorer JEM performance (i.e., higher exposure misclassification) compared to agents with high probability of exposure in a few occupations. The concordance comparison between CANJEM and expert assessment was conducted for ever/never exposure to an agent; it is likely that the concordance between the two methods would be lower if we were to assess more refined exposure parameters (e.g., intensity of exposure) as there would be more opportunity for exposure misclassification to occur. Finally, our comparison of CANJEM-derived estimates with expert assessment method is considered as a "best-case scenario" for CANJEM-derived estimates. This is because CANJEM was built from the database of expert assessments conducted as part of several case-control studies in Montreal, one of which was the Montreal lung cancer study. We do not have data to estimate how much lower the sensitivity and specificity would be if we applied CANJEM to a different study population for which expert assessments of exposure were available.

In our third manuscript, we used CANJEM to assess lifetime occupational exposures to 15 prevent agents exposed by women from ten case-control studies of lung cancer. As mentioned previously, exposure misclassification ensues when using a JEM to assign occupational exposure based on job titles. Still, misclassification resulting from use of a JEM (CANJEM in this case) can be expected to be non-differential by disease status since both cases and controls would be assigned the same exposure for a given job title. In this study, to partially remediate the magnitude of exposure misclassification introduced by CANJEM, we classified exposures with a probability below but relatively close to the chosen cutpoint as "uncertain exposures" and removed them from the reference unexposed category in all lung cancer regression analyses. In addition, we carried out sensitivity analyses replacing CANJEM-50% with a lower probability cutpoint (CANJEM-25%) to define exposure and obtained very similar results. Nonetheless, some exposure misclassification would remain, and this would likely lead to attenuated risk estimates for associations between ever exposure to an agent and lung cancer risk. Another methodological limitation pertains to the use of the current version of CANJEM to estimate female occupational exposures. Since the version used in this thesis was built using source exposure data from both male and female workers, the exposure estimate output of CANJEM would not be able to distinguish any potential exposure differences, if it exists, between a female or a male worker with the same job title. Therefore, exposure misclassification within certain jobs held by women in this pooled study might occur, but still, this misclassification should be non-differential by disease status.

In our lung cancer risk analyses using the Montreal and the international pooled studies (manuscripts 1 and 3), we analyzed women's lung cancer risks associated with ever exposure, duration of exposure, and cumulative exposure to occupational agents. The ever exposure parametrization does not allow the differentiation of lung cancer risks associated with different levels of exposure, and while the cumulative exposure and duration of exposure metrics allowed for a more detailed examination of potential varying cancer risks associated with different levels of exposure to agent, they do not capture all aspects of exposure characteristics. The cumulative exposure (CE) metric, as an example, does not allow the distinction between patterns of exposure that ultimately resulted in the same level of lifetime cumulative exposure to agents, since it includes considerations for intensity, frequency, and duration of exposure. A worker can be classified as having a high CE exposure to an agent either through a long duration of exposure at a low or median intensity, or through a short duration of exposure at a high intensity, as long as their lifetime exposure level is above the median level in the exposed group. Although cumulative exposure has been used as the default exposure metrics for cancer risk assessments (189), it is possible that, given the different modes of action for carcinogenicity, the peak exposure level for some agents is a more significant determinant of future cancer risks (189). An assessment of cancer risk based on CE might be biased in this scenario. One aspect of occupational exposure assessment that has rarely been discussed in studies examining lung cancer risk factors from male-dominated occupations is the discussion on whether to include homemakers in the agent-cancer regression analysis of interest; due to historically gender roles, men have been assumed to be the main breadwinner of a household and thus it is unlikely for them to be homemakers. However, this is a challenge faced by studies of occupational exposures among women. Women who are homemakers — although without an assigned job title according to occupation classification systems — share many similar chemical exposures in the domestic setting to women at work in service occupations, including cooks, waitresses, and cleaning maids, while the measures of exposure would be very particular to any individual homemaker. In our agent-lung cancer analyses (manuscripts 1 and 3), we excluded from the analyses women who were lifelong homemakers because data from the Montreal study have revealed a heightened lung cancer risk among female homemakers when compared to women who had worked outside the home, after adjustment for smoking and socioeconomic status (manuscript in preparation). This indicates that there are likely other lung cancer risk factors that homemakers are more exposed to than working women, however, we had limited information regarding women's domestic exposure. This lack of consideration of domestic exposure would also impact the exposure assessment of working women that we had included in our analyses of the Montreal study and the pooled study. Women were only assigned exposure had it come from a recognized job title based on the International Standard Classification of Occupations (ISCO), any similar exposures occurred from performing homemaking activities, were not considered to be exposed, which, could lead to an underestimation of exposure to certain chemicals.

9.3.3 Confounding

Confounding is another potential source of systematic bias that could have occurred during the analyses of manuscripts 1 and 3 on the associations between occupational exposures among women and lung cancer risks. Since cancer is a multifaceted chronic disease, there exist many other factors that could affect the disease outcome and are also associated with occupational risk factors. Smoking is one of the most important risk factors for lung cancer, in both sexes. In both manuscripts, we were able to adjust for participant's lifetime smoking history in all of our analyses using a comprehensive approach; however, we did not have data regarding involuntary smoking (second-hand or environmental tobacco smoking), which, is carcinogenic to humans (190). In manuscript 1, we observed an increased risk of lung cancer among women working in waitressing jobs in the Montreal case-control studies. It is possible that this increased risk is partially explained by their exposure to passive smoking at work. Before the ban of indoor smoking in public spaces in Montreal in 2016, it was common for servers to be routinely exposed to environmental tobacco smoke in restaurants and bars. There are also possibilities for residual confounding due to smoking or uncontrolled confounding due to not adjusting for other occupational exposures in our agent-cancer analyses. In our manuscript 3 lung cancer histological subtype analyses, there were statistically significant positive associations between exposure to several agents and risk for squamous cell carcinoma (SqCC) or small cell lung carcinoma (SCLC), but not for adenocarcinoma (AdCa). AdCa is less strongly associated with smoking compared to the other two examined lung subtypes (19). It is possible that the increased risks observed for different agents and SqCC or SCLC were partially attributed to residual confounding due to smoking.

There is also a possibility of uncontrolled confounding due to other factors in the analyses of manuscripts 1 and 3. In manuscript 1, to minimize the potential for confounding bias, we initially explored whether adding additional potential confounders including domestic exposures to traditional heating and cooking fuels, as well as more SES covariates including education and income to the statistical models would modify the OR estimates of any of the examined occupational agents with lung cancer. But since none of these additional adjustments changed the OR by more than 10%, they were dropped from the final models. Because we had access to many variables in the Montreal study, and fewer variables in the international dataset of ten studies, we used slightly different sets of covariate adjustments in manuscripts 1 and 3. For both analyses, however, the most important potential confounders (i.e., age and smoking) were available and adjusted for. We lacked information regarding participants' family history of lung cancer in the Montreal study and the international dataset of ten studies. Although family history of lung cancer is identified in the thesis literature review as a risk factor for lung cancer, it is unlikely to be a confounder in our analyses of manuscripts 1 and 3, as this factor is not known to be associated with our exposures of interest (i.e., occupational agents). In manuscript 1, we observed that cases in this study tended to have a shorter lifetime work duration than controls, but we have decided not to adjust for participants' lifetime work duration as a confounder in statistical analyses. Although this factor is associated with occupational exposures, it is not directly associated with lung cancer, other than through its association with age (i.e., longer lifetime work duration can be a proxy for older age), which has already been adjusted for in the models. For both manuscripts 1 and 3, we did not adjust for exposure to occupational lung carcinogens identified in the thesis literature review, as the prevalences of exposure to these agents are very low in both of our study populations of female workers. We also chose not to mutually adjust each of our examined occupational agents for each other in our analyses. Since there is a possible correlation of exposure among different agents of interest, the inclusion of exposure to other agents in the covariate list could lead to over-adjustment in multi-exposure models. However, we are aware that by not mutually adjusting for the agents, there may be confounding if there are true risk factors among the selected agents. For example, the elevated risk seen in women exposed to metallic dust in manuscript 3 might be partially attributable to co-exposure to iron compounds. Given the hypothesis-generating nature of the agent-cancer analyses presented in manuscripts 1 and 3, as empirical data from women is sparse, our priority was to identify any potential associations between prevalent occupational exposure in women and their lung cancer risk; acknowledging that there might be possibility for type 1 error. Further research using other databases are needed to confirm our findings.

