Concawe
Environmental Science for European Refining

Report

Report no. 15/18

Air Pollution and Lung Cancer: A Review of Issues Affecting the Interpretation of the Epidemiological Literature

ISBN 978-2-87567-096-0

Air Pollution and Lung Cancer: A Review of Issues Affecting the Interpretation of the Epidemiological Literature

Prepared by:

J. A. Tomenson (Causation Ltd)
D. Morgott (Pennsport Consulting)
G. B. Copley (ExxonMobil Biomedical Sciences, Inc.)

for the Concawe Health Management Group:

A. Steneholm (Chair)
M. Banton
S. Blazquez Borras
P. Boogaard
W. Doig
H. Eggers
I. Fragkou
P. Gostomski
F. Guerra
R. Hambach
J. Lewis
R. Marschall
V. Moukrioti
G. Pizzella
M. Samuelsen
B. Savonius
R. Schnatter
B. Schubert
S. Schunder-Tatzber
J. Urbanus
M. Vaissiere
C. Verge Lopez
H. Ketelslegers (Science Executive for Health)
M. Trantallidi (Research Associate for Health)

Brussels
December 2018

ABSTRACT

The following report examines key issues that need to be considered when evaluating the attributes and implications of studies examining the association between ambient air pollution and lung cancer. Following a brief general discussion of the types of epidemiology studies that can be used to investigate the association between an environmental or occupational exposure and a particular health outcome, the report goes on to examine specific topics that need to be considered when evaluating the strength and weakness of any relationships that are purported to exist. Areas of focus include exposure estimation, confounding, quantitative risk assessment, heterogeneity, and plausibility. Each of these topics is explored in detail and information is provided showing how reported relative risk estimates may have been impacted by the failure to fully evaluate or consider specific methodological, procedural, or interpretive characteristics of the study. As such, the aim of the report is to highlight some generally overlooked areas of inquiry that need to be addressed in order to frame and draw conclusions from the results of a chronic health effects investigation focusing on lung cancer and air pollutant exposures.

KEYWORDS

Confounding, exposure estimation, misclassification, heterogeneity, modelling, plausibility

INTERNET

This report is available as an Adobe pdf file on the Concawe website (www.concawe.eu).

CONTENTS

SUMMARY

In 2013 the authors of this report were observers at the IARC monograph meeting (Volume 109) to assess the carcinogenicity of outdoor air pollution. In preparation for the meeting, some time was spent reviewing the strengths and limitations of the many potentially relevant epidemiological studies, and assessing whether they would permit a causal interpretation. This report attempts to capture some of those discussions with the aim of highlighting some generally overlooked areas of inquiry that need to be addressed in order to frame and draw conclusions from the results of studies of the chronic health effects of air pollution exposures, especially studies of lung cancer.

Section 2 looks generically at the different study designs that have been used to examine the association between lung cancer and exposure to outdoor air pollution, and discusses why the results of some studies are more informative than others. It also highlights the studies that the IARC Working Group considered to be the most informative and gave greatest weight to in their evaluation. These were all studies that possessed quantitative exposure data, primarily cohort studies within the general population, but also some population-based case-control studies that were considered sufficiently informative.

Section 3 looks in detail at that exposure data and the methodologies used to derive it. Two of the most commonly used approaches involve either interpolating measurements from central monitoring sites or emission sources to an address using methods with varying degrees of sophistication, or predicting pollution at an address from land use characteristics, population density, and traffic patterns (land use regression models). It is noted that even if average pollution levels at the doorsteps of subjects were known exactly for the time periods of interest, it is unlikely that they will correlate well with personal inhalation concentration because of the use of only the residence location to estimate exposure (and not the workplace, school, or transportation environments etc.), and the differences between indoor and outdoor air at these locations.

The lack of proper validation of exposure estimates and failure to estimate exposure during the relevant time period (at least 10-15 years previously in the case of lung cancer) are discussed. It is also noted that many studies make an assumption that the within-area contrasts in the spatial distribution defined by their exposure estimates may be a reasonable surrogate for the real contrasts in exposure levels, and hence can be used to investigate associations between cancer and air pollution. This may be justifiable to some degree, but the same assumption has also been used to give credibility to risk estimates per unit of pollution, even in the case where the proxy measure is an estimate of current exposure at the baseline address of participants. Some investigators have recognised that risk estimates based on a recent exposure contrast might be too high with decreasing air pollution concentrations and contrasts over time. Nevertheless, risk estimates from unsuitable studies are often used to estimate the number of deaths likely to result from air pollution. The potential impact of exposure misclassification error is also discussed, which as a recent review of the use of geographically modelled environmental exposure estimates in epidemiological studies correctly noted, is too often dismissed based on the erroneous assumption that an accurately classified exposure would have led to an even stronger estimate of risk (Chang et al., 2014).

It is not difficult to see that lung cancer might appear to be due to air pollution if the people most exposed to air pollution are more likely to be smokers. This problem, known as confounding, is especially important in studies of air pollution

and lung cancer because of the large effect of smoking relative to that of pollution. However, smoking is not the only potential confounder, and even studies of lung cancer risk from outdoor air pollution exposure among never smokers can be biased by confounding from occupational exposure to lung carcinogens and other social class-related factors (Samet et al., 2009). Section 4 looks at how well study investigators have adjusted for smoking and other potential confounders with a focus on the reliability of the confounder measurements used, and the likelihood that confounding remains after adjustment (residual confounding). The majority of studies of particulates and lung cancer possess individual level smoking information, but some rely on proxy measures such as pre-existing illnesses related to smoking habits. Nevertheless, it is noted that residual confounding by smoking is likely in many studies because the smoking information may be many years out of date, and residual confounding by smoking is also possible in studies of never smokers because of misreporting of smoking status and consequent misclassification of current and ex-smokers as never smokers. It is also observed that many investigators claim to have controlled for occupational exposure to lung carcinogens, but the measures used are generally poor or not reported adequately. A related issue is determining whether the associations observed in single pollutant models are confounded by exposure to other pollutants found in ambient air.

Section 5 briefly discusses exposure –response and what the animal studies indicate about potential mechanisms that may be operating and their implications for the shape of the response curve. The few studies which have considered alternatives to a linear exposure-response model have concluded that there is no evidence of a marked deviation from linearity (Hamra et al., 2014), but some had limited information to reach such a conclusion.

Section 6 looks at the consistency of the evidence linking exposure to outdoor pollution with lung cancer. A causal interpretation is generally considered to be strengthened when studies of dissimilar populations, exposure characteristics or research methods yield similar measures of effect, and the IARC Working Group noted that “both cohort and case-control studies with exposures assessed in the population setting, involving millions of subjects and many thousands of lung cancer cases in different parts of the world, consistently showed an association between exposure to outdoor pollution and the risk of lung cancer, in both sexes and after adjustment for the main potential confounders”. However, considerable heterogeneity in study results should be expected in some circumstances because of heterogeneity in study populations, research methodology and the exposure of interest. For instance, particulate matter varies considerably in chemical make-up between locations and hence one might not expect to see the same effects per unit of particulate exposure in studies from different locations. In addition, the effect of exposure to $PM_{2.5}$ might be expected to be more variable than exposure to PM_{10} as it includes a higher proportion of mutagenic species, many of which are products of combustion, and the smaller particles penetrate more deeply into the lung and are more likely to be retained, whilst the coarser PM_{10} consists mainly of minerals and biological materials (Hamra et al., 2014). However, less heterogeneity might be expected to be seen in the results of studies where the same material is studied at different locations e.g. studies of ozone, SO_2 and NO_2. Consistency or the absence of significant heterogeneity is also important when judging how widely the results from a meta-analysis can be generalised to other populations. However, it is demonstrated that considerable heterogeneity may be present which is not detected by standard statistical tests.

Section 7 asks what observable consequences would be expected if air pollution causes lung cancer, and examines whether they are in fact seen in the epidemiological studies i.e. whether the studies are coherent with the theory. This

is formally known as coherency and, like consistency, it is one of the nine Bradford Hill criteria used to establish causality (Hill, 1965). For instance, would one expect to observe a concentration-response relationship between estimates of exposure to particulates in air pollution and lung cancer for smokers when the intake of particulates from outdoor air pollution is tiny compared to the dose from cigarette smoking? Other coherence issues that are considered relate to the findings for adenocarcinoma (the most common cancer observed in smokers) versus those for squamous and small cell lung cancer, and the strength of associations observed for men versus women. Other aspects considered include whether the associations observed between non-malignant respiratory diseases including chronic obstructive pulmonary disease and outdoor air pollution are coherent with those for lung cancer and whether the results seen in occupational studies are coherent with those from population studies. Another question considered is whether measures of different pollutants are reflecting a causal agent or are merely acting as an indicator of pollution (Section 8).

In conclusion, the report shows that there are many issues and attributes that need to be examined when assessing or designing a study evaluating the relationship between particulate exposures and lung cancer. It is particularly important to assess whether the exposure contrast defined by the exposure estimates is a reasonable and validated surrogate for the true exposure contrast. Further, if the study findings are to be used for quantitative risk assessment or global burden calculations, then it is also necessary that the exposure assessment provides a realistic description of the exposures people actually experienced during the biologically-relevant exposure period, and this is unlikely to be achieved if outdated proximity-based approaches are used. Bias and confounding also need to be ruled out with more certainty than can be achieved if information on key potential confounders such as smoking or occupational exposures is at best collected at baseline, or unavailable as is the case for one of the studies considered most informative by IARC. It is correct to attach weight to the findings of meta-analyses, especially combined results for never smokers where confounding by smoking is less of an issue, but implausible and unexplained associations of a comparable or greater magnitude observed in smokers and former smokers weaken this body of evidence. Finally, the epidemiological evidence is not sufficient to distinguish between $PM_{2.5}$ and PM_{10}, or even NO_2 which is presumably only a surrogate for exposure, and better studies are needed to provide a more helpful conclusion for understanding risk than the IARC evaluation that PM is carcinogenic to humans.

1. INTRODUCTION

Scientists and scientific panels are often called upon to evaluate the strength of an epidemiological study independent of establishing any basis for causality. Their goal is to ensure that every precaution has been taken to minimize the influence of extraneous errors on the salient observations emerging from the study. This is an especially important exercise in the air pollution arena where studies often suffer from a host of limitations that may compromise the findings. Whereas, it is often quite easy to generically identify those common sources of error or bias that routinely affects the quality of an epidemiological investigation, it is far more difficult, even for the study investigators, to examine specific sources of error in a fashion that is meaningful for a final risk determination. In some cases, sensitivity analyses may help to address the issue e.g. restricting the analysis to never smokers to exclude the possibility of confounding by smoking, but some assumptions are much more difficult to test. For example, it is often assumed that spatial patterns of air pollution change slowly in a city and that exposure assessment performed today can be a good surrogate of exposure occurring in the past or in the future (Cesaroni et al. 2012). To be successful, reviewers need access to the original data to verify the findings and ensure that adequate precautions have been taken to control for potential sources of bias through an in-depth consideration of the methodological underpinnings of the study (Peng et al., 2006). This effort requires a keen eye, a willingness to dig deep into the particulars of the study design, and a healthy dose of skepticism. Faced with this effort, many are either unwilling or unable to undertake such a huge task given (i) the multitude of studies that may exist for a single health endpoint, (ii) the need to collect and examine an extensive amount of often difficult to acquire raw data; and (iii) the absence of broadly accepted criteria for grading or evaluating the strengths and weaknesses of a particular study. Yet, without such an undertaking it is very difficult to guarantee that the findings en masse are truly indicative of a public health problem that requires the attention of policy makers.

These issues are not unique to the air pollution arena and have reached dramatic proportions in clinical settings where there is increasing awareness of the limitations inherent in the design of observational studies (Grimes and Schulz, 2012) and the frequent inability of these studies to yield reproducible results (Ioannidis, 2008). This has led some researchers to call for the pre-registration of study protocols prior to their initiation to help increase transparency and eliminate the a posteriori data dredging that often takes place (Dal-Re et al., 2014, Young and Karr, 2011). Although pre-registration of study protocols would help eliminate the overzealous interpretation of the findings from a clinical evaluation, efforts are also needed to harmonize the reporting requirements as well. For instance, efforts are now underway to broaden the implementation of new journal reporting requirements such as STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) that describes a host of factors that need to be addressed in a new original research publication (von Elm et al., 2007). Whereas, these innovations are currently confined to the clinical sciences, their implementation in the air pollution arena would ease the burden of evaluating the public health relevance of new studies purporting to show an association between an exposure and any of a wide array of adverse health outcomes (Rooney et al., 2014, Sheehan et al., 2016).

The discussion that follows takes an in-depth look at some of the often over-looked drawbacks that can affect the risk magnitudes reported in recently published epidemiology studies focusing on lung cancer. The goal behind this effort was to reinforce the need for caution and careful scrutiny when examining the results from the myriad of air pollution studies addressing the relationship between exposure

and risk. The issues/topics described are not intended to provide an exhaustive list of every possible source of bias, but they do highlight many of the most important. The topics are presented without prejudice and are aimed at raising awareness of some potential problem areas that may affect ultimately impact the reliability of the findings, be they positive or negative. The report was prepared with the recognition that confounding and bias can never be entirely eliminated, merely contained and controlled in a manner that is transparent and objective. Increasing the rigor used to evaluate the findings from individual studies will give many scientists working outside the confines of classical epidemiology a more favourable impression of the methodologies employed and the conclusions attained.