9.3.4 Statistical precision

In manuscript 1, we examined the association between prevalent occupational agents and lung cancer in women using data from the Montreal case-control study, which included 422 cases and 577 controls. This study is one of the few to examine the associations between occupational agents and lung cancer among women that went beyond job title analysis. However, the small sample size limited the study's statistical power to detect risks. Due to the moderate to low prevalence of exposure to selected agents and the relatively low number of highly exposed participants, lung cancer risk estimates were imprecise with wide confidence intervals.

9.3.5 Generalizability

Our findings regarding the exposure concordance performance between different exposure assessment methods and the estimated lung cancer risks associated with selected occupational agents only apply to female occupational exposures and female workers. Findings from this thesis should not be generalized to that for male workers without careful validation.

The generalizability of our findings to studies that might use CANJEM to assess these agents in other female populations would depend on how similar occupational exposure profiles are in those populations to that of the Canadian female population, conditional on occupation. Since not all Job Exposure Matrices (JEMs) are built using the same methodology, the concordance performance for each agent reported in this manuscript should not be generalized to determine whether the use of other JEMs can serve as suitable alternatives to expert assessment of these agents. In the third manuscript, CANJEM was used to assess occupational exposures among participants from Canada, Europe, and New Zealand, all of which went through industrialization during similar time periods and thus share comparable occupational exposures within a given job title. Depending on the agent of interest, occupational exposure estimated using CANJEM might not be generalizable to developing countries; in fact, due to a general lack of or less stringent occupational health and safety measures and regulations, the exposure level to many occupational agents is likely much higher in many developing countries than that seen in Canada.

9.4 Strengths

The major strength of manuscript 1 lies in the expert assessment of detailed lifetime occupational exposure of women in this Montreal case-control study of lung cancer. Given the lack of a real gold standard method for retrospective occupational exposure assessment, this expert assessment approach is considered as the best available method for assessing occupational exposures in case-control studies, when compared to other retrospective exposure assessment methods such as job title analysis or subject's self-reported exposure or JEMs, typically derived in different populations (145). The expert assessed exposures allowed us to go beyond conventional job title analysis and provided a rare opportunity to explore the association between a large number of prevalent occupational agents among women and lung cancer risk. An additional strength of the study is the use of a comprehensive smoking index (CSI) for lifetime smoking history adjustment. The CSI is a parsimonious measure of lifetime smoking history, incorporating smoking status, duration, time since cessation and intensity, and this measure has previously been validated for use in this Montreal lung cancer study (106).

The unique strength of manuscript 2 when compared with previous attempts to evaluate the performance of JEMs in published literature, is the focus on common occupational exposures among women, a long-neglected but growing part of the workforce. This manuscript is also an important reference publication because it provides useful insights for future users of CANJEM when choosing occupational exposures in women that are suitable to examine using this exposure assessment tool.

The final manuscript of this thesis examined women's lifetime occupational exposures to various agents using a large-scale combined dataset of ten case-control studies of lung cancer. This is so far one of the largest occupational case-control studies of lung cancer among women. Using CANJEM, we were able to assess lifetime occupational exposures from 25,679 jobs held by participants. Such an endeavour of assigning agent-specific exposure in a large study population with lifetime occupational histories would not have been feasible using case-by-case expert assessment due to cost and time constraints. Moreover, the use of CANJEM, provides a reproducible methodology which offers a transparent and systematic way to translate job titles into specific exposures, guaranteeing a standardized exposure assessment within and between different studies (182). Given that the lung cancer data was combined from different case-control studies, we opted to use random-effects meta-analysis instead of pooled logistic regression as the main analysis to examine lung cancer association with each agent. Compared to pooled analysis, meta-analysis provides a better control for confounding since it allows the effect of confounders to differ by study center (181); in addition, the meta-analysis approach also allowed the use of all available information for model adjustment, including different socio-economic proxy variables collected in participating centers. In sensitivity analysis, we repeated our agent-cancer analyses using pooled logistic regression models, and similar results to the main analyses were observed. The application of both statistical analyses enabled the confirmation of consistent results across statistical approaches and lent greater credibility to the study's findings.

Finally, all three studies contribute to paving the way in identifying potential female occupational risk factors for lung cancer for future research. This thesis provided a portrait of women's prevalent occupational exposures and generated hypotheses on potential lung carcinogens to women at the workplace, which can be tested in future occupational cancer research.

9.5 Future perspectives

This thesis explored the validity of using CANJEM as an alternative to expert assessment of occupational exposures in women, and the subsequent association between exposure to prevalent agents and female workers' risk of lung cancer. Much remains to be understood regarding women's occupational risk factors for lung cancer and this thesis is the first attempt to explore the validity of using CANJEM to assess female occupational exposures.

One important aspect of CANJEM that we could not examine in this thesis is the comparison of exposure concordance between the "gold-standard" expert assessment and a female-specific CANJEM, which would be built on occupational source data exclusively from female workers. At the time of analyses for manuscript two, a parallel project working to incorporate female occupational source data from a Montreal-based breast cancer study (105) into CANJEM was underway and the version of CANJEM with this new data added only became available during the analyses of manuscript three. However, the number of female occupational source data available in CANJEM is still limited compared to that of male occupational source data. With more female data added to CANJEM in the future, it would then be feasible to construct a female-specific CANJEM with enough data points to produce more precise estimates of women's occupational exposures. In addition, it would be interesting to compare whether exposure differs between women and men within the same job title, using the female-specific vs. the male-specific CANJEM.

Given the hypothesis-generating nature of our lung cancer analyses, there is a need for future studies to replicate our findings of the potentially increased risks of lung cancer associated with exposure to metallic dust, iron compounds, isopropanol, and organic solvents, perhaps through conducting occupational cohorts of women working in light manufacturing industries. Furthermore, it is equally vital to produce additional evidence on lung cancer risks associated with agents that we did not find an association with, as well as to expand the research to include agents beyond the list examined in this thesis. It would also be interesting to study co-exposures to multiple correlated agents assessed using CANJEM and their associated lung cancer risk among female workers. In our first manuscript on expert-assessed occupational exposures and lung cancer risk, we conducted additional Principal Component Analysis to explore the pattern of women's occupational co-exposures and their association with lung cancer. However, this additional analysis did not provide a clear insight into women's exposure patterns and were thus not included in the manuscript. Future studies, perhaps with the use of more advanced methods to assess co-exposures/mixtures, could shed more light on women's occupational exposure risk profiles.