Much has been written about the bias and confounding that can plague an epidemiological investigation and it is not the purpose of this treatise to review and rehash all of these pitfalls; rather the purpose is to urge a more balanced approach to the evaluation of variables and issues that may be impacting the results. This is always a somewhat contentious and controversial subject that can cause consternation with those who are heavily invested in a particular set of beliefs (Young and Karr, 2011). However, there is now enough evidence to show that the best science is not always applied when honestly and objectively evaluating the results from air pollution studies. Whereas some believe that the results from observational studies are inherently flawed due to their inability to adequately control for confounding at all levels of design and implementation, others are more pragmatic and stress the need for transparency when publishing the results from a particular study so that results can be placed in perspective. Despite the increased willingness of many investigators to share their data to ensure its reproducibility, others are more hesitant and restrict access based on confidentiality concerns (Barnett et al., 2012, Tenopir et al., 2011). As is often the case in science, the quality of an investigation can be viewed as lying along a continuum from good to bad, with the exact position open to debate. Factors such as model selection bias; regional heterogeneity, disease and exposure misclassification, White Hat bias, compositional clustering, multiple comparisons, publication bias, loss to follow-up, inadequate latency, pollutant collinearities, and recall bias are but a few of the issues that need to be considered when grading the quality of an observational study. When these topics are ignored or overlooked, the risks of a false positive or false negative finding increases and the full value of the research investigation may not be realized.

The challenges of studying the relationship between air pollution metrics and health begins with the problems associated with exposure modelling, heterogeneity, and bias. The following pages contain a description of these and other areas of concern within the context of those studies aimed at exploring the relationship between air pollution and lung cancer. It is the intent of the authors to show that the procedures and practices used to interpret the results from these studies could be made more rigorous through an in-depth evaluation of specific methodological, procedural, and expositional characteristics of a chronic health effects investigation.

2. STUDY METHODOLOGIES

This section looks at the different study designs that have been used to examine the association between lung cancer and exposure to outdoor air pollution, and discusses why the results of some studies are more informative than others. Some of the studies include a large number of subjects in the population setting and the most informative have individual-level information on cancer outcomes, cancer risk factors and estimated exposures to outdoor air pollution. The information from these studies can be supplemented by investigations focusing on workers who have been exposed occupationally to outdoor air pollution. Another study design includes populations who have potentially been exposed to local sources of industrial emissions. In total, there are a large number of epidemiology studies that may be considered relevant to an examination of the association of outdoor air pollution with lung cancer risk. However, this report will focus on the body of literature that the IARC Working Group considered to be the most informative and gave greatest weight to in their evaluation (IARC, 2016). The last part of this section reviews and discusses these studies in some detail.

2.1. OCCUPATIONAL STUDIES

Occupational epidemiology studies contrast with environmental epidemiology studies by focusing on the incidence and prevalence of a disease or illness in a workplace setting rather than a community environment. Whereas occupational and environmental epidemiology studies both employ similar methodological designs, they differ in several fundamental respects that can be exploited to improve the reliability and statistical power of an investigative study looking at the relationship between an exposure metric and an adverse outcome (Checkoway et al., 2007). Perhaps the single most important difference between environmental and occupational epidemiology studies concerns the nature and magnitude of the exposures (Brunekreef et al., 2008). Since exposure levels within the workplace are generally higher than what is found in ambient air, health studies with a group of employees are inherently more sensitive than their environmental counterpart, which are frequently unable to accurately define the nature of the exposure-response relationship throughout the ambient concentration range. Consequently, it is not uncommon for epidemiologists to use the results from workplace studies to describe the exposure-response functions needed for a chronic risk characterization. Another potential advantage of conducting epidemiology studies using workplace personnel is that employee exposures are restricted to a limited number of substances encountered at a particular work site. This becomes a benefit when the study population is not exposed to excessive amounts of relevant co-pollutants while away from the jobsite. Under these conditions, the possibility of confounding from pollutant co-exposures is reduced and the biostatistical analysis of the collected data is simplified. This is often not the case in an environmental epidemiology study because ambient air may contain a myriad of chemical and physical agents that may act additively or synergistically to obscure the true nature of the exposure-response relationship.

This is not to say, however, that occupational epidemiology studies are completely free of the bias and confounding that can plague the results from an environmental study. As with environmental epidemiology studies, there is a distinct possibility that exposure misclassification may be affecting the reliability of any findings. This problem arises because oftentimes insufficient personal exposure information is available for the employees of interest. Under these circumstances, researchers are faced with the option of estimating the exposure levels on either an individual or group basis (Loomis and Kromhout, 2004). Although estimates of individual exposure

are always preferred there is always the possibility of recall bias when interviews and questionnaires are used to obtain a semi-quantitative estimate of the exposure magnitude. Likewise, appreciable misclassification may occur when the exposure estimates are derived using expert panels or job exposure matrices to examine a group of individuals working under similar conditions. In fact, a comparison of exposure estimates obtained using an expert panel, a job exposure matrix, and self-reports showed poor to fair agreement across the methods (Benke et al., 2001). Although there continues to be problems with the exposure estimation techniques used with occupational epidemiology studies, the resulting bias from over- or under-estimations of inhalation exposure has been shown to be chemical specific with some substances showing a higher degree of concordance than others (Offermans et al., 2012, Teschke et al., 2002).

In addition to these issues, occupational studies are restricted in the types of individuals that can be examined, since workplace employees constitute healthy adult men and women populations. As such, occupational studies are unable to examine the associations that may exist in the very young, the very old, or the debilitated. Examinations of individuals suffering from rare acute and chronic health conditions found in the general population necessarily require an environmental approach. In addition to restrictions in the population of individuals that can be studied, occupational epidemiology studies include a particular type of selection bias termed the "healthy worker effect" that needs to be controlled to the extent possible (Pearce et al., 2007). If the healthy worker effect is not adequately addressed, morbidity and mortality risks will be lowered because only relatively healthy individuals are being examined. Likewise, a healthy worker survivor effect prolongs survival and reduces loss to follow-up relative to their non-working counterparts.

2.2. COHORT STUDY

This type of study identifies a group of people and follows them over a period of time to see how their exposures affect their outcomes. Cohort studies are normally used to look at the effect of suspected risk factors that cannot be controlled experimentally, for example the effect of air pollution on lung cancer. The objective is to estimate the risks of various diseases among the cohort relative to background risks among persons not exposed to the same environmental factors. Investigators have noted that cohort studies, among all types of epidemiological study designs, are the most accepted among the scientific community for two main reasons; firstly cohort studies usually include everyone from the available study population (rather than only a sample) and secondly because they most closely resemble a controlled randomised experiment such as an animal toxicology experiment or a randomised clinical trial (Checkoway et al., 2004).

There are two main types of cohort studies: prospective and retrospective (or historical). In a prospective cohort study, a group of people is identified at the time the study is being conducted and they are followed over a period of time to see how their exposures affect their outcomes. In general, prospective designs are better suited for examining health outcomes that develop within a relatively short period of time and less commonly used for studies of cancer and other diseases that have long induction and latency periods. However, many of the major cohort studies of air pollution and lung cancer such as the European Study of Cohorts for Air Pollution Effects (ESCAPE) (Raaschou-Nielsen et al., 2013), the study based on data collected by the American Cancer Society (ACS) as part of the Cancer Prevention Study II (CPS-II) (Krewski et al., 2009, Pope et al., 2002, Turner et al., 2011); and the Harvard Six Cities (H6C) study (Laden et al., 2006b, Lepeule et al., 2012) have a prospective design. In a retrospective study, a group of people are identified at some time in

the past and the cohort is then follower over historical time to estimate disease rates. The basic study design features of retrospective cohort studies are the same as those of prospective studies. Retrospective studies rely on exposure data and/or outcomes that have already been collected (through medical records or as part of another study). Data used in this way may not be as reliable as data collected prospectively as it relies on the accuracy of records made at the time and on the recall of historical events, which can be inaccurate. Retrospective cohort studies are increasingly using administrative data (e.g., healthcare system data) due to the relatively low data acquisition costs and the convenience (i.e., no recruitment, no enrolment into study, no primary data collection). Perhaps the most notable example of administrative data being used for cohort studies is the suite of studies based on the US Medicare Cohort (Zeger et al., 2008). Often overlooked is the fact that these databases were not developed for research purposes, so data are not available for many candidate covariates/confounders of interest. This requires researchers to find or design surrogates for these covariates, often of ecological origin (e.g., percent home ownership in a census tract as a stand-in for socioeconomic status) that might not always be indicative of a particular person's circumstances. Likewise, environmental pollution exposure data are often absent, requiring a separate effort to acquire that information. Pollutant data are typically acquired for a timeline that runs concurrently with the retrospective "follow-up" which excludes the etiologically relevant and significantly higher exposures that occurred prior to the period under study. Instead, observed effects are attributed to proximate exposure concentrations, essentially an out-of-sample extrapolation. This issue is covered in more detail in section 3.2.3 of this report. Other examples of retrospective cohort studies of air pollution include the Trucking Industry Particle Study (TrIPS) which examined associations between mortality and air pollution in a cohort that was identified using trucking industry employment records (Hart et al., 2011). Two others include the Clinical Practice Research Datalink (CPRD) study that used a national cohort of adults registered with family practitioners in England, and the Rome Longitudinal Study (RoLS) which identified subjects using census records (Carey et al., 2013, Cesaroni et al., 2013).

2.3. CASE-CONTROL STUDY

In this type of study, a group of individuals who have a particular condition or disease (cases) and a group of individuals that do not (controls) are identified without knowledge of prior exposure history and compared with respect to existing or past exposure. In contrast to a cohort study, which selects subjects who are initially free of disease and follows them over time to determine rates of disease in the absence or presence of exposure, a case-control study selects subjects on the basis of the presence or absence of the disease under study. However, both methods allow the investigator to estimate the effect of exposure on the risk of disease. Controls can be selected as a simple random sample or matched to cases on a group or individual basis with respect to one or more potential confounding factors (i.e. factors related to both exposure and health outcome such as smoking in studies of lung cancer and air pollution) (Checkoway et al., 2004). However, it should be noted that the aim of matching in case-control studies is not to prevent confounding (bias resulting from not controlling for the effect of confounders - see Section 4). It is usually performed to achieve a more efficient statistical design for control of confounding (i.e. to achieve a more precise estimate of effect) (Rothman et al., 2008). For example, Vineis et al. performed a case-control study to estimate the relationship between air pollution and lung cancer and matched three controls per case (Vineis et al., 2006). The matching criteria were gender, age (65 years), smoking status, country of recruitment and the time elapsed between recruitment and diagnosis. The authors noted that matching was introduced to allow strict control of potentially confounding variables, as all the selected risk factors may

have been stronger in their effect than air pollution. Regardless of whether matching is used to select controls, it is essential that the controls are sampled from the entire source population that gives rise to the cases and that specific steps are taken to reduce or eliminate any selection bias that can distort the magnitude of any observed relationships (Rothman et al., 2008). A distinction is sometimes made between the sources from which cases and controls are selected. In population-based case-control studies, all cases of the study disease occurring within a defined area during a specified period of time are ascertained and controls are a sample of disease-free individuals from the same area. In hospital-based studies, all cases of the disease of interest in a hospital population are ascertained during a specified period of time, and the controls are selected from persons admitted to the same hospital population for conditions other than the studied disease.

Case-control studies tend to be ranked below cohort studies in terms of weight of evidence. Ascertainment and selection bias can be a particular problem in case-control studies and may result in misleading associations and biased estimates of relative risk. However, these problems are not unique to case-control studies. Bias in the estimation of exposure is also a problem, in particular subject recall bias and interviewer bias. There have been relatively few population-based case-control studies of lung cancer and air pollution compared to the number of cohort studies. However, there have been several case-control studies of populations potentially exposed to local sources of industrial emissions. The most informative of these include the Canadian National Enhanced Cancer Surveillance System (CNECSS) case-control study and a retrospective case-control study of lung cancer in residents of Stockholm, Sweden (Hystad et al., 2013, Nyberg et al., 2000).

2.4. ECOLOGICAL STUDIES

Ecological studies are studies in which data are available for groups only, rather than for the individuals within the groups. Many of the early studies of air pollution were ecological studies because individual exposures to environmental factors, such as pollutants in air or water, were difficult and expensive to obtain, whereas pollution measures taken from monitoring sites were routinely available. For example, Chinn et al. looked at the relationship between routinely available mortality rates for 116 counties and London boroughs of England and Wales for the period 1969-1973 and mean winter levels in 1971 of smoke and sulphur dioxide (SO_2) in the same boroughs collected in a national survey (Chinn et al., 1981). Some later studies such as the H6C study and the ACS CPS-II used a semi-ecological design in which individual outcomes and confounders were available, but with an ecological exposure measure (Dockery et al., 1993, Pope et al., 1995). A major problem of ecological studies is that relationships at an ecological level cannot be assumed to hold for the individuals within the groups, and ecological studies usually suffer from the unavailability of data necessary for adequate control of confounding. Ecological studies in epidemiology have yielded important insights, but it is easy to draw incorrect conclusions from aggregate data. This problem of incorrect inference is known as the ecological fallacy and Greenland and Robins provide a review of the issues (Greenland and Robbins, 1994). One important issue is that a risk factor may be an ecological confounder even if it is not associated with exposure in every group (i.e. at an individual level, a risk factor must be associated with exposure and the outcome in order to be a confounder). In contrast a risk factor that is a confounder within groups may not be an ecological confounder. These problems mean that ecological studies may produce results that are of questionable validity and are usually only regarded as hypothesis-generating because of concern about bias (Rothman et al., 2008). However, there are certain situations, albeit limited, in which ecological studies can be particularly informative (Savitz, 2012).

2.5. EPIDEMIOLOGICAL STUDIES FOCUSSED ON IN THIS REPORT

The IARC Working Group gave greatest weight to cohort studies with quantitative exposure data within the general population. Among the North American cohort studies of this nature, results from the H6C study, the never smokers analyses of the ACS CPS-II study, and the prospective California Teachers study (CTS) were judged to be the most informative (Lepeule et al., 2012, Lipsett et al., 2011, Turner et al., 2011). The ESCAPE and RoLS studies were judged to be among the most informative European cohort studies (Cesaroni et al., 2013, Raaschou-Nielsen et al., 2013). It was also noted that the most influential European cohort studies included investigations from Norway, the Netherlands and Denmark (Beelen et al., 2008a, Naess et al., 2007, Nafstad et al., 2003, Raaschou-Nielsen et al., 2010, Raaschou-Nielsen et al., 2011a). The most useful cohort studies from other areas of the world included two studies from Japan including the Three-Prefecture Cohort Study (TPCS), and a study from New Zealand (Hales et al., 2012, Katanoda et al., 2011, Yorifuji et al., 2013). Overall, the ESCAPE and ACS CPS-II were considered to be the most revealing cohort studies.

The IARC Working Group categorized case-control studies according to whether the main type of exposure was from all sources including traffic pollution, or specific industrial sources. Studies examining all sources of air pollution were grouped according to whether the exposure methodology used was qualitative (or semi-quantitative) or quantitative. Case-control studies conducted before 1990 were considered to have too many limitations to be informative and were given little weight. Also case-control studies of populations who have potentially been exposed to local sources of industrial emissions were considered to be less informative because of the unique nature of their exposures and because of other methodological limitations similar to those observed with the early case-control studies. However, three population-based studies conducted in Canada, Poland and Sweden possessed quantitative exposure measures and were considered to be suitably informative (Hystad et al., 2013, Jedrychowski et al., 1990, Nyberg et al., 2000).

The other category of studies reviewed by the IARC Working Group were occupational cohort and case-control studies of outdoor workers whose exposure was considered to be of the same nature as that of the general population. The occupations included professional drivers, traffic police, mail carriers, and filling station attendants. However, studies of workers exposed to specific sources of pollution such as diesel or gasoline engine emissions such as underground workers were not reviewed as they had been analysed in a previous IARC monograph, although it was noted that these sources are important contributors to urban air pollution (IARC, 2013). These studies were also not discussed in the evaluation of human carcinogenicity data made by the IARC Working Group. Ecological studies of lung cancer and industrial emissions were also not reviewed. This is not surprising given that case-control studies of industrial emissions were considered to be less informative than other case-control studies, although the Working Group did review them for other cancer endpoints.

This report focuses on the studies listed above that the IARC Working Group considered to be the most informative about the relationship between the general air pollution mixture and lung cancer. In addition, the report will focus on studies included in a meta-analysis of the lung cancer risk associated with exposure to particulate matter (PM) in outdoor air, which has recently been published as a result of the IARC review (Hamra et al., 2014). It was noted by the authors that the studies included in this analysis were a key component of the epidemiological evidence reviewed by the IARC Working Group in its evaluation of the carcinogenicity of PM,

and that the quantitative analyses complemented the qualitative classification of the evidence by the IARC Working Group (Hamra et al., 2014). The study selection process was identical to that of the IARC review, except that one recently accepted paper by Puett et al. was included in the analysis because it met the criteria for inclusion (Puett et al., 2014). A second meta-analysis originating with the IARC review has also been published which was an analysis of studies examining exposure to measures of nitrogen oxides and other measures of traffic exposure and lung cancer (Hamra et al., 2015). Although the literature search was performed three months after the IARC review, no other relevant papers were identified.

3. EXPOSURE ASSESSMENT

Exposure assessment is perhaps the single largest source of error in many environmental epidemiology studies focusing on the acute and chronic health effects of air pollutants. Estimates of outdoor pollution concentration at the residence location are usually used as a surrogate for pollution exposure inside the home, which is where most time is spent. Investigators generally use any of several different metrics to define the exposure magnitude. These include techniques that define exposure classes based on geographical location, spatial and temporal extrapolation approaches, and model-based procedures (e.g. land use regression) to derive estimates for a residence location at a particular point in time. This section discusses the methodologies that are used to estimate the residential exposure of a subject and assesses whether their predictive ability is as claimed.

In addition, the section discusses the methods used to derive an estimate of relevant individual exposure with a particular focus on lung cancer. Even if average pollution levels at the doorsteps of subjects were known exactly for the time periods of interest, the measures of pollution exposure derived by the investigators of most studies would still be subject to error because of the reliance on self-reported residential histories and incomplete coverage of lifetime exposure. More importantly, it is unlikely that doorstep pollution levels will correlate well with personal inhalation concentration because of the use of only the residence location to estimate exposure (and not other microenvironments such as the workplace, school, or transportation environments etc.) and the effect of infiltration at the residence location (and other microenvironments where subjects spend their time).

This section also deals with the impact of measurement error. High predictive ability is often claimed for the exposure surrogate, but the impact of measurement error is often ignored. Sheppard et al. noted that most air pollution epidemiology studies report estimates of health effects conditional on measured or predicted exposures without regard to how these exposure estimates were obtained ("plug-in" exposures) (Sheppard et al., 2012). These issues are considered further below, but focus on only a small component of the measurement error.

Finally, this section discusses how risk estimates can be interpreted and used to evaluate the true magnitude of a health problem.

3.1. ESTIMATING EXPOSURE

3.1.1. Methodologies

Many exposure methodologies have been employed over the years to combat known deficiencies in available data and they continue to evolve as new geostatistical and biostatistical insight is brought to bear on the problem of exposure misclassification. As a result, exposure measurement within the context of ambient air pollution has developed into a separate and unique discipline.

The increasing sophistication and complexity of the exposure models has caused some to question whether exposure measurement techniques are becoming too complex for use by anyone except a few dedicated experts (Briggs, 2007). Others, on the other hand, feel that the opposite is true and that exposure models can help reduce misclassification by explaining how ambient concentrations of a pollutant change over space and time (Jerrett et al., 2005). In the absence of detailed knowledge of an individual's personal exposure throughout the day, modelling provides a useful alternative, but there are pitfalls that must be closely scrutinized

lest the results contain an unacceptably high degree of bias that invalidates the findings. This is not an easy exercise to perform since a large variety of exposure models are available for use in an environmental epidemiology study. For instance, the most recent entry into the field of exposure modelling has been the introduction of satellite remote sensing technologies that allow discrete measurement of individual pollutants at precise locations (Prud'homme et al., 2013). Although they have been shown to equate well with measurements taken at monitoring sites, they cannot, as yet, account for all of the microenvironmental inhalation exposures that individuals experience as they move through the day. In addition to remote sensing, there is a litany of methods that can be adapted for use in a study of air pollutant health effects; and the final selection is often dictated by the study design, the background of the senior investigator, and the availability of expert collaborators.

Available air pollutant exposure models yield both continuous and categorical metrics depending on the focus of the investigation. Figure 1 provides a hierarchical breakdown of these exposure metrics moving from best to worst.

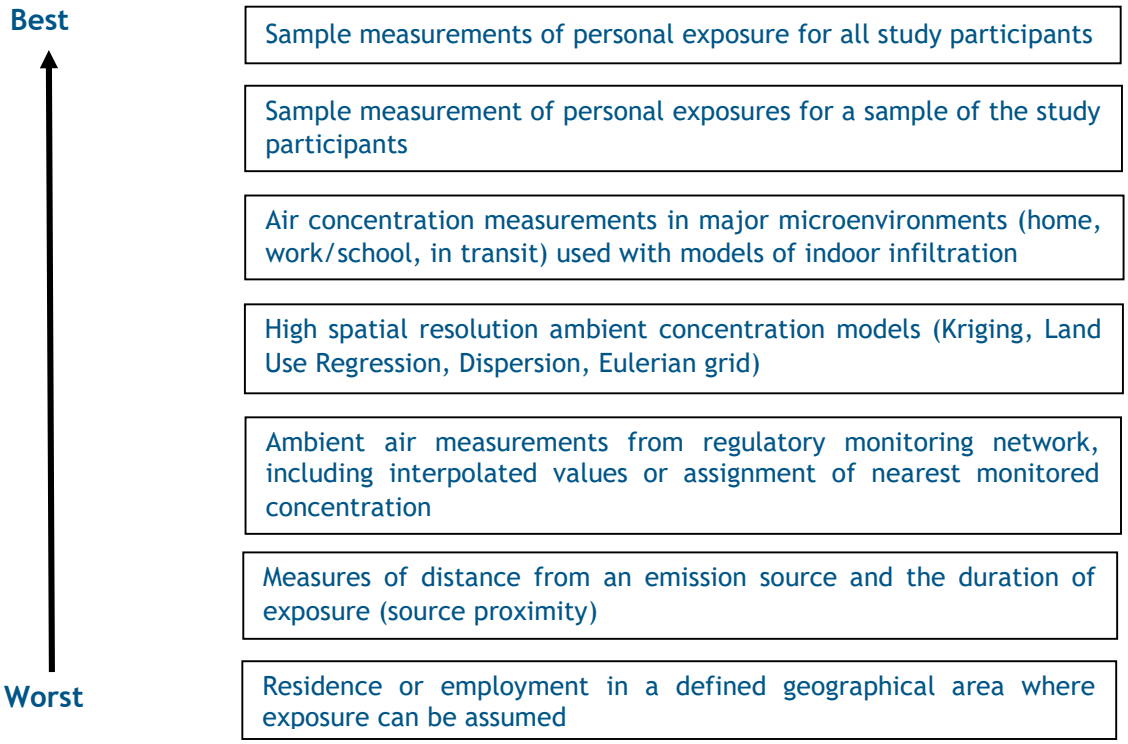
The worst exposure metric that has repeatedly been shown to yield highly biased results is self-reported estimates of exposure (Kuehni et al., 2006). A slight improvement can be attained by using proximity metrics, but these are also highly suspect and subject to considerable bias (Lipfert and Wyzga, 2008). The proximity exposure methods typically focus on the relationship between traffic-related emissions and the distance to an individual's residential location. The approach equates proximity to an emission source (i.e. a roadway) with exposure intensity. This categorical approach involves the creation of somewhat arbitrary distance and traffic density categories that can be assigned to the individual and the roadway, respectively.

Although these simple proximity models are still employed to some degree they have been largely supplanted by more advanced continuous measures of exposure that are capable of taking into consideration topographical characteristics of the local landscape and meteorological factors such as wind speed and direction. These so called interpolation models rely heavily on geostatistical techniques including geographic information systems (GIS), kriging (i.e. a type of geostatistical regression), and inverse distance weighting (IDW) to construct a surface map that interpolates air measurements from central monitoring sites or emission sources over an entire geographical region (Cesaroni et al., 2008).

The methods involve the assignment of precise spatial coordinates to an individual's residential location and at a grid of measurement sites or emission sources. These allow the estimation of an exposure concentration over the entire grid with some degree of accuracy. Interpolation methods are relatively easy to construct and have the advantage of providing estimated values and standard errors (i.e. uncertainty determinants) at all unmeasured locations including a subject's place of residence.

Although interpolation methods are an improvement over proximity models, they do not consider the impact that topography can have on localized spatial changes in pollutant concentration (de Mesnard, 2013). In other words, the terrain is assumed to be flat and the concentration is assumed to be relatively homogeneous and independent of local topography. Accurate use of these interpolation techniques requires a fairly dense network of monitoring stations, which generally only exists in an urban centre. Despite these problems, kriging provides exposure estimates that are vastly superior to those that simply assign an exposure measurement that is equivalent to the value from the nearest monitoring site (Son et al., 2010).