9.6 General conclusion

Over the past decades, occupational research had mainly focused on "dirty" blue-collar jobs mostly held by male workers due to gender segregation in the workforce. However, many occupational exposures that could be potentially carcinogenic in prevalent jobs held by women but not men have never been thoroughly examined in occupational epidemiological studies. A lack of publication on women's occupational risk factors to diseases such as lung cancer does not mean that women are not exposed to potentially dangerous chemicals or physical agents at work. Unfortunately, women's occupations are less likely to be unionized, and are often less subjected to occupational hygiene inspections and the same levels of protection and regulation to occupations held by men in the "dirty industries" (191). These problems are even more precarious in developing countries where a sharp increase of women joining the workforce has been observed over the past decades (191). Enormous number of women are now working in insalubrious conditions without proper workplace protective measures and worker's compensation for job-related diseases. The implication of occupational risk factors of lung cancer among women is not limited within the occupational arena, as many of the chemicals to which workers are exposed at work found their way to the general environment, either via industrial emissions or discharges, or the use in consumer products, continuous effects into hazard identification of lung carcinogens in women remain crucial to improving public health.

I want to end this thesis going back to the case of the radium girls. Without their brave fights raising awareness for worker's protection and compensation and demands for justice, more women would have suffered from radium poisoning. Scientists from all fields of medicine and occupational health and safety played a crucial part in aiding these women in winning the case against the radium corporation and banning radium use for painting watch dials. Unfortunately, in spite of the fact that more women are entering the workforce, occupational risk factors for women, especially for chronic illnesses, are still largely overlooked in the scientific community. Since the 1990s, electronic waste (e-waste) recycling has become a fast-growing business and most of the e-wastes were illegally dumped or traded and ended up in informal e-waste recycling communities in developing countries such as China, India, Ghana, and Nigeria, where e-waste workers — who are often women and their children — use primitive recycling methods including sorting (often improperly), heating, and burning using acid baths to extract precious metals such as gold and silver to resell (192-194). Because these are largely informal part-time jobs, the working women and their children are rarely supplied with proper personal protective equipment and are thus exposed to a high level of heavy metals detrimental to health, such as lead, cadmium, nickel, and lithium (193). Our mission as occupational epidemiologists is to identify occupational hazards so that harm can be reduced or prevented and health can be promoted, and to hold industry accountable for its actions. In addition to scientific efforts to identify occupational hazards, a lot more effort in the social and political realm is needed to achieve this goal. As scientists, workers, and conscious citizens, we must continue to fight together on this issue.

References

1. Clark C. Radium girls, women and industrial health reform: 1910-1935: Univ of North Carolina Press; 1997.

2. Moore K. The Radium Girls: The Dark Story of America's Shining Women: Sourcebooks, Inc.; 2017.

3. Zamay TN, Zamay GS, Kolovskaya OS, Zukov RA, Petrova MM, Gargaun A, et al. Current and prospective protein biomarkers of lung cancer. Cancers. 2017;9(11):155.

4. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. Introduction to the 2015 World Health Organization classification of tumors of the lung, pleura, thymus, and heart. Journal of Thoracic Oncology. 2015;10(9):1240-2.

5. Duma N, Santana-Davila R, Molina JR, editors. Non–small cell lung cancer: epidemiology, screening, diagnosis, and treatment. Mayo Clinic Proceedings; 2019: Elsevier.

6. Myers DJ, Wallen JM. Lung adenocarcinoma. StatPearls [Internet]: StatPearls Publishing; 2021.

7. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. Annals of global health. 2019;85(1).

8. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. Journal of thoracic oncology. 2011;6(2):244-85.

9. Lortet-Tieulent J, Soerjomataram I, Ferlay J, Rutherford M, Weiderpass E, Bray F. International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. Lung cancer. 2014;84(1):13-22.

10. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, et al. Projected estimates of cancer in Canada in 2020. Cmaj. 2020;192(9):E199-E205.

11. North CM, Christiani DC, editors. Women and lung cancer: what is new? Seminars in thoracic and cardiovascular surgery; 2013: Elsevier.

12. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

13. Brenner DR, Poirier A, Woods RR, Ellison LF, Billette J-M, Demers AA, et al. Projected estimates of cancer in Canada in 2022. CMAJ. 2022;194(17):E601-E7.

14. Thun MJ, Carter BD, Feskanich D, Freedman ND, Prentice R, Lopez AD, et al. 50-year trends in smoking-related mortality in the United States. N engl J med. 2013;368:351-64.

15. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. Nature Reviews Cancer. 2007;7(10):778-90.

16. Cruz CSD, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clinics in chest medicine. 2011;32(4):605-44.

17. Malhotra J, Malvezzi M, Negri E, La Vecchia C, Boffetta P. Risk factors for lung cancer worldwide. European Respiratory Journal. 2016:ERJ-00359-2016.

18. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. Lung cancer. 2001;31(2):139-48.

19. Pesch B, Kendzia B, Gustavsson P, Jöckel KH, Johnen G, Pohlabeln H, et al. Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case–control studies. International journal of cancer. 2012;131(5):1210-9.

20. Devesa SS, Bray F, Vizcaino AP, Parkin DM. International lung cancer trends by histologic type: male: female differences diminishing and adenocarcinoma rates rising. International journal of cancer. 2005;117(2):294-9.

21. Kligerman S, White C. Epidemiology of lung cancer in women: risk factors, survival, and screening. American journal of roentgenology. 2011;196(2):287-95.

22. Shankar A, Dubey A, Saini D, Singh M, Prasad CP, Roy S, et al. Environmental and occupational determinants of lung cancer. Translational lung cancer research. 2019;8(Suppl 1):S31.

23. Altekruse S, Kosary C, Krapcho M, Neyman N, Aminou R, Waldron W, et al. SEER Cancer Statistics Review, 1975-2007, National Cancer Institute. Bethesda, MD. 2010.

24. Straif K, Baan R, Grosse Y, Secretan B, El Ghissassi F, Cogliano V. Carcinogenicity of household solid fuel combustion and of high-temperature frying. The lancet oncology. 2006;7(12):977-8.

25. Siemiatycki J, Xu M. Occupational causes of cancer. Handbook of Disability, Work and Health. 2019:1-25.

26. Thomas L, Doyle LA, Edelman MJ. Lung cancer in women: emerging differences in epidemiology, biology, and therapy. Chest. 2005;128(1):370-81.

27. Berardi R, Verdecchia L, Paolo MDP, Giampieri R, Scartozzi M, Pierantoni C, et al. Women and lung cancer: clinical and molecular profiling as a determinate for treatment decisions: a literature review. Critical reviews in oncology/hematology. 2009;69(3):223-36.

28. Paulus JK, Zhou W, Kraft P, Johnson BE, Lin X, Christiani DC. Haplotypes of estrogen receptor-beta and risk of non-small cell lung cancer in women. Lung Cancer. 2011;71(3):258-63.

29. Egleston BL, Meireles SI, Flieder DB, Clapper ML, editors. Population-based trends in lung cancer incidence in women. Seminars in oncology; 2009: Elsevier.

30. Heiss G, Wallace R, Anderson GL, Aragaki A, Beresford SA, Brzyski R, et al. Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin. Jama. 2008;299(9):1036-45.

31. Zhang Y, Zhihua Y, Li S, Yan W, Baosen Z. Menstrual factors, reproductive factors and lung cancer risk: a meta-analysis. Chinese Journal of Lung Cancer. 2012;15(12):701.