Figure 1 Hierarchical arrangement of exposure assessment methodologies used in environmental epidemiology studies (Brauer et al., 2008)



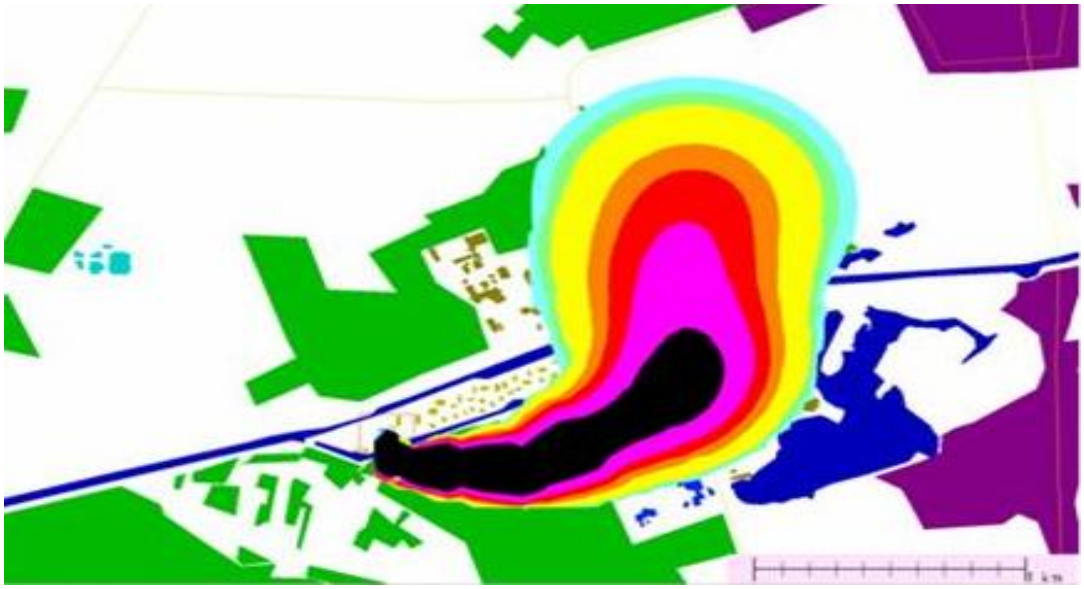
Land use regression (LUR) models circumvent the problems seen with interpolation methods by including land use characteristics, population density, and traffic patterns into the mapping scheme (Ryan and LeMasters, 2007). They are increasingly being used because of their ability to resolve the spatial distribution of a pollutant on very fine scale and their acknowledged superiority over models that solely rely on interpolation or proximity. In addition, LUR models often depend on tailored sampling campaigns to obtain additional exposure monitoring data beyond what is available from sparsely distributed central site monitoring stations. The sampling campaigns typically occur during several weekly periods and the sites are coded using geographic information systems maps (Hoek et al., 2008a). The list of potential predictor variables for LUR development is very large and includes altitude, meteorology, land coverage, topography, population density, road density, road type, commercial land uses, and distance to local pollution sources. Dozens of these models have been developed for different regions of Europe, the US, and Canada; however, the models vary in quality and the degree of validation that has been performed. Perhaps the biggest limitation of LUR models is, however, their lack of transferability and the need to generate a new model for each location and time period of interest (Vienneau et al., 2010).

Although LUR models are felt to perform well, it is essential that monitoring data is collected from a sufficient number of monitoring sites and that an adequate number of predictor variables are included in the model (Basagana et al., 2013, Wang et al., 2012). A minimum of 80 monitoring locations is necessary to achieve reasonable estimates that are neither biased nor highly variable (Basagana et al., 2012). The use of a small number of monitoring sites together with a large number of predictor variables has been shown to artificially increase the correlation coefficient between actual and predicted measurements. Some have noted that correlation coefficients can be inflated by 50% or more if an independent data set is not used for validation

(Johnson et al., 2010). Another drawback to the use of LUR models is their inability to account for temporal variability, which some have corrected for by constructing separate models for discrete time periods lasting up to several years (Cesaroni et al., 2012, Eeftens et al., 2012).

Unlike LUR exposure models, dispersion models rely on emission rates to determine the concentration at a remote location. They accomplish this by assuming that the emissions are distributed in the form of plume that is affected by topography and meteorological conditions (Jerrett et al., 2005). Figure 2 provides a diagrammatic representation of how a plume is dispersed once released from an emission source. Dispersion models are data intensive and have not been used a great deal in air pollution studies. They have an advantage over LUR models because they use emissions inventory information rather the results from monitoring stations to estimate exposure. Consequently, they can consider temporal changes in pollution levels in a better fashion than LUR models. Studies suggest that dispersion models are as accurate as proximity models, but their performance against LUR models is still somewhat in question with one study suggesting better performance, a second indicating a poorer result, and a third showing no difference (Beelen et al., 2010, Brauer et al., 2008, Cyrys et al., 2005, de Hoogh et al., 2014).

Figure 2 *Plume dispersion as represented in a pollutant dispersion model*



A final type of airshed model that is finding some use in exposure estimation is called an Eulerian grid or chemical transport model. These are 3-dimensional models that divide a regional air parcel into grids that are as small as $(1\ km)^3$ in size. The model includes an emission module, a detailed meteorological module, and a photochemistry module that considers the complex chemical interactions that can occur as an air parcel ages. The most commonly employed Eulerian grid model is called CMAQ (Community Multi-scale Air Quality) and is used by the regulatory community to examine chemical transport and fate in a compliance setting (Bravo et al., 2012). These models are very sophisticated to run and take great effort to populate with information. The biggest advantage with CMAQ is that it allows exposure predictions at remote locations, far removed from any monitoring sites

(Wu et al., 2011). On the other hand, CMAQ is not able to provide enough spatial resolution to examine exposures within a neighborhood (Marshall et al., 2008). Given the advantages and disadvantages associated with the alternative modelling techniques, the current trend in exposure estimation is to develop complex hybrid models that merge the output from several models to reduce uncertainty (Beevers et al., 2013). Alternatively, the output from individual models can be melded in a Bayesian statistical framework (Akita et al., 2014). Exposure modelling in environmental epidemiology will continue to improve with the generation of increasingly complex approaches to offset concerns of misclassification. It is imperative, therefore, that future studies focus on providing some measure of model uncertainty through validation and sensitivity analysis in order to understand the amount of modelling bias that still exists.

3.1.2. Validation of models

LUR models have been used increasingly for describing small-scale spatial variation in air pollution concentrations and estimating exposures for individual participants in a cohort study. However, the models may not predict exposure at the residence location as well as suggested by the explained variance (r^2) figures which are often reported. The ESCAPE project has looked extensively at the performance of the LUR estimates and Eeeftens et al. reported on the performance of LUR models developed for four measures of PM: the fractions of PM smaller than 2.5 µm ($PM_{2.5}$) and smaller than 10 µm (PM_{10}), the coarse fraction calculated as the difference between PM_{10} and $PM_{2.5}$ (PM_{coarse}), and a measure of the blackness of $PM_{2.5}$ ($PM_{2.5}$ absorbance). For $PM_{2.5}$, Eeftens et al. reported that the median r^2 was 71% with a range across 20 European study areas of 35–94% (Eeftens et al., 2012). The estimates were derived from models for individual study areas developed to explain spatial variation in annual average concentrations of $PM_{2.5}$ at only 20 measurement sites. However, the available predictor variables included a huge number of GIS-derived predictor variables (e.g., traffic intensity, population, and land-use) and regional background concentrations. When these authors used leave-one-out cross-validation (LOOCV) to assess model fit, the median LOOCV r^2 fell to 60% with a range of 21% to 78%. The authors noted that there was a considerable risk of over fitting when evaluating a large number of predictor variables to explain the concentrations at relatively few sites.

Other studies have shown that LUR models based on a limited number of training sites perform well in internal LOOCV, but do worse in external hold-out validation (HOV) against independent measurements set aside for model evaluation (Basagana et al., 2012, Basagana et al., 2013, Wang et al., 2013). Basagana et al. reported an in sample r^2 of 0.82 which fell to 0.02 when calculated out of sample for a LUR model developed for 20 measurement sites and variable selection from 100 predictor variables (Basagana et al., 2013). In a separate publication, Basagana et al. noted that it was very easy to find models with very high cross-validation r^2 even when no relationship existed; especially when there was a small number of measurement sites and a large number of potential predictors (Basagana et al., 2012). HOV is preferable as it better reflects the predictive power of the model at locations where no measurements were taken, such as addresses of subjects in an epidemiological study. However, Wang et al. noted that they were not aware of HOV r^2 being calculated for particulate matter LUR models because sampling of PM requires more effort and usually the number of sampling sites is not sufficient to allow for a separation into training and test data set (for validation purposes) of sufficient size. Wang et al. also evaluated LUR models for nitrogen dioxide (NO_2) developed using training sets of 20 sampling sites and test data sets of a similar size, and two PM components ($PM_{2.5}$ absorbance and copper in PM_{10}). These PM components only had sampling site data but were highly correlated with NO_2; so a

good indication of performance was given by calculating the correlation with NO_2 data from test sites. In the case of NO_2, the median model r^2 was 0.88 and the median LOOCV r^2 was 0.83, but the median HOV r^2 fell to 0.52. The training set correlations between LUR estimates for the two PM components and NO_2 levels suggested similar reductions in HOV r^2 for the PM components.

Although it has not been possible to calculate HOV r^2 for the $PM_{2.5}$ LUR estimates in the ESCAPE study areas, some indication of performance can be derived from another ESCAPE validation study by Wang et al. (Wang et al., 2014b). In this study, the investigators evaluated LUR models for NO_2, $PM_{2.5}$ and $PM_{2.5}$ absorbance by combining standardized measurement data from 17 PM and 23 NO_2 ESCAPE study areas across 14 European countries to derive a global European model rather than models for individual study areas. Models were evaluated with LOOCV and HOV, and the transferability of the models was investigated by successively excluding each study area from model building. In the case of $PM_{2.5}$, 356 PM sites were available for modelling from 17 study areas and HOV used 75% training and 25% test sites for the PM metrics. The European $PM_{2.5}$ model had a high r^2 of 0.86, but the regional background concentration explained a large fraction (71%) of variation in $PM_{2.5}$. Between-area differences for $PM_{2.5}$ were much higher than those for the traffic-related pollutants NO_2 and $PM_{2.5}$ absorbance, where the regional background concentration explained 8% and 28% of variation, respectively. The LOOCV and HOV r^2s were only slightly lower, 81% and 80% respectively. However, the investigators also reported other statistics which were more relevant for the ESCAPE study where cohorts were located within a city or small area, and cohort-specific epidemiological analyses were conducted. Wang et al (2014) reported within-area r^2s (the within-area variation explained by the European model and described as $Model_{intra}$ r^2) which are more comparable to the r^2 of individual study areas calculated by Eeftens et al (Eeftens et al., 2012). The median and IQR of the $Model_{intra}$ r^2s was 0.48 and 0.16, respectively. Wang et al also reported what they describe as transferability r^2s ($TRANS_{intra}$ r^2) (Wang et al., 2014b). These show the performance of models that used all monitoring data excluding one area at the time. Hence, 17 PM models were built until each of the study areas had been excluded once from model building and the transferred models were applied directly to the sites of the area that was left out. The median and IQR of the $TRANS_{intra}$ r^2s was 0.42 and 0.17, respectively. These provide a better indication of the performance of $PM_{2.5}$ LUR estimates at the ESCAPE study areas than the r^2 and LOOCV r^2 reported by Eeftens et al., and provide some indication of what the HOV r^2 would be if there had been sufficient PM data available to calculate them.

Other sources of error which are often overlooked include error in the exposure measurements at the sites used to develop the LUR models. For instance, Eeftens et al. noted that the results from the three measurements at each site were averaged to estimate the annual average, adjusting for temporal variation using a centrally located background reference site, which was operated for a whole year (Eeftens et al., 2012). Eeftens et al. also fitted separate models in the ESCAPE study for $PM_{2.5}$, PM_{10}, and PM_{coarse}, even though PM_{coarse} = PM_{10} - $PM_{2.5}$. In addition, the $PM_{2.5}$:PM_{10} ratio is likely to have been fairly constant over many study areas, but the LUR models for $PM_{2.5}$ and PM_{10} contained very different predictor variables in most study areas. In addition, $PM_{2.5} \leq PM_{10}$, but this may not the case for modelled values [see results for EPIC-Athens in Figure 2 of Raaschou-Nielson et al. (2013)]. Nevertheless, LUR models are better understood than some of the spatial interpolation techniques such as dispersion modelling (DM). De Hoogh et al. is one of few studies that have compared the performance of estimates of individual air pollution exposure derived using LUR and DM, techniques which are commonly used for in population studies (de Hoogh et al., 2014). De Hoogh et al. showed that for the ESCAPE study areas, LUR and DM estimates correlated well on average for NO_2,

but only moderately for PM_{10} and $PM_{2.5}$, with large variability across areas. The median (range) Pearson correlation coefficients between LUR and DM estimates for the annual average concentrations of NO_2, PM_{10} and $PM_{2.5}$ were 0.75 (0.19-0.89), 0.39 (0.23-0.66) and 0.29 (0.22-0.81) based on 13, 7 and 4 study areas, respectively. DM predicted a moderate to large proportion of the measured variation for NO_2, but less for PM_{10} and $PM_{2.5}$. Overall, there is good evidence that estimates of individual air pollution exposure derived using LUR and other methods may perform much worse than claimed by many study investigators.

3.2. EXPOSURE METRICS

Exposure estimation techniques are needed when personal measurements are either unavailable, prohibitively expensive to collect, or analytically impossible. Exposure estimation may either be qualitative (yes/no), semi-quantitative (high/medium/low), or quantitative (continuous variable) depending on the needs of the investigator, the methodological design, and the amount of funding available. Simple qualitative estimates of exposure are typically used in an occupational setting in conjunction with a screening level assessment of potential hazards in the workplace. This type of exposure estimate is typically obtained using a worker questionnaire, personal interview, or from work histories. Qualitative exposure estimates are by their very nature limited in their ability to assess the strength and robustness of an association with adverse health outcome. Semi-quantitative estimations on the other hand are capable of supplying more detailed information on the nature of the concentration-response function, but still lack clarity because of the arbitrary weighting that is present when classes are created to define each exposure group (Stewart and Herrick, 1991). This subjectivity reduces the utility of these measurements and prevents their use in a quantitative risk assessment. As such, the classification criteria used to define each exposure category may result in a large degree of misclassification and uncertainty because of the arbitrary cut-points used to distinguish one exposure class from another (Hornung, 1991). Quantitative exposure estimation techniques provide the best option for assessing the degree of exposure at both the individual and population levels. Within the air pollution arena, a wide variety of exposure models have been developed to aid in the exposure estimation at either the individual or population level.

Modern exposure estimation techniques in the air pollution arena are highly dependent on the creation and validation of air pollutant exposure models capable of supplying robust approximations of personal exposure (Zou et al., 2009). At their core, exposure models represent an indirect method of exposure assessment that may entail the use of any of a variety surrogate indicators as a proxy for the actual exposure of interest. Examples include the distance from a roadway, traffic density, or road density (Lipfert et al., 2006, Rose et al., 2009). These surrogate measures may, however, result in considerable misclassification in complex environments where the terrain or meteorology may affect the dispersion of pollutants from the emission source (Jerrett et al., 2005). Since all exposure estimation models suffer from some degree of bias, it is essential that the variability and uncertainty of the estimations are documented in some fashion using either a qualitative or quantitative technique. This includes an assessment of the spatial and temporal variability of the estimates together with an evaluation of the errors that may be affecting the linkage between the exposure source and the human receptor. Three types of uncertainty can affect the risk estimates arising from the use of exposure estimation models (Fryer et al., 2006). They include i) scenario uncertainty, ii) model uncertainty, and iii) parameter uncertainty. Scenario uncertainty can arise when spatial and temporal variations have not been adequately considered in an exposure model. Model uncertainty develops when there are inconsistencies between model representations of the events leading to an exposure and the

processes and events that actually take place. Parameter uncertainty can be easily investigated using Monte Carlo simulations; however, this quantitative approach is rarely applied in the air pollution arena. Since many of the exposure estimation models available for use in air pollution epidemiology studies have only seen a modest degree of validation via intra-model comparisons of the predictions, there continues to be some anxiety about the dependability of the exposure estimates derived from the use of air pollution models.

3.2.1. General principles

Accurate measures of personal exposure are vital for establishing relationships between environmental air pollutants and an acute or chronic disease state. These measurements are rarely if ever collected, however, because of the prohibitively high costs of monitoring hundreds or even thousands of individuals on a daily basis and the logistical constraints imposed by the use personal sampling devices (Zou et al., 2009). As such, environmental epidemiologists are often faced with using an indirect exposure surrogate that is representative of personal exposure. These surrogates fall into one of two categories depending upon whether the focus is on traffic-related emissions from local roadways or regional ambient air pollutants measured at a central air pollution monitoring station. In each case, residential location serves as a proxy for the individual since it is assumed that most people spend the majority of their time at home (Huang and Batterman, 2000). Although this is supported by time-activity studies showing that Americans and Europeans spend an average of 67% and 58% of their time indoors at home, the assumption fails to capture the other factors that may affect the exposure magnitude (Leech et al., 2002, Schweizer et al., 2007). In particular, place of residence surrogates do not consider the time spent outdoors, home penetration factors, indoor sources of pollutants, commuting in heavy traffic, or secondary occupational exposures (Baxter et al., 2013). These factors alone can result in substantial exposure misclassification independent of any modelling bias.

Residential locations can be used in conjunction with any of an array of exposure modelling tools to refine the estimates and provide some temporal and spatial resolution. Geocoding techniques are often used to precisely describe the map coordinates of a residence relative to an emission source or a monitoring site (Nuckols et al., 2004). As explained further in Section 3.2.2, geocoding of residential locations is subject to positional errors that need to be determined and conveyed in order to understand potential uncertainties. Unfortunately, the positional error associated with geocoding is rarely if ever conveyed in the exposure modelling. Most epidemiology studies focusing on traffic-related pollutants use proximity to the source as the most basic measure of exposure. This approach has largely supplanted self-reported estimates of traffic density, which is highly prone to bias (Kuehni et al., 2006). Alternatively, some studies will also include other traffic-related metrics to better describe the strength of the emission source and provide some additional insight on the nature of any relationships that are identified. Whereas, proximity models simply measure the centreline distance from a home to a roadway to evaluate the magnitude of an exposure, the inclusion of metrics such as traffic volume or traffic density help reduce the uncertainty and provide some refinement in the nature of the exposure. Dispersion modelling will also occasionally be used to relate actual traffic emission rates to pollutant levels at a particular residence. The value of dispersion models resides in their ability to factor wind speed and direction into a determination of pollutant levels at a receptor site that is located some distance from an emission source. Table 1 provides additional information on the strengths and limitations of the modelling approaches used in traffic-related epidemiology studies (Batterman et al., 2014).

Table 1 Strengths and limitations of metrics used to characterise roadway emissions

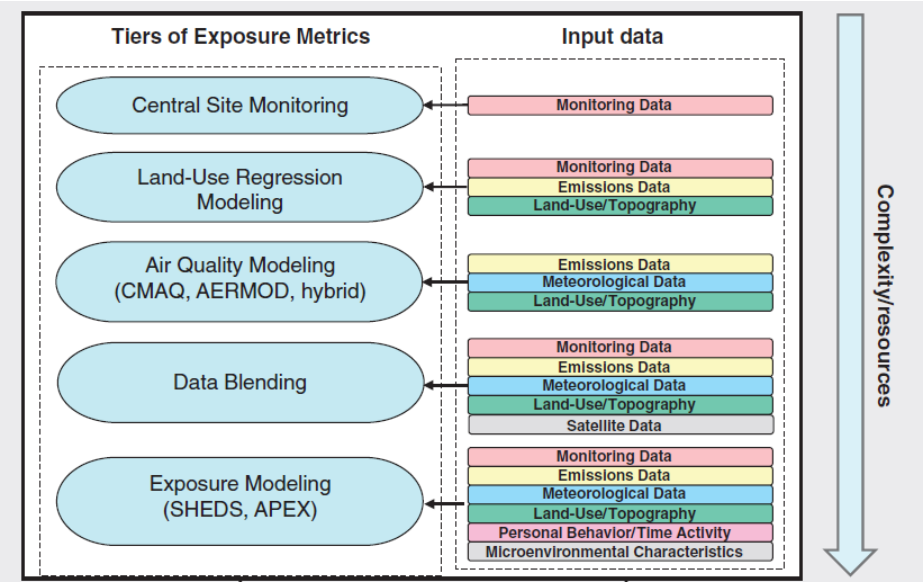
Metric	Strengths	Limitations
Distance to major road	• Simple to construct. • Low data needs. • Can potentially distinguish roads with varying traffic volume, vehicle mix, or other characteristics.	• Distance limit used as cutoffs for classifying homes/receptors is arbitrary. • May not consider traffic volume, vehicle mix, and other factors. • Sensitive to distance calculation, e.g., using road edge or centerline.
Total traffic volume on nearby roads	• Relatively simple to construct. • Reasonably good volume estimates on major roads. • Can select period of day, e.g., rush-hour.	• Traffic volume estimates needed. • Distance criterion used to determine road is arbitrary. • Does not provide metric for low traffic groups.
Diesel or gasoline traffic volume on nearby roads	• Relatively simple to construct. • May relate to PM emissions from diesel traffic. • Can select period of day.	• Difficult to estimate diesel traffic volume accurately. • Does not account for type of diesel vehicles and emissions. • Limitations described for total traffic also apply.
Local traffic density	• Includes local traffic emissions that might affect receptor. • Traffic density expressed a vehicle kilometer mile traveled per day, which is easily interpretable and can be generalized. • Large range across sites. • Can be applied to irregular shaped sources and receptors. • Can select period of day. • Relevant to traffic analysis zones used by planners.	• Moderately high data needs. • Computationally intensive. • Sensitive to distance criterion, which is assigned somewhat arbitrarily. • Uncertainty of traffic estimates on all but major roads. • Excludes smaller roads.
Emissions on local roads	• Incorporates vehicle emissions of pollutants of interest. • Reflects vehicle mix on roads. • Strengths described for local traffic density also apply.	• Results depend on pollutant, to an extent. • High data needs. • Computationally intensive. • Difficult to estimate emissions accurately.
Pollutant concentration predictions	• Incorporates effects of emissions, meteorology, and location in physically-based approach. • Quantifies and apportions concentrations due to each source, e.g., traffic. • Can be derived for specific periods of day, season or year, e.g., daily predictions at rush hour periods. • Inter-study comparisons are possible and meaningful.	• Results depend on pollutant, averaging time, and statistic. • High data needs. • Computationally intensive. • Uncertainty not well characterized. • Results potentially sensitive to many factors, including home placement.

Studies focusing on regional air pollutants often rely on the measurements taken at central site monitoring stations. These values are sometimes used directly without any spatial mapping of the results to a nearby residential location. This can result in an appreciable amount of measurement error, since topological, meteorological, and socioeconomic factors can modulate airborne concentrations and impact the levels observed at locations that are just a short distance from a central site. To reduce the error, investigators often rely on a suite of exposure models to obtain a better estimate of exposure at a particular address. Figure 3 lists many of the models that can be used in association with the values from monitoring sites (Ozkaynak et al., 2013). The models are listed according to the level of complexity and the types of input information needed to populate the model. In all but a single case, monitoring data provides the basis of exposure estimations. Land use regression models are complex representations of a residential area that take into consideration topography, traffic patterns and local geography to map out the exposure level that exists at any set of coordinates (Ryan and LeMasters, 2007). Air

shed and dispersion models are different in that they rely on emission rates to determine the concentration at a remote receptor location. Data blending models are hybrid models that use information from a variety of sources including satellite measurements and remote sensors in a Bayesian statistical framework that is capable of handling diverse types of information. The highest modelling tier includes several sophisticated approaches capable of considering an individual's behavioural pattern as well as home penetration rates, and local meteorology. The approach includes the US EPA's Stochastic Human Exposure and Simulation (SHEDS) and Air Pollutant Exposure (APEX) models, which are aimed at collected personal exposure information rather than exposures at a particular location.

Although these modelling methods are capable of reducing spatial and temporal misalignment in exposure estimations, much depends on the study design and data availability. Ecological time-series studies of acute conditions such as hospitalizations or emergency department visits are adversely affected by any temporal misalignment in the exposure assignments at a particular location and care must be taken to select an exposure model that has been thoroughly evaluated (Sheppard et al., 2012). Although methods exist to validate an exposure model and provide some measure of uncertainty, oftentimes this analysis if overlooked or treated in a very superficial manner (Baxter et al., 2010). It is also tempting to assume that the suite of exposure models available for spatial and temporal mapping yield similar exposure metrics that are directly comparable; but, in fact, the results can be quite different depending on the precision inherent in the model and the degree of spatial variability in the pollutant. For instance, it has been shown that the spatial variability of nitrogen dioxide is quite high due to all of the local emission sources, so the results obtained with a land use regression model can differ from those obtained using an urban air shed or dispersion model (Gulliver et al., 2011a). PM_{10}, on the other hand, is regionally distributed in a more uniform fashion so an exposure model does not require as much spatial resolution and alternative modelling approaches may yield similar results (Marshall et al., 2008). Although land use regression models are becoming more popular because of their fine spatial resolution, they have one big limitation; the model cannot be extrapolated across time or space and is limited to a specific location and time period (Vienneau et al., 2010).

Figure 3 Modelling tiers used in conjunction with measurements from central site monitors (Dionisio et al., 2013).