32. Linseisen J, Rohrmann S, Miller AB, Bueno‐de‐Mesquita HB, Büchner FL, Vineis P, et al. Fruit and vegetable consumption and lung cancer risk: updated information from the European Prospective Investigation into Cancer and Nutrition (EPIC). International journal of cancer. 2007;121(5):1103-14.

33. Tsyganov M, Pevzner A, Ibragimova M, Deryusheva I, Litviakov N. Human papillomavirus and lung cancer: an overview and a meta-analysis. Journal of Cancer Research and Clinical Oncology. 2019;145(8):1919-37.

34. Li Y-J, Tsai Y-C, Chen Y-C, Christiani DC, editors. Human papilloma virus and female lung adenocarcinoma. Seminars in oncology; 2009: Elsevier.

35. Srinivasan M, Taioli E, C. Ragin C. Human papillomavirus type 16 and 18 in primary lung cancers—a meta-analysis. Carcinogenesis. 2009;30(10):1722-8.

36. Siemiatycki J. Historical overview of occupational cancer research. Occupational Cancers: Springer; 2014. p. 1-20.

37. Guha N, Loomis D, Guyton KZ, Grosse Y, El Ghissassi F, Bouvard V, et al. Carcinogenicity of welding, molybdenum trioxide, and indium tin oxide. The Lancet Oncology. 2017;18(5):581-2.

38. Field RW, Withers BL. Occupational and environmental causes of lung cancer. Clinics in chest medicine. 2012;33(4):681-703.

39. Anttila S, Boffetta P, Straif K. Occupational cancers: Springer; 2014.

40. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. JNCI: Journal of the National Cancer Institute. 1981;66(6):1192-308.

41. Jöckel K, Ahrens W, Bolm-Audorff U, Jahn I, Pohlabeln H. Occupationally-induced lung cancer--a quantitative evaluation for the North Germany area. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). 1997;59(4):275-8.

42. Steenland K, Burnett C, Lalich N, Ward E, Hurrell J. Dying for work: the magnitude of US mortality from selected causes of death associated with occupation. American journal of industrial medicine. 2003;43(5):461-82.

43. Marant Micallef C, Charvat H, Houot M-T, Vignat J, Straif K, Paul A, et al. Estimated number of cancers attributable to occupational exposures in France in 2017: an update using a new method for improved estimates. Journal of Exposure Science & Environmental Epidemiology. 2021:1-7.

44. Standard industrial classification manual. Ottawa: Information Canada: Dominion Bureau of Statistics; 1970.

45. National occupational classification, 2011. Ottawa: Minister of Industry. 978-0-662-45368-0. Statistics Canada; 2012.

46. Standard Occupational Classification (SOC) System. U.S. Bureau of Labor Statistics; 2014.

47. International standard classification of occupations. Revised Edition 1968. Geneva: International Labour Office (ILO); 1969.

48. Pukkala E, Martinsen JI, Lynge E, Gunnarsdottir HK, Sparén P, Tryggvadottir L, et al. Occupation and cancer–follow-up of 15 million people in five Nordic countries. Acta oncologica. 2009;48(5):646-790.

49. Guo J, Kauppinen T, Kyyrönen P, Lindbohm ML, Heikkilä P, Pukkala E. Occupational exposure to diesel and gasoline engine exhausts and risk of lung cancer among Finnish workers. American journal of industrial medicine. 2004;45(6):483-90.

50. Amr S, Wolpert B, Loffredo CA, Zheng Y-L, Shields PG, Jones R, Harris CC. Occupation, gender, race, and lung cancer. Journal of occupational and environmental medicine/American College of Occupational and Environmental Medicine. 2008;50(10):1167.

51. Atramont A, Guida F, Mattei F, Matrat M, Cenée S, Sanchez M, et al. Professional Cleaning Activities and Lung Cancer Risk Among Women: Results From the ICARE Study. Journal of occupational and environmental medicine. 2016;58(6):610-6.

52. Jahn I, Ahrens W, Brüske‐Hohlfeld I, Kreuzer M, Möhner M, Pohlabeln H, et al. Occupational risk factors for lung cancer in women: Results of a case‐control study in Germany. American journal of industrial medicine. 1999;36(1):90-100.

53. Robinson CF, Sullivan PA, Li J, Walker JT. Occupational lung cancer in US women, 1984–1998. American journal of industrial medicine. 2011;54(2):102-17.

54. Miettinen OS, WANG J-D. An alternative to the proportionate mortality ratio. American Journal of Epidemiology. 1981;114(1):144-8.

55. Cancer IAfRo. Some flame retardants and textile chemicals, and exposures in the textile manufacturing industry. IARC monographs on the evaluation of carcinogenic risks to humans. 1990;48.

56. Lai PS, Christiani DC. Long-term respiratory health effects in textile workers. Current opinion in pulmonary medicine. 2013;19(2):152-7.

57. Huang X. Cotton dust exposure and risk of lung cancer: A meta-analysis of observational studies. Medicine. 2020;99(14).

58. Checkoway H, Lundin J, Costello S, Ray R, Li W, Eisen E, et al. Possible pro-carcinogenic association of endotoxin on lung cancer among Shanghai women textile workers. British journal of cancer. 2014;111(3):603-7.

59. Astrakianakis G, Seixas NS, Ray R, Camp JE, Gao DL, Feng Z, et al. Lung cancer risk among female textile workers exposed to endotoxin. Journal of the National Cancer Institute. 2007;99(5):357-64.

60. Checkoway H, Ray R, Lundin J, Astrakianakis G, Seixas N, Camp J, et al. Lung cancer and occupational exposures other than cotton dust and endotoxin among women textile workers in Shanghai, China. Occupational and environmental medicine. 2011;68(6):425-9.

61. Elliott L, Loomis D, Dement J, Hein MJ, Richardson D, Stayner L. Lung cancer mortality in North Carolina and South Carolina chrysotile asbestos textile workers. Occupational and environmental medicine. 2012;69(6):385-90.

62. Humans IWGotEoCRt. Diesel and Gasoline Engine Exhausts and Some Nitroarenes. IARC monographs on the evaluation of carcinogenic risks to humans. 2014;105:9.

63. Boffetta P, Dosemeci M, Gridley G, Bath H, Moradi T, Silverman D. Occupational exposure to diesel engine emissions and risk of cancer in Swedish men and women. Cancer Causes & Control. 2001;12(4):365-74.

64. Kwak K, Paek D, Park JT. Occupational exposure to formaldehyde and risk of lung cancer: A systematic review and meta‐analysis. American journal of industrial medicine. 2020;63(4):312-27.

65. Brownson RC, Alavanja MC, Chang JC. Occupational risk factors for lung cancer among nonsmoking women: a case-control study in Missouri (United States). Cancer Causes & Control. 1993;4(5):449-54.

66. Mahboubi A, Koushik A, Siemiatycki J, Lavoué J, Rousseau M-C. Assessment of the effect of occupational exposure to formaldehyde on the risk of lung cancer in two Canadian population-based case—control studies. Scandinavian journal of work, environment & health. 2013:401-10.

67. Mattei F, Guida F, Matrat M, Cenée S, Cyr D, Sanchez M, et al. Exposure to chlorinated solvents and lung cancer: results of the ICARE study. Occupational and environmental medicine. 2014;71(10):681-9.

68. Cassidy A, Mannetje At, Van Tongeren M, Field JK, Zaridze D, Szeszenia-Dabrowska N, et al. Occupational exposure to crystalline silica and risk of lung cancer: a multicenter case–control study in Europe. Epidemiology. 2007;18(1):36-43.