3.2.2. Historical reconstruction of exposures

Historical reconstruction of exposures is generally performed in conjunction with retrospective studies examining the association of an emission source or pollutant level with a chronic health effect such as lung cancer (Nuckols et al., 2004). Typically, these studies use residence locations as a proxy for personal exposure since it is assumed that most people spend the majority of their time at home. Although these techniques add value over what can be achieved when relating a health effect to a residential location at the time of enrolment or diagnosis, they still suffer from various forms of bias. The intent behind exposure reconstruction is ether the development of a cumulative exposure metric over the time span of the study or the temporal tracking of an individual's proximity to an emission source (Rushton, 2009). As such, temporal and spatial changes in exposure are taken into consideration. The reconstruction techniques used in environmental epidemiology generally differ from those used in occupational studies, which often rely on work histories, job titles, questionnaires, and job exposure matrices to develop a record of exposure (Sahmel et al., 2010). Modern retrospective environmental studies, on the other hand, often rely on the use of GIS (Geographic Information System) to spatially map an individual's residential location relative to nearby monitoring stations or emission sources. This approach has largely supplanted the use of questionnaires whereby participants were asked to provide estimates of their exposure in a qualitative (exposed or not exposed) or semi-quantitative (low, medium, or high) manner.

Regardless of the study goals, a key aspect of exposure reconstruction is the identification of a participant's residential history during the period of interest (Hughes and Pruitt, 2014). This information is typically obtained using a questionnaire or through personal interviews; however, this approach is labour intensive and subject to recall bias. Some individuals invariably fail to remember all of the residential locations, correct addresses, or lengths of residence at each location. Alternatively, the residential history can be obtained from commercial sources or government agencies such as a motor vehicle department, but this approach is often cost prohibitive (Jacquez et al., 2011). Another issue of concern

is the failure by most investigators to check the accuracy of their information or to cite the percentage of individuals who showed high residential mobility, which may directly impact the reliability of the residential record.

Following the creation of a residential history, the information is geocoded to obtain the coordinates for each residential location (Beyea and Hatch, 1999). Geocoding can take place by any of a variety of techniques, which all take advantage of GIS maps that are available for many locations (Zandbergen, 2008). An entire area of science has emerged in response to the widespread interest in geocoding. As such, epidemiologists often collaborate with experts in geospatial science who are able to create the geocoded residential locations that are needed for an exposure determination. Geocoding essentially involves the grid mapping of a residential address using the information available from a GIS (Healy and Gilliland, 2012). This mapping is, however, prone to error since the residential location is interpolated from those GIS coordinates (longitude and latitude) that are known with certainty. In some cases, this error can be quite high especially when aerial photography is relied upon for geocoding (Schootman et al., 2007). A review of geocoding studies found a mean positional error of 58-96 m in urban areas and 129-614 m in rural locations where there is a greater distance between homes (Jacquez, 2012). An investigation of the impact of address geocoding errors found an appreciable impact on exposure determinations. A median positional geocoding error of 41 meters resulted in unacceptably high overestimates of traffic-related exposures in a study of children residing near busy roadways (Zandbergen, 2007). A total of 391 children from the study were found to actually reside within 50 m of a high traffic corridor, but street geocoding indicated that 1413 children were within this zone. This positional error resulted in a large bias in the odds ratios for children at risk from exposure to traffic-related pollutants. Geocoding using the parcel centroids (centre point within the property boundary) instead of a property address have been shown to be far superior, yielding positional errors that are an order of magnitude smaller than those found when the street address was used. This approach, however, is not widely used in exposure reconstruction since it requires the collection of additional information on the property lines for each residential location (Cayo and Talbot, 2003). Likewise, although the use of a global positioning system (GPS) receiver for mapping the coordinates of a residential location provides an unambiguous approach to geocoding, it is rarely applied because of the extra labour required (Bonner et al., 2003). Since many retrospective studies relying on GIS for exposure reconstruction fail to include precise details regarding the geocoding methods employed or their positional accuracy, a high degree of exposure misclassification can be assumed to exist. The failure of exposure reconstruction methods to consider time away from the home, time spent indoors, and employment history increases the likelihood of an exposure misclassification beyond what the positional error can cause.

An often cited example of the practical value of historical exposure reconstruction was performed in conjunction with a retrospective case-control study of lung cancer in citizens from Stockholm, Sweden (Nyberg et al., 2000). Geocoding was performed on residential addresses that were identified via a questionnaire sent to patients or the next of kin (Bellander et al., 2001). An 85 % response rate was attained after mail reminders and phone calls. Gaps in knowledge were filled using information from local parishes and tax authorities. Geocoding was also performed for a large number of known traffic-related emission sources as well as point (industrial areas and power plants) and area (merchant vessels and airplanes) sources. Dispersion modelling was used to relate source emissions to individual time-weighted average exposures over a 30-year period. An automated procedure was used for geocoding a majority of the emission sources and residential addresses. The average positional accuracy of the method was 50 m and 90 % of the coordinates were within 100 m.

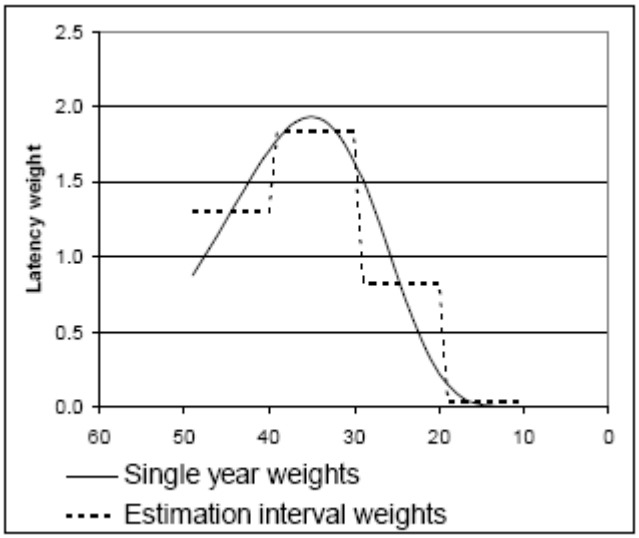
Although the authors correctly noted that the likelihood of exposure misclassification was less likely than with other population-based studies of lung cancer that relied on an ecological design, they failed to acknowledge the potentially large bias that continued to exist because of respondent recall errors and the positional inaccuracy. The authors also failed to conduct an important, but often omitted, sensitivity analysis to validate the exposure determinations. The omission of a sensitivity analysis that showed how positional errors affected the exposure estimates is seen as a serious limitation in this study (Jacquez, 2012).

There have been few in-depth evaluations of the impact of recall and geocoding errors on the exposure predictions resulting from reconstruction efforts. The results from a recent case-control study of lung cancer risks for Canadian citizens exposed to air pollutants suggest that exposure misclassification may be a serious problem with all reconstruction efforts (Hystad et al., 2012). Using proximity indicators as a surrogate for exposure magnitude, the study found that residential recall bias was especially high for residential histories more than 20 years old and that the bias was greater in control participants than in the cases. This caused a differential error in the exposure estimates that resulted in an overestimation of exposures in the cases relative to the controls. Another problem with this study was the geocoding of residential locations using postal code centroids rather than street addresses, which is known to be far less accurate (Goldberg et al., 1993). Although more studies are needed to gauge the real impact of geocoding errors on exposure and risk determinations, a recent study comparing the use of questionnaires versus GIS based residential proxies found that they yielded comparable results in a case-control study of lung cancer (Cordioli et al., 2014). These results suggest that the errors associated with geocoding are as large as those found when using self-reported estimates of exposure and that that the only advantage in using address geocoding is that the results are more objective.

3.2.3. Biologically relevant exposure period

Studies on longer-term exposure to air pollution and cancer risk generally use exposure estimates for some convenient time point or window, but there is often incomplete coverage (or no coverage) of the biologically relevant exposure period (BREP). For instance, many prospective studies estimate exposure at the recruitment/baseline address and in some cases exposure at that address at the end of follow-up. Few studies have acknowledged that the relevant time window for exposure in the case of lung cancer probably occurred before recruitment, especially if the follow up period is short. Proximate air pollution estimates often do not represent the BREP, and most studies ignore the exposures occurring before the follow-up begins. In the case of particulate matter, historic exposure concentrations may be much higher than the concentration at recruitment. For studies on lung cancer, exposures during the most recent 10 years are unlikely to be relevant given the latency for that cancer, and even exposure 10-20 years previously will be relatively unimportant. For example, Hutchings and Rushton used the latency weights shown in Figure 4 when predicting future deaths from lung cancer due to silica exposure. In the case of lung cancer, it is likely the BREP is anywhere from 15 to 50 years before death (Hutchings and Rushton, 2011).

Figure 4 Lognormal distribution latency weights (single year and estimation interval) used by Hutchings and Rushton.



The degree of exposure misclassification depends on a number of factors including the design of the study and the exposure metric used. The most influential prospective studies include the H6C study, the ACS CPS-II study and the ESCAPE study. The exposure estimates for the H6C and ACS CPS-II studies were ecological, although the number of metropolitan areas with estimated exposures and the number of subjects differ greatly between the studies. The H6C study included six areas and approximately 8000 subjects and the ACS CPS-II study included 126 areas and approximately 540,000 subjects with a $PM_{2.5}$ exposure that lasted for at least part of the study period (Pope et al, 2002). In contrast, the ESCAPE study of lung cancer included 17 cohorts from nine European countries (312,944 subjects), but the study pooled hazard ratios from within the cohort analyses (Raaschou-Nielsen et al., 2013). Subjects from 14 cohorts had individual estimates of proximate $PM_{2.5}$ exposures calculated using land use regression (LUR) models.

All of the above studies have used Cox regression models, but the exposures have been treated differently both within and between studies. Some studies have treated exposure as a time dependent variable reflecting mean exposures within a time window before death. For instance, Laden et al. fit their models for mean exposure in the year of death, whereas Lepeule et al. fit their models for mean exposure in the 3 years before death (Laden et al., 2006b, Lepeule et al., 2012). Krewski et al. optimized their models for a 15 year time window, but also used estimates based on 1-5 year, 6-10 year, and 11-15 year time windows (Krewski et al., 2009). For many pollutants, the difference between exposures during such a time window and exposures in the BREP may be very large. Even if spatial distributions of pollutants remain constant (see below), the levels of some pollutants like $PM_{2.5}$ have fallen considerably in recent years in many areas. Consequently, the risk per unit of $PM_{2.5}$ will be overestimated. The problem is well illustrated by the latest update of the H6C study (Lepeule et al., 2012). In this study, $PM_{2.5}$ was included in Cox regression models as an annual time-dependent variable and a 1- to 3-year moving average was used for lung-cancer mortality (i.e. a 3-year time window). The latest update studied mortality between 1982 and 2008 and the investigators estimated annual exposure for the six cities between 1974 and 2009. The earliest air pollution monitoring results were available from 1979 for five

cities and 1980 for the other city, and these estimates were assumed to be applicable to the years before monitoring began. Figure 1 of Lepeule et al. shows estimated annual mean $PM_{2.5}$ levels during 1974–2009 for the six cities in the study. It is noticeable that although there was a huge range in 1979/1980 levels in the six cities when monitoring began, there was virtually no difference between levels in the six cities in 2009. The study reported a strong association between lung cancer and exposure to $PM_{2.5}$ at the baseline address in the 3-year window before death. For lung cancer, a RR of 1.37 (95% CI 1.07-1.75 per 10 μg/m^3 increase in $PM_{2.5}$) was reported. However, in 2009, the 1- to 3-year moving average $PM_{2.5}$ levels for the six cities ranged from about 10 to 14 μg/m^3, whereas mean exposure levels for the six cities between 1974 and 2009 ranged from 11.4 to 23.6 μg/m^3. Mean $PM_{2.5}$ levels during the BREP for a 2009 lung cancer death (1959-1994) would have had a much wider range, and a range of at least 30 μg/m^3 is plausible given that $PM_{2.5}$ levels at Steubenville were approximately 40 μg/m^3 and falling sharply at the start of the monitoring period in 1979. Consequently, the RR estimated using 1- to 3-year moving average $PM_{2.5}$ levels will greatly overestimate the true risk per unit of $PM_{2.5}$.

Lepeule et al. reported RRs for four periods of follow up (obtained by dividing the follow-up period into four equally spaced time periods) (Lepeule et al., 2012). Only the RR for the last period of follow up (2001-2009) was significantly different from unity (RR 2.84; 95% CI 1.06–7.59: 10 μg/m^3 increase in $PM_{2.5}$). Except for the city of Portage, Figure 1 of Lepeule et al. indicates that mean $PM_{2.5}$ level at the baseline address during the BREP for a subject dying during 2001-2009 will be higher than the 1- to 3-year moving average used in the analysis. The range of 1- to 3-year moving average $PM_{2.5}$ levels fell from about 8 μg/m^3 in 2001 to approximately 3 μg/m^3 in 2009 with a mid-period value of approximately 5 μg/m^3. Hence, the range of 1- to 3-year moving average $PM_{2.5}$ levels in 2005 is likely to be at least six times less than the range of mean exposures during the BREP. Although the RR for the period 2001-2009 appears large, it would fall to 1.19 if exposure estimates were scaled upwards by a factor of six. In addition, air pollution levels at the six cities were not very well separated throughout the period from 2001 to 2009, and had virtually converged by 2009. Consequently, the air pollution estimates for some of the six cities during the period 2001-9 had a different ordering to that based on their mean $PM_{2.5}$ levels during the BREP. For instance, Portage almost certainly had the lowest mean $PM_{2.5}$ level of the six cities during the BREP, but Watertown and Topeka had lower 1- to 3-year moving average $PM_{2.5}$ levels from 2001-2009. It cannot be concluded that the RR for 2001-2009 would have been statistically elevated or even greater than unity if a measure of exposure during the BREP had been used in the analysis. However, the RR for this period of follow-up has a strong influence on the RR for the entire follow up period because of increasing rate of lung cancer deaths with follow-up.