69. Kreuzer M, Heinrich J, Kreienbrock L, Schaffrath Rosario A, Gerken M, Wichmann HE. Risk factors for lung cancer among nonsmoking women. International journal of cancer. 2002;100(6):706-13.

70. Nurminen MM, Jaakkola MS. Mortality from occupational exposure to environmental tobacco smoke in Finland. Journal of occupational and environmental medicine. 2001;43(8):687-93.

71. Stayner L, Bena J, Sasco AJ, Smith R, Steenland K, Kreuzer M, Straif K. Lung cancer risk and workplace exposure to environmental tobacco smoke. American Journal of Public Health. 2007;97(3):545-51.

72. Sun X-w, Dai X-d, Lin C-Y, Shi Y-b, Ma Y-Y, Li W. Environmental tobacco smoke (ETS) and lung cancer among nonsmoking women in Harbin, China. Lung Cancer. 1996(14):S237.

73. Veglia F, Vineis P, Overvad K, Boeing H, Bergmann MM, Trichopoulou A, et al. Occupational exposures, environmental tobacco smoke, and lung cancer. Epidemiology. 2007;18(6):769-75.

74. Johnson KC, Hu J, Mao Y, Group CCRER. Lifetime residential and workplace exposure to environmental tobacco smoke and lung cancer in never‐smoking women, Canada 1994–97. International Journal of Cancer. 2001;93(6):902-6.

75. Kauppinen TP. Assessment of exposure in occupational epidemiology. Scandinavian journal of work, environment & health. 1994:19-29.

76. Teschke K, Olshan A, Daniels J, De Roos A, Parks C, Schulz M, Vaughan T. Occupational exposure assessment in case–control studies: opportunities for improvement. Occupational and environmental medicine. 2002;59(9):575-94.

77. Ge CB, Friesen MC, Kromhout H, Peters S, Rothman N, Lan Q, Vermeulen R. 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. 2018;62(9):1047-63.

78. Stewart PA, Lees PS, Francis M. Quantification of historical exposures in occupational cohort studies. Scandinavian journal of work, environment & health. 1996:405-14.

79. Siemiatycki J, Day N, Fabry J, Cooper J. Discovering carcinogens in the occupational environment: a novel epidemiological approach. J Natl Cancer Inst. 1981;66:217-25.

80. Fritschi L, Siemiatycki J, Richardson L. Self-assessed versus expert-assessed occupational exposures. American journal of epidemiology. 1996;144(5):521-7.

81. Gérin M, Siemiatycki J, Kemper H, Bégin D. Obtaining occupational exposure histories in epidemiologic case-control studies. Journal of Occupational Medicine. 1985:420-6.

82. Friesen M, Lavoue J, Teschke K, Van Tongeren M. Occupational exposure assessment in industry- and population-based epidemiological studies. Exposure assessment in environmental epidemiology. 2015;2.

83. Siemiatycki J. Job-exposure matrices. In: Encyclopedia of Epidemiologic Methods, eds. MH Gail and J Benichou. Wiley; 2000.

84. McHugh ML. Interrater reliability: the kappa statistic. Biochemia medica. 2012;22(3):276-82.

85. Betansedi CO, Vaca Vasquez P, Counil E. A comprehensive approach of the gender bias in occupational cancer epidemiology: A systematic review of lung cancer studies (2003‐2014). American journal of industrial medicine. 2018;61(5):372-82.

86. Zahm SH, Blair A. Occupational cancer among women: where have we been and where are we going? American journal of industrial medicine. 2003;44(6):565-75.

87. Hohenadel K, Raj P, Demers PA, Zahm SH, Blair A. The inclusion of women in studies of occupational cancer: a review of the epidemiologic literature from 1991–2009. American journal of industrial medicine. 2015;58(3):276-81.

88. Cook MB, Dawsey SM, Freedman ND, Inskip PD, Wichner SM, Quraishi SM, et al. Sex disparities in cancer incidence by period and age. Cancer Epidemiology and Prevention Biomarkers. 2009;18(4):1174-82.

89. OECD.Stat. Employment: Labour force participation rate, by sex and age group: Organization for economic co-operation and development; 2017 [cited 2018 July 26]. Available from: https://stats.oecd.org/index.aspx?queryid=54741.

90. Gochfeld M. Sex differences in human and animal toxicology: toxicokinetics. Toxicologic pathology. 2017;45(1):172-89.

91. Kim CS, Hu S. Regional deposition of inhaled particles in human lungs: comparison between men and women. Journal of Applied Physiology. 1998;84(6):1834-44.

92. Bennett WD, Zeman KL, Kim C. Variability of fine particle deposition in healthy adults: effect of age and gender. American journal of respiratory and critical care medicine. 1996;153(5):1641-7.

93. Eng A, Mannetje At, McLean D, Ellison-Loschmann L, Cheng S, Pearce N. Gender differences in occupational exposure patterns. Occupational and environmental medicine. 2011;68(12):888-94.

94. Scarselli A, Corfiati M, Di Marzio D, Marinaccio A, Iavicoli S. Gender differences in occupational exposure to carcinogens among Italian workers. BMC Public Health. 2018;18(1):1-10.

95. Pintos J, Parent M-E, Richardson L, Siemiatycki J. Occupational exposure to diesel engine emissions and risk of lung cancer: evidence from two case–control studies in Montreal, Canada. Occup Environ Med. 2012:oemed-2012-100964.

96. Siematycki J, Day N, Fabry J, Cooper J. Discovering carcinogens in the occupational environment: a novel epidemiologic approach National Cancer Institute. 1981;66:217-25.

97. Gérin M, Siemiatycki J, Kemper H, Bégin D. Obtaining occupational exposure histories in epidemiologic case-control studies. Journal of occupational medicine: official publication of the Industrial Medical Association. 1985;27(6):420-6.

98. Siemiatycki J. Risk factors for cancer in the workplace: Boca Raton: CRC Press; 1991.

99. Siemiatycki J, Fritschi L, Nadon L, Gerin M. Reliability of an expert rating procedure for retrospective assessment of occupational exposures in community‐based case‐control studies. American journal of industrial medicine. 1997;31(3):280-6.

100. Fritschi L, Nadon L, Benke G, Lakhani R, Latreille B, Parent ME, Siemiatycki J. Validation of expert assessment of occupational exposures. American journal of industrial medicine. 2003;43(5):519-22.

101. Siemiatycki J, Wacholder S, Richardson L, Dewar R, Gérin M. Discovering carcinogens in the occupational environment: methods of data collection and analysis of a large case-referent monitoring system. Scandinavian Journal of Work, Environment & Health. 1987;13:486-92.

102. Pintos J, Parent ME, Rousseau MC, Case BW, Siemiatycki J. 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. 2008;50(11):1273-81.

103. Lacourt A, Cardis E, Pintos J, Richardson L, Kincl L, Benke G, et al. INTEROCC case-control study: lack of association between glioma tumors and occupational exposure to selected combustion products, dusts and other chemical agents - art. no. 340. BMC Public Health. 2013;13(340):12.

104. Labreche F, Goldberg MS, Valois MF, Nadon L. Postmenopausal breast cancer and occupational exposures. Occup Environ Med. 2010;67(4):263-9.

105. Goldberg MS, Labreche F, Weichenthal S, Lavigne E, Valois M-F, Hatzopoulou M, et al. The association between the incidence of postmenopausal breast cancer and concentrations at street-level of nitrogen dioxide and ultrafine particles. Environmental research. 2017;158:7-15.

106. Leffondré K, Abrahamowicz M, Xiao Y, Siemiatycki J. Modelling smoking history using a comprehensive smoking index: application to lung cancer. Statistics in medicine. 2006;25(24):4132-46.

107. Siemiatycki J, Lavoué J. Availability of a new job-exposure matrix (CANJEM) for epidemiologic and occupational medicine purposes. Journal of occupational and environmental medicine. 2018;60(7):e324-e8.

108. Sauvé J-F, Siemiatycki J, Labrèche F, Richardson L, Pintos J, Sylvestre M-P, et al. Development of and Selected Performance Characteristics of CANJEM, a General Population Job-Exposure Matrix Based on Past Expert Assessments of Exposure. Annals of work exposures and health. 2018.

109. Sauvé J-F, Siemiatycki J, Labrèche F, Richardson L, Pintos J, Sylvestre M-P, et al. Development of and selected performance characteristics of CANJEM, a general population job-exposure matrix based on past expert assessments of exposure. Annals of work exposures and health. 2018;62(7):783-95.

110. Pasquet R. Methodological considerations of the Canadian job-exposure matrix and the evaluation of the risk of brain cancer in relation to occupational exposure to metallic compounds. 2019.

111. Friesen MC, Coble JB, Lu W, Shu X-O, Ji B-T, Xue S, et al. 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. 2012;56(1):80-91.

112. Bassig BA, Friesen MC, Vermeulen R, Shu X-O, Purdue MP, Stewart PA, et al. Occupational exposure to benzene and non-Hodgkin lymphoma in a population-based cohort: the Shanghai Women's Health Study. Environmental health perspectives. 2015;123(10):971-7.

113. Parent M-E, Turner MC, Lavoué J, Richard H, Figuerola J, Kincl L, et al. Lifetime occupational exposure to metals and welding fumes, and risk of glioma: a 7-country population-based case–control study. Environmental Health. 2017;16(1):1-10.

114. Dewar R, Siemiatycki J, Gérin M. Loss of statistical power associated with the use of a job-exposure matrix in occupational case-control studies. Applied Occupational and Environmental Hygiene. 1991;6(6):508-15.

115. Flegal KM, Brownie C, Haas J. The effects of exposure misclassification on estimates of relative risk. American journal of epidemiology. 1986;123(4):736-51.

116. Ramanakumar AV, Parent M-E, Siemiatycki J. Risk of lung cancer from residential heating and cooking fuels in Montreal, Canada. American journal of epidemiology. 2007;165(6):634-42.

117. Brüske-Hohlfeld I, Möhner M, Pohlabeln H, Ahrens W, Bolm-Audorff U, Kreienbrock L, et al. Occupational lung cancer risk for men in Germany: results from a pooled case-control study. American journal of epidemiology. 2000;151(4):384-95.

118. Consonni D, De Matteis S, Lubin JH, Wacholder S, Tucker M, Pesatori AC, et al. Lung cancer and occupation in a population-based case-control study. American journal of epidemiology. 2010;171(3):323-33.

119. Jöckel K-H, Ahrens W, Jahn I, Pohlabeln H, Bolm-AudorfF U. Occupational risk factors for lung cancer: a case-control study in West Germany. International journal of epidemiology. 1998;27(4):549-60.

120. Guida F, Papadopoulos A, Menvielle G, Matrat M, Févotte J, Cénée S, et al. Risk of lung cancer and occupational history: results of a French population-based case-control study, the ICARE study. Journal of occupational and environmental medicine. 2011:1068-77.

121. Scélo G, Constantinescu V, Csiki I, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, et al. Occupational exposure to vinyl chloride, acrylonitrile and styrene and lung cancer risk (Europe). Cancer causes & control. 2004;15(5):445-52.

122. Corbin M, McLean D, Mannetje At, Dryson E, Walls C, McKenzie F, et al. Lung cancer and occupation: A new zealand cancer registry‐based case–control study. American journal of industrial medicine. 2011;54(2):89-101.
123. Brenner DR, Boffetta P, Duell EJ, Bickeboeller H, Rosenberger A, McCormack V, et al. Previous lung diseases and lung cancer risk: a pooled analysis from the International Lung Cancer Consortium. American journal of epidemiology. 2012;176(7):573-85.
124. Richiardi L, Boffetta P, Simonato L, Forastiere F, Zambon P, Fortes C, et al. Occupational risk factors for lung cancer in men and women: a population-based case–control study in Italy. Cancer Causes & Control. 2004;15(3):285-94.
125. Labreche F, Goldberg MS, Valois M-F, Nadon L. Postmenopausal breast cancer and occupational exposures. Occupational and environmental medicine. 2010;67(4):263-9.
126. Davis P, McLeod K, Ransom M, Ongley P, Pearce N, Howden‐Chapman P. The New Zealand Socioeconomic Index: developing and validating an occupationally‐derived indicator of socio‐economic status. Australian and New Zealand journal of public health. 1999;23(1):27-33.
127. Ahrens W, Merletti F. A standard tool for the analysis of occupational lung cancer in epidemiologic studies. International journal of occupational and environmental health. 1998;4(4):236-40.
128. Mirabelli D, Chiusolo M, Calisti R, Massacesi S, Richiardi L, Nesti M, Merletti F. Database of occupations and industrial activities that involve the risk of pulmonary tumors. Epidemiologia e prevenzione. 2001;25(4-5):215-21.
129. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? Psychological methods. 2006;11(2):193.
130. Richiardi L, Barone-Adesi F, Merletti F, Pearce N. Using directed acyclic graphs to consider adjustment for socioeconomic status in occupational cancer studies. Journal of Epidemiology & Community Health. 2008;62(7):e14-e.
131. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evidence-based mental health. 2019;22(4):153-60.
132. Wong MC, Lao XQ, Ho K-F, Goggins WB, Shelly L. Incidence and mortality of lung cancer: global trends and association with socioeconomic status. Scientific Reports. 2017;7(1):14300.

133. ILO. International Standard Classification of Occupations. Geneva, Switzerland: International Labour Office; 1968.

134. . Vienna, Austria2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

135. Applebaum KM, Ray RM, Astrakianakis G, Gao DL, Thomas DB, Christiani DC, et al. Evidence of a paradoxical relationship between endotoxin and lung cancer after accounting for left truncation in a study of Chinese female textile workers. Occupational and environmental medicine. 2013;70(10):709-15.

136. Christensen KY, Lavoué J, Rousseau M-C, Siemiatycki J. Lack of a protective effect of cotton dust on risk of lung cancer: evidence from two population-based case-control studies. BMC cancer. 2015;15(1):212.

137. Muscat JE, Stellman SD, Richie Jr JP, Wynder EL. Lung cancer risk and workplace exposures in black men and women. Environmental research. 1998;76(2):78-84.

138. Xue Y, Jiang Y, Jin S, Li Y. Association between cooking oil fume exposure and lung cancer among Chinese nonsmoking women: a meta-analysis. OncoTargets and therapy. 2016;9:2987.

139. Chen T-Y, Fang Y-H, Chen H-L, Chang C-H, Huang H, Chen Y-S, et al. Impact of cooking oil fume exposure and fume extractor use on lung cancer risk in non-smoking Han Chinese women. Scientific reports. 2020;10(1):1-10.

140. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 71. Re-evaluation of Some Organic Chemicals, Hydrazine and Hydrogen Peroxide. Lyon: IARC (International Agency for Research on Cancer); 1999.

141. Walker JT, Bloom TF, Stern FB, Okun AH, Fingerhut MA, Halperin WE. Mortality of workers employed in shoe manufacturing. Scandinavian journal of work, environment & health. 1993:89-95.

142. Warden H, Richardson H, Richardson L, Siemiatycki J, Ho V. Associations between occupational exposure to benzene, toluene and xylene and risk of lung cancer in Montréal. Occupational and Environmental Medicine. 2018;75(10):696-702.

143. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 100F. Chemical Agents and Related Occupations. Lyon: IARC (International Agency for Research on Cancer); 2012.

144. Kjærheim K, Andersen A. Cancer incidence among waitresses in Norway. Cancer Causes & Control. 1994;5(1):31-7.

145. McGuire V, Nelson LM, Koepsell TD, Checkoway H, Longstreth Jr W. Assessment of occupational exposures in community-based case-control studies. Annual review of public health. 1998;19(1):35-53.

146. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and intelligent laboratory systems. 1987;2(1-3):37-52.

147. Burstyn I. Principal component analysis is a powerful instrument in occupational hygiene inquiries. Annals of Occupational Hygiene. 2004;48(8):655-61.

148. Kaiser HF. The application of electronic computers to factor analysis. Educational and psychological measurement. 1960;20(1):141-51.

149. Siemiatycki J, Day NE, Fabry J, Cooper JA. Discovering carcinogens in the occupational environment: A novel epidemiologic approach. Journal of the National Cancer Institute. 1981;66(2):217-25.

150. Siemiatycki J, Dewar R, Richardson L. Costs and statistical power associated with five methods of collecting occupation exposure information for population-based case-control studies. Am J Epidemiol. 1989;130(6):1236-46.

151. Bertin M, Thébaud-Mony A, Counil E. Do women and men have the same patterns of multiple occupational carcinogenic exposures? Results from a cohort of cancer patients. Annals of Work Exposures and Health. 2018;62(4):450-64.

152. Kennedy SM, Koehoorn M. Exposure assessment in epidemiology: does gender matter? American journal of industrial medicine. 2003;44(6):576-83.

153. Marinaccio A, Corfiati M, Binazzi A, Di Marzio D, Scarselli A, Ferrante P, et al. The epidemiology of malignant mesothelioma in women: gender differences and modalities of asbestos exposure. Occupational and environmental medicine. 2018;75(4):254-62.

154. Pintos J, Parent M-E, Richardson L, Siemiatycki J. Occupational exposure to diesel engine emissions and risk of lung cancer: evidence from two case-control studies in Montreal, Canada. Occupational and environmental medicine. 2012;69(11):787-92.

155. Dominion Bureau of Statistics. Standard industrial classification manual. Ottawa: Information Canada. 1970.

156. Siemiatycki J, Fritschi L, Nadon L, Gerin M. Reliability of an expert rating procedure for retrospective assessment of occupational exposures in community-based case-control studies. American journal of industrial medicine. 1997;31(3):280-6.

157. Siemiatycki J, Lavoue J. Availability of a new job-exposure matrix (CANJEM) for epidemiologic and occupational medicine purposes. Journal of occupational and environmental medicine. 2018.

158. Sauve JF, Siemiatycki J, Labreche F, Richardson L, Pintos J, Sylvestre MP, et al. Development of and selected performance characteristics of CANJEM, a general population job-exposure matrix based on past expert assessments of exposure. Annals of work exposures and health. 2018.

159. Dupont WD. Power calculations for matched case-control studies. Biometrics. 1988:1157-68.

160. Stevenson M, Stevenson MM, BiasedUrn I. Package 'epiR'. 2020.

161. Xu M, Ho V, Siemiatycki J. Role of occupational exposures in lung cancer risk among women. Occupational and environmental medicine. 2021;78(2):98-104.

162. Benke G, Sim M, Fritschi L, Aldred G, Forbes A, Kauppinen T. Comparison of occupational exposure using three different methods: hygiene panel, job exposure matrix (JEM), and self reports. Applied Occupational and Environmental Hygiene. 2001;16(1):84-91.

163. Offermans NSM, Vermeulen R, Burdorf A, Peters S, Goldbohm RA, Koeman T, et al. Comparison of expert and job-exposure matrix-based retrospective exposure assessment of occupational carcinogens in the Netherlands Cohort Study. Occupational and Environmental Medicine. 2012;69(10):745-51.

164. Peters S, Vermeulen R, Cassidy A, Mannetje At, van Tongeren M, Boffetta P, et al. Comparison of exposure assessment methods for occupational carcinogens in a multi-centre lung cancer case–control study. Occupational and environmental medicine. 2011;68(2):148-53.

165. Siemiatycki J. Historical overview of occupational cancer research. Occupational cancers: Springer; 2020. p. 1-20.

166. Loomis D, Guha N, Hall AL, Straif K. Identifying occupational carcinogens: an update from the IARC Monographs. Occupational and environmental medicine. 2018;75(8):593-603.

167. Saritha V, Grace Theodora K, Vijaya Rachel K. Essential Role of Occupational Hazards in Cancer among Women. Handbook of Oxidative Stress in Cancer: Mechanistic Aspects. 2020:1-13.

168. Ali I, Högberg J, Hsieh J-H, Auerbach S, Korhonen A, Stenius U, Silins I. Gender differences in cancer susceptibility: role of oxidative stress. Carcinogenesis. 2016;37(10):985-92.

169. De Matteis S, Consonni D, Lubin JH, Tucker M, Peters S, Vermeulen RC, et al. Impact of occupational carcinogens on lung cancer risk in a general population. International journal of epidemiology. 2012;41(3):711-21.

170. Field J, Smith D, Duffy S, Cassidy A. The Liverpool lung Project research protocol. International journal of oncology. 2005;27(6):1633-45.

171. Siemiatycki J, Wacholder S, Richardson L, Dewar R, Gérin M. Discovering carcinogens in the occupational environment: methods of data collection and analysis of a large case-referent monitoring system. Scandinavian journal of work, environment & health. 1987:486-92.

172. Lacourt A, Cardis E, Pintos J, Richardson L, Kincl L, Benke G, et al. 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. 2013;13(1):1-11.

173. Lacourt A, Labrèche F, Goldberg MS, Siemiatycki J, Lavoué J. Agreement in occupational exposures between men and women using retrospective assessments by expert coders. Annals of Work Exposures and Health. 2018;62(9):1159-70.

174. Xu M, Ho V, Lavoue J, Richardson L, Siemiatycki J. Concordance of Occupational Exposure Assessment between the Canadian Job-Exposure Matrix (CANJEM) and Expert Assessment of Jobs Held by Women. Annals of Work Exposures and Health. 2022.

175. Liu A. Is the impact of cigarette smoking on lung cancer risk different between males and females?: McGill University (Canada); 2012.

176. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond “p< 0.05”. Taylor & Francis; 2019. p. 1-19.

177. Hayes RB. The carcinogenicity of metals in humans. Cancer Causes & Control. 1997;8(3):371-85.

178. Wild P, Bourgkard E, Paris C. Lung cancer and exposure to metals: the epidemiological evidence. Cancer Epidemiology. 2009:139-67.

179. Engholm G, Palmgren F, Lynge E. Lung cancer, smoking, and environment: a cohort study of the Danish population. Bmj. 1996;312(7041):1259-63.