Other studies have not treated exposure as a time-dependent variable, although exposure during the BREP will vary over time. For instance, Turner et al. studied lung cancer mortality from 1982 to 2008 in never smokers included in the ACS CPS-II study (Turner et al., 2011). Three ecological measures of particulate air pollution were used for subjects living in each Metropolitan Statistical Area (MSA): mean $PM_{2.5}$ levels between 1979 and 1983 (available for 61 MSA); mean $PM_{2.5}$ levels between 1999 and 2000 (available for 117 MSA); and the average of these measures (available for 53 MSA). It is clear from Figures 17 and 18 of Krewski et al. that the annual mean $PM_{2.5}$ levels of many MSAs fell between 1972 and 2000, and Turner et al. reported mean $PM_{2.5}$ levels of 21.1 μg/m^3 between 1979 and 1983, 14.0 μg/m^3 between 1999 and 2000, and an averaged level over the two periods of 17.6 μg/m^3 (Krewski et al., 2009, Turner et al., 2011). The fully adjusted HRs for lung cancer mortality in relation to each 10 μg/m^3 increase in mean $PM_{2.5}$ concentrations were 1.15 (95% CI 0.99-1.35), 1.27 (95% CI 1.03-1.56) and 1.19 (95% CI 0.97-1.47) for the three

exposure measures, respectively. The HRs from the three analyses are more consistent than they at first appear. If the HR based on 1999-2000 exposure and the HR based on the average exposure measure are adjusted to the mean exposure concentration for 1979-1983 period, then the HRs are reduced to 1.17 and 1.16, respectively; close to the HR of 1.15 based on 1979-1983 exposure. Mean exposure during 1979 and 1983 will be closer in magnitude to exposure during the BREP than exposure between 1999 and 2000 (which is not in the BREP for any subject). However, we do not know how well the two measures correlate at different time points during the study period and hence cannot conclude that results obtained using a BREP based exposure measure would be similar to those obtained using 1979-1983 exposure measure.

In the case of the ESCAPE study, the exposure measure was individual level modelled exposure during 2008-2011 at the address of the subject at recruitment (mainly in the 1990s) (Raaschou-Nielsen et al., 2013). The mean follow-up period was 12.8 years, so the BREP of most subjects occurred before recruitment. In this study, analyses were conducted within centres and a pooled estimate of effect obtained. The authors claimed that the results of recent research in Rome, the Netherlands, and Vancouver showed that the spatial distribution of air pollution is stable over 10-year periods, but only cited a study of spatial contrasts in NO_2 in Rome which was reportedly stabile over a 12 year period (Cesaroni et al., 2012). Cesaroni et al. cited studies from the Netherlands and Oslo which reported stabile spatial contrasts in NO_2 and nitrogen oxides (NO_x) over shorter periods of eight years and three years, respectively (Eeftens et al., 2011, Madsen et al., 2011). Raaschou-Nielsen et al. also cited a study which showed high correlations between traffic intensities in 1986 and 1996 on Dutch streets. Beelen et al. and Gulliver et al. independently reported that spatial models for black smoke provided reasonable predictions, even going back to the 1960s (Beelen et al., 2007, Gulliver et al., 2011b, Raaschou-Nielsen et al., 2013). However, the $PM_{2.5}$ exposure trends of the H6C and ACS CPS-II studies suggested that the spatial distribution of $PM_{2.5}$ levels in 2008-2011 might be considerably different to the spatial distribution during the BREP. Information on time trends in exposure levels at the different centres is not provided, but it seems likely exposure levels in 2008-2011 would be closest to those during the BREP for the centres which had the lowest exposure in 2008-2011. Subjects from the Scandinavian and UK centres had the lowest estimated exposure to $PM_{2.5}$, and the meta-HR for these seven centres was 0.94 (95 % CI 0.64-1.38, p = 0.74) per 5 $\mu g/m^3$ $PM_{2.5}$ (i.e. no evidence of an exposure effect). By contrast, the meta-HR for the seven centres with the highest estimated exposure in 2008-2011 was 1.30 (95 % CI 1.01-1.67, p = 0.04) per 5 $\mu g/m^3$ $PM_{2.5}$. The biggest changes in $PM_{2.5}$ levels have probably occurred for these centres, especially the Italian and Greek centres, and the HRs for these centres may have been much lower if exposure during the BREP had been used in Cox regressions instead of exposure during 2008-2011. Morfeld et al. also noted the heterogeneity in these results and reported similar findings when the centres were split according to whether they were north of the Alps or elsewhere (Morfeld et al., 2013). The northern centres included the seven centres with the lowest exposure and the two Netherlands centres.

Some studies have focussed on changes of address during the follow up period e.g. Raaschou-Nielsen et al. and Cesaroni et al., although their follow-up periods were short (i.e. an average of 12.8 years in the former study, nine years for the latter) (Cesaroni et al., 2013, Raaschou-Nielsen et al., 2013). The investigators hypothesised that they should expect to see stronger associations among subjects who did not change address throughout follow-up as this minimised misclassification of long-term exposure relevant to the development of lung cancer. However, a conventional lagged analysis for a cancer with a long latent period such as lung cancer would disregard most if not all of the exposure during the follow up period

of many studies, and there seems no good reason to expect to see a stronger relationship with lung cancer for subjects who didn't move during a short follow up period. Indeed, if relocation is based on preference for “cleaner air”, then stronger effects might be seen among subjects who change address during the follow up period. In a sensitivity analysis conducted within the ESCAPE study, Oudin et al. showed good concordance between the average exposure at all addresses during the follow-up period (mean = 12.7 years) and exposure at the baseline address for a Swedish population (Oudin et al., 2012). However, temporal changes in exposure were ignored, and spatial variation in exposure levels over the study area was low. Consequently, this finding isn’t very relevant for lung cancer incidence/mortality studies.

Changes in address and exposure before the start of follow-up are likely to be much more relevant than those during the follow-up period for studies of lung cancer incidence/mortality. This is certainly true for studies with very short follow-up periods, (e.g. the large New Zealand cohort mortality study with a follow-up period of only three years) (Hales et al., 2012). With such a short follow-up period, average or cumulative exposure during the BREP is much more likely to differ from exposure at the baseline address. However, few studies have collected residential histories for their subjects which would allow exposure during the BREP to be estimated. One of these studies, the CNECSS case-control study by Hystad et al., collected 20 year residential histories for cases and controls to estimate annual residential exposure to various pollutants over a 20-year exposure period (Hystad et al., 2013). The investigators noted that the long latency periods associated with lung cancer make epidemiological analyses particularly challenging, especially for air pollution where spatial and temporal variation in both residential mobility and air pollution concentrations may produce significant exposure misclassification if not properly incorporated into the exposure assessment approach (Hystad et al., 2012). However, it may not be straightforward to estimate exposure during the BREP. In a separate article, these same authors noted that residential histories were available for an earlier exposure period for many subjects, but few air pollution measurements and no geographic data were available for these years (Hystad et al., 2013). Hystad et al. reported substantial exposure misclassification if the study entry address was used instead of residential history (Hystad et al., 2012). When examining exposure misclassification based on incorrectly assigned exposure quintiles, 50%, 49% and 46% of individuals where classified into a different $PM_{2.5}$, NO_2 and ozone quintile, respectively. However, the correlations between ambient air pollution exposures derived from study entry residential addresses only, and exposures derived from residential histories ranged from 0.70 to 0.76. Other studies which have attempted to estimate long term exposure include Danish studies of the association between NO_x (as a proxy for particulate matter exposure from traffic emissions) and various cancer endpoints including lung cancer by Raaschou-Nielson et al. and a Swedish study of the association between lung cancer and NO_2 and SO_2 (Bellander et al., 2001, Nyberg et al., 2000, Raaschou-Nielsen et al., 2010, Raaschou-Nielsen et al., 2011a, Raaschou-Nielsen et al., 2011b).

3.2.4. Temporality and time-dependent exposure metrics

Temporality (i.e. exposures occurring before the health effect) is an important consideration when assessing causality. In most studies of air pollution and lung cancer, it is likely that some exposure will have occurred before lung cancer was diagnosed, although exposure may be estimated for the baseline address after death. Ostro et al. noted that most previous cohort studies had assigned the same exposure period to all study subjects, regardless of when deaths occurred, and consequently the estimated exposures for some study participants in several studies occurred after their deaths (Ostro et al., 2010). For some studies e.g. the ESCAPE

study, the estimated exposures for all study participants occurred after their deaths, although as noted in section 3.2.3, this may not be a problem if assumptions about a stable spatial distribution for air pollution are valid (Raaschou-Nielsen et al., 2013). In some cases, the estimated exposure may have occurred 16 to 19 years after the subject developed lung cancer. In the ACS CPS-II study, the earliest lung cancer death may have occurred in 1982 when follow-up started, but many analyses are based on average exposure concentrations during 1999-2000. Ostro et al. also noted that exposures have usually been assigned to participants based on their residential address at enrolment only, without taking into account exposure changes that may have occurred throughout the study period or when participants relocated, and often exposure was measured for only a subset of the years during which the cohort was followed.

The analysis of the CTS study by Ostro et al. provides a good example of what can go wrong if the temporal nature of exposure is neglected (Ostro et al., 2010). In an attempt to reduce exposure misclassification, these authors estimated exposures beginning prior to the cohort follow-up period, continuing to the end of the study or until the participant died or relocated out of state, and incorporated updated exposure assignments when the subjects moved. Each subject was assigned a monthly exposure value based on the monitor nearest their geocoded residential address and the values for all person-months of exposure were summed and divided by the total months of exposure to create an average measure of overall long-term exposure. However, in the Cox regression analysis, exposure was not initially treated as a time-dependent variable and this resulted in apparently highly significant associations between $PM_{2.5}$ exposure and the four mortality endpoints. This was because measured concentrations of several pollutants in California declined substantially during the 2002-2007 follow-up period including annual average $PM_{2.5}$ concentrations which decreased by around 30%. This decrease in ambient $PM_{2.5}$ concentrations resulted in lower average exposure estimates for cohort members who survived to the end of the study. This meant that the exposure assigned to a participant who died at time t in the Cox regression analysis tended to be greater for events occurring early in the observation period compared with the long-term average exposures of the participants who comprised the remainder of the risk set (i.e., those who were still part of the cohort study at time t and who subsequently experienced lower ambient pollution levels). The CTS study was reanalysed using time-dependent pollution metrics which meant that the exposure estimates for everyone remaining alive in the risk set were recalculated at the time of each death in order to compare their average exposures up to that time with that of the individual who had died. In this way, decedents and survivors comprising the risk set had similar periods of pollution exposure, without subsequent pollution trends influencing the surviving women's exposure estimates. In the time-dependent exposure analysis, the HRs for $PM_{2.5}$ were much reduced. For example, the HR for all-cause mortality for a 10 $\mu g/m^3$ change in $PM_{2.5}$ was initially 1.84 (95% CI 1.66-2.05), but fell to 1.06 (95% CI 0.96-1.16) when the time-dependent exposure metric was used.

It isn't suggested that other studies have made the same error that Ostro et al. initially made (Ostro et al., 2010). For example, in another analysis of the CTS study Lipsett et al. focussed on lung cancer using the same time dependent exposure metric as Ostro et al (Lipsett et al., 2011). Nevertheless, the results of some studies involve temporal assumptions that have not been validated, including the assumptions made by studies such as ESCAPE and ACS CPS-II studies that results based on exposure distributions occurring potentially many years after the death of some subjects will be valid. The incorrect initial analysis of Ostro et al. is also of particular interest in that it demonstrates that error misclassification can result in much stronger associations and not the attenuation that investigators claim.

3.2.5. Personal concentration level versus exposure at the gatepost

The major epidemiological studies of air pollution and cancer risk have not been able to estimate the personal exposure of subjects. Instead, an estimate of outdoor pollution at the residence of a subject is often used as a proxy for personal exposure, or another indicator such as proximity to major roadways. Even if measurements of outdoor pollution at the residence of a subject were available, there would still be exposure misclassification because personal exposure depends on factors such as time-activity patterns, infiltration of pollution indoors, and indoor sources of pollution. Little is known about how well land use regression (LUR) models predict average personal exposure, but a study by Montagne et al. showed that within three cities (Barcelona, Utrecht and Helsinki) there was virtually no correlation between personal concentration measurements of $PM_{2.5}$ and LUR-modelled estimates (see Table 2) (Hoek et al., 2008b, Montagne et al., 2013). Indeed, the slope of the relationship was negative in two cities. The results were similar for within-city correlations between personal exposure measurements of $PM_{2.5}$ and outdoor measurements taken at the home address, although the correlation for Helsinki was barely statistically significant. The range in measured $PM_{2.5}$ concentrations for each city was small, especially in Helsinki, but between-city variation was much larger. Consequently, the correlation between personal exposure measurements of $PM_{2.5}$ and LUR-modelled estimates was stronger when data were pooled across the cities. When indicator variables for city were added to the model, $PM_{2.5}$ personal exposure did not remain significantly associated with the modelled concentrations (i.e. the model did not add further explanatory power than a single city background value from the between-city analyses). LUR-modelled estimates of NO_x concentrations were also not a good proxy for personal exposure. Additionally, the NO_x model did not add further explanatory power than a single city background value from between-city analyses. LUR-modelled soot and NO_2 estimates were a better proxy for personal exposure in within-city analyses and gave additional explanatory power in pooled analyses than a single city background value.

Table 2 Relationship between measured average outdoor/personal concentrations and LUR-modelled outdoor concentrations for individual cities and pooled data (from Montagne et al, 2013).

City	N	Personal vs. LUR-modelled outdoor		Personal vs. Measured outdoor		Measured outdoor vs. LUR-modelled outdoor	
		r^2	β	r^2	β	r^2	β
Utrecht	15	0.06	-0.37	0.09	-0.34	0.43	1.09*
Helsinki	15	0.08	1.08	0.32	0.23*	0.21	0.72
Barcelona	15	0.00	-0.15	0.10	-0.12	0.10	0.37
Pooled	45	0.35	1.01**	-	-	0.81	0.97**

*p<0.05, ** p<0.01

Montagne et al. concluded that, over larger ranges of outdoor concentrations, modelled or measured outdoor concentrations are good proxies for personal exposures of soot, NO_2, NO_x, and $PM_{2.5}$ (Montagne et al., 2013). The pooled sample of LUR-modelled $PM_{2.5}$ estimates from Utrecht, Helsinki and Barcelona had a range of 7.7 to 25.1 μg/m^3 (mean = 14.6 μg/m^3). This range is wider than the range of such estimates from all of the ESCAPE study centres, apart from EPIC-Athens (Raaschou-Nielsen et al., 2013). In addition, the pooled sample of Montagne et al. only includes 45 subjects, whereas the study groups of the ESCAPE centres range from 2,384 to 108,018. For a normally distributed variable, a sample of 45

observations would have a range which is on average 36% narrower than a sample of 2384 observations and 52% lower than a sample of 108,018. On that basis, the dispersion of LUR-modelled $PM_{2.5}$ estimates from all the ESCAPE centres is much less than the range reported by Montagne et al. and it cannot be assumed that the estimates will be a good proxy for personal exposure. Furthermore, 8 of the 14 ESCAPE study centres (SNAC-K, SALT, Sixty, SDPP, DCH, EPIC-MORGEN, EPIC-PROSPECT, VHM&PP) had ranges with widths between 5 and 10 $\mu g/m^3$, even narrower than the range reported for Barcelona (11.5 $\mu g/m^3$) by Montagne et al., and less than the range for Utrecht if adjusted for sample size. The study by Montagne et al. suggests that there is no reason to expect an association between lung cancer incidence and LUR-modelled $PM_{2.5}$ estimates in these eight centres with such low spatial variation, but the meta-HR for these eight centres is 1.22 (95% CI 0.93-1.55) per 5 $\mu g/m^3$ $PM_{2.5}$, similar to the meta-HR for all 14 centres of 1.18 (95% CI 0.96-1.46) per 5 $\mu g/m^3$ $PM_{2.5}$.

There will be a differential impact on studies that have a large between-subject variation in exposure (e.g. H6C and ACS CPS-II) than for those with low between-subject variation in exposure (e.g. a single city in Northern Europe). Studies such as H6C and ACS CPS-II rely on between-centre variation in pollution rather than within-centre variation. Nevertheless, the range of exposure estimates can still be narrow in some analyses. For example, the latest update of the H6C study used 1- to 3-year moving average $PM_{2.5}$ results for the six cities in a Cox regression analyses (Lepeule et al., 2012). There was a huge range in exposure levels between levels in the six cities when monitoring began in 1979/1980, but there was virtually no difference between levels in the six cities in 2009. One of the analyses was performed for the period 2001-2009 and 1- to 3-year moving average $PM_{2.5}$ results for the six cities had a range of about 8 $\mu g/m^3$ in 2001, which fell to a range of approximately 3 $\mu g/m^3$ in 2009. However, the range was considerably wider at the start of follow-up in 1982. Turner et al. also reported a wide range in 1979-1983 averaged $PM_{2.5}$ levels for 61 MSA included in one set of analyses (10.3-37.8 $\mu g/m^3$), but the range of 1999-2000 averaged $PM_{2.5}$ levels for 117 MSA included in other analyses was narrower (5.8–22.2 $\mu g/m^3$) (Turner et al., 2011). It is more reasonable to assume that this measure is a reasonable proxy for personal exposure, but this ignores other factors such as the BREP and residential history of a subject.

3.2.6. Back extrapolation

Back extrapolation is a historical exposure reconstruction technique that focuses less on residential geocoding and more on trend analysis and pollutant collinearities to estimate exposure levels for periods of time when sufficient monitoring data was either unavailable or of poor quality. As such, it can be used to reconstruct the exposure profile in a retrospective examination of cancer risk for a cohort of individuals living in an area of particular interest. The approach often relies on the use of statistical imputation techniques that relate the pollutant of interest to an emission factor or an emission source that has been more accurately tracked both spatially and temporally. For instance, changes in traffic density on local roadways can be used to estimate particulate exposures after first establishing the mathematical relationship that exists between the pollutant of interest and the proxy. The resulting equations may then be used to estimate exposures at other time periods when the traffic patterns were known with some certainty, but the pollutant of interest has not been monitored (Bellander et al., 2001). Alternatively, a co-pollutant may be used as a surrogate if information is available showing that the two vary in a regular and predictable fashion (Beyea et al., 2008). This later approach was employed to estimate historical $PM_{2.5}$ exposures in the TPCS study from Japan (Katanoda et al., 2011). Although monitoring levels were not available for $PM_{2.5}$, suspended particulate matter measurements had been taken at various

locations since the early 1970's. Since the ratio of $PM_{2.5}$ to suspended particulate matter (SPM) was shown to range from 0.6 to 0.8 at several locations and various time periods, a factor of 0.7 was uniformly applied to the SPM results to obtain measurements of $PM_{2.5}$ over a twenty-year span. The main problem with this approach, however, is the uncertainties that result when using a single adjustment factor to describe the $PM_{2.5}$/SPM relationship at all locations and all time periods of interest. A somewhat similar, but more precise, approach was used to estimate $PM_{2.5}$ exposures in a Canadian lung cancer cohort study (Hystad et al., 2012). In this case, a random effects model was constructed that related $PM_{2.5}$ concentrations to levels of total suspended particulates (TSP) in each of the cities of interest. The model was constructed using $PM_{2.5}$ and TSP measurements available from 1984 to 2000. The modelling results were then used with available TSP measurements to back extrapolate $PM_{2.5}$ for a twenty-year period lasting from 1975 to 1994. An advantage of this approach is the ability to take regional differences in the $PM_{2.5}$/TSP relationship into consideration. Still the veracity of the back extrapolation is difficult to establish with any degree of certainty. Oftentimes considerable uncertainty and misclassification may occur if sufficient care is not taken to ensure that the approach is well vetted and standardized to the extent possible (Sahmel et al., 2010).

Although a variety of approaches can be used to perform a back extrapolation, the method of choice is more dependent upon the goals of the investigation rather than the availability of relevant datasets that can be used to establish how a particular pollutant concentration has changed over time. Although back extrapolations can be performed for each individual in a study or at the population level, their use in the air pollution arena has, up until recently, been a relatively rare occurrence given the abundance of historical measurements from urban monitoring stations. Exposure back extrapolation are, however, much more common in retrospective studies in the workplace where historical exposure databases can be used to supplement any available personal monitoring information (Hornung et al., 1996). In this circumstance, great care needs to be taken to ensure that the available historical information is truly representative of the personal exposures that have taken place in past years. Typical pitfalls include the use of historical measurements that were collected as part of a sampling campaign to ensure compliance with an occupational exposure limit. The measurements may not be reflective of exposures throughout the workplace. Other dangers include the failure to correct for any personal protective devices being worn such as a respirator, and the absence of any adjustment for changes in the sampling procedure or analytical method. These factors can all have a dramatic impact on the reliability of the back extrapolation. When the uncertainty is known to be high, an exposure assessor needs to consider using a semi-quantitative approach to document historical exposures and obtaining a relative estimate of exposure magnitude. Under these circumstances, job-exposure matrices, expert judgment, and surrogate exposure techniques may be applied (Dosemeci et al., 1990). Surrogate exposure measurements, such as the length of time employed, the department where employed, or the type of machinery being used, introduce their own set of concerns and may not provide a dramatic improvement in reliability (Stewart and Herrick, 1991). Regardless of the techniques used to back extrapolate past exposures, it is imperative that variability and uncertainty are documented in some manner using an accepted approach (Tielemans et al., 2002).

3.3. IMPACT OF ERRORS IN MEASUREMENT

In most air pollution studies, estimates of exposure are based on available measurement data and models which predict exposure at a location during a particular time window. The models may involve spatial interpolation models based

on data from monitoring networks, but interpolation models also include dispersion models which address the inputs of the system (e.g., information about point sources of pollutants or environmental parameters). Alternatively, land use regression models which incorporate GIS-predictors of pollution concentration such as highways/traffic and industrial use of land are used to estimate exposure. Whatever approach is used, it is clear that the true measure of concentration cannot be accurately measured and the exposure estimates are subject to complex measurement error. Sheppard et al. discuss exposure measurement error in air pollution studies and note that inference about health effects can be incorrect when the measured or predicted exposure used in the analysis is different from the underlying true exposure (Sheppard et al., 2012). A full discussion of the consequences of measurement error is beyond the scope of this report, but it is a well-recognised problem in epidemiological research and Buzas et al. provide a comprehensive description of the problems that epidemiologists face when trying to infer disease patterns using noisy or indirect measurements of risk factors or covariates (Buzas et al., 2014).

One important consideration is the impact of measurement error on exposure-response estimates. This is difficult to assess because the exposure estimates are subject to complex errors that are not fully understood. However, when exposure estimates are included as explanatory variables in a regression model for a health outcome, it is likely that the variability of the estimated regression coefficients results will be underestimated, and the estimated coefficients will suffer from bias. Szpiro et al. developed a general framework for measurement error in spatial prediction, and demonstrated that for a fairly general class of exposure models there are two components to the measurement error: a Berkson-like component and a classical-like component (Szpiro et al., 2011a). The Berkson-like component of error results from smoothing the exposure surface using a model that may not account for all sources of variation and can be thought of as the part of the true exposure that is not predictable from the model. This component behaves like Berkson error in that it inflates the standard deviation of the health effect estimate and introduces little or no bias (Buzas et al., 2014). However, unlike conventional Berkson error, it may not be completely independent of the predicted exposures, and it may also be correlated in space. Another component of error comes from uncertainty in estimating the exposure model parameters. It is similar to classical measurement error in that it is a source of variability in the predicted exposures and can introduce bias in health effect estimates as well as change their standard errors. The two components are clearly apparent for a simple situation where a linear regression model predicts the effect of true exposure on a health outcome, and a LUR model based on a selected set of potential predictors is used to estimate the true exposure at the residential locations of study participants. Two types of error are introduced by using exposure estimates instead of the true exposure. One component is the error of prediction that will be present even if the coefficients of the LUR model are known exactly. This component is Berkson error and it increases the variance of the estimated coefficient in the health model, but does not bias it. The second component results from estimating the coefficients of the LUR model and is usually termed classical error. This component also increases the variance of the estimated coefficient in the health model, but can also introduce bias.

Basagana et al. presented findings for a LUR model with five predictor variables fitted to measurement data from 20, 40 or 80 locations (Basagana et al., 2013). The random error variance was set to two different values that resulted in proportions of explainable variability in the true exposure of 50% and 75%. The bias in the estimated health regression coefficient was in the form of attenuation towards the null hypothesis, but fell as the number of measurement sites used to build the LUR model increased and the attenuation factor ranged between 0.74 and 0.99 (see

Tables 1 and 2 of Basagana et al.). The degree of underestimation of the standard error of the health regression coefficient also fell as the number of measurement sites used to build the LUR model increased, and ranged between 3.4 and 8.6 fold. However, when variable selection from 20 or 100 predictor variables was allowed in the development of the estimated LUR model, much greater attenuation of the health model coefficients occurred (the attenuation factor ranged between 0.23 and 0.90), and the naive standard errors of the health model coefficient severely underestimated the true variation in all scenarios (between 14.7 and 25.3 fold). The strongest attenuation was found when the estimated LUR model could select among 100 predictor variables, the model was developed with 20 measurement sites, and the proportion of explainable variability in the true exposure was 50%. However, the naive standard