180. Lenters V, Basinas I, Beane-Freeman L, Boffetta P, Checkoway H, Coggon D, et al. Endotoxin exposure and lung cancer risk: a systematic review and meta-analysis of the published literature on agriculture and cotton textile workers. Cancer Causes & Control. 2010;21(4):523-55.

181. Basagana X, Pedersen M, Barrera-Gomez J, Gehring U, Giorgis-Allemand L, Hoek G, et al. Analysis of multicentre epidemiological studies: contrasting fixed or random effects modelling and meta-analysis. International Journal of Epidemiology. 2018;47(4):1343-54.

182. Peters S. Although a valuable method in occupational epidemiology, job-exposure matrices are no magic fix. Scandinavian Journal of Work, Environment & Health. 2020;46(3):231-4.

183. Siemiatycki J. Exposure assessment in community-based studies of occupational cancer. Occupational Hygiene. 1996;3(1):41-58.

184. Siemiatycki J. Job-exposure matrices. In: Encyclopedia of Epidemiologic Methods, eds MH Gail and J Benichou: Wiley; 2000.

185. Richiardi L, Boffetta P, Merletti F. Analysis of nonresponse bias in a population-based case–control study on lung cancer. Journal of clinical epidemiology. 2002;55(10):1033-40.

186. Stang A, Jöckel K. Low response rate studies may be less biased than high response rate studies. Am J Epidemiol. 2004;159:204-10.

187. Slattery ML, Edwards SL, Caan BJ, Kerber RA, Potter JD. Response rates among control subjects in case-control studies. Annals of epidemiology. 1995;5(3):245-9.

188. Xu M, Richardson L, Campbell S, Pintos J, Siemiatycki J. Response rates in case-control studies of cancer by era of fieldwork and by characteristics of study design. Annals of Epidemiology. 2018;28(6):385-91.

189. Guidelines for Carcinogen Risk Assessment. United States Environmental Protection Agency; 2005.

190. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, Volume 83 Tobacco smoke and involuntary smoking: World Health Organization, International Agency for Research on Cancer; 2004.

191. Siemiatycki J, Xu M. Occupational causes of cancer. Handbook of Disability, Work and Health. 2020:127-51.

192. Grant K, Goldizen FC, Sly PD, Brune M-N, Neira M, van den Berg M, Norman RE. Health consequences of exposure to e-waste: a systematic review. The lancet global health. 2013;1(6):e350-e61.

193. Kim S, Xu X, Zhang Y, Zheng X, Liu R, Dietrich K, et al. Metal concentrations in pregnant women and neonates from informal electronic waste recycling. Journal of exposure science & environmental epidemiology. 2019;29(3):406-15.

194. McAllister L, Magee A, Hale B. Women, e-waste, and technological solutions to climate change. Health & Hum Rts J. 2014;16:166.

Appendices

Appendix 1. Ethics approval for the Montreal case-control study of lung cancer.

McGill

Faculty of Medicine
3655 Promenade Sir William Osler #633
Montreal, QC H3G 1Y6

Faculté de médecine
3655, Promenade Sir William Osler #633
Montréal, QC H3G 1Y6

Fax/Télécopieur: (514) 398-3870
Tél/Tel: (514) 398-3124

December 10, 2019

Dr. Jack Siemiatycki
CHUM - Pavillon S
850, rue St-Denis, Bureau S03-448
Montreal (Québec) H2X 0A9

RE: IRB Study Number A12-E06-99A
Case-Control Study of Occupational Risk Factors for Lung Cancer

Dear Dr. Siemiatycki,

Thank you for submitting an application for Continuing Ethics Review for the above-referenced study.

The study progress report underwent review, and Full Board re-approval was provided on December 9, 2019. The ethics certification renewal is valid until **December 7, 2020.**

The Investigator is reminded of the requirement to report all IRB approved protocol and consent form modifications to the Research Ethics Offices (REOs) for the participating study sites, when applicable. Please contact the individual research ethics offices for instructions on how to proceed. Research funds may be withheld, and/or the study's data may be revoked for failing to comply with this requirement.

Please promptly notify the IRB office of any study modifications or unanticipated events that may occur to the study prior to the next annual renewal. Study modifications cannot be implemented prior to an ethics review and approval by the Board.

Regards,

[redacted]

Roberta Palmour, PhD
Chair
Institutional Review Board

cc: Natasha Graham
SMH-REB
A12-E06-99A

Appendix 2. List of variables available in CANJEM for each combination of occupation, time period, and agent.

Variable	Description
idchem	Agent code
agent.label	Agent label
isco.code	Occupation code
isco.label	Occupation label
ntot	Number of individual jobs with this occupation code
nsub	Number of individual subjects with this occupation code
nexp	Number of individual jobs with this occupation code that are exposed
nexp.s	Number of individual subjects with this occupation code who are exposed
p	Proportion of jobs with this occupation code that are exposed (nexp / ntot)
n.R1	Number of exposed jobs in this occupation code with low reliability
n.R2	Number of exposed jobs in this occupation code with medium reliability
n.R3	Number of exposed jobs in this occupation code with high reliability
n.C1	Number of exposed jobs in this occupation code with low intensity
n.C2	Number of exposed jobs in this occupation code with medium intensity
n.C3	Number of exposed jobs in this occupation code with high intensity
n.F1	Number of exposed jobs in this occupation code with less than 2 hours exposed per week
n.F2	Number of exposed jobs in this occupation code with 2-12 hours exposed per week
n.F3	Number of exposed jobs in this occupation code with 12-39 hours exposed per week
n.F4	Number of exposed jobs in this occupation code with 40 hours or more exposed per week
p.R1	Proportion of jobs in this occupation code with low reliability
p.R2	Proportion of jobs in this occupation code with medium reliability
p.R3	Proportion of jobs in this occupation code with high reliability
p.C1	Proportion of jobs in this occupation code with low intensity
p.C2	Proportion of jobs in this occupation code with medium intensity
p.C3	Proportion of jobs in this occupation code with high intensity

p.F1	Proportion of jobs in this occupation code with less than 2 hours exposed per week
p.F2	Proportion of jobs in this occupation code with 2-12 hours exposed per week
p.F3	Proportion of jobs in this occupation code with 12-39 hours exposed per week
p.F4	Proportion of jobs in this occupation code with 40 hours or more exposed per week
Cmoy.1	Mean intensity (1-2-3 scale)
Cmoy.3	Mean intensity (1-3-9 scale)
Cmoy.5	Mean intensity (1-5-25 scale)
Cmoy.10	Mean intensity (1-10-100 scale)
Dmoy.1	Mean frequency weighted intensity (1-2-3 scale)
Dmoy.3	Mean frequency weighted intensity (1-3-9 scale)
Dmoy.5	Mean frequency weighted intensity (1-5-25 scale)
Dmoy.10	Mean frequency weighted intensity (1-10-100 scale)
Cmed.1	Median intensity (1-2-3 scale)
Cmed.3	Median intensity (1-3-9 scale)
Cmed.5	Median intensity (1-5-25 scale)
Cmed.10	Median intensity (1-10-100 scale)
Dmed.1	Median frequency weighted intensity (1-2-3 scale)
Dmed.3	Median frequency weighted intensity (1-3-9 scale)
Dmed.5	Median frequency weighted intensity (1-5-25 scale)
Dmed.10	Median frequency weighted intensity (1-10-100 scale)
Fmoy	Median frequency (continuous)
Fmed	Mean frequency (continuous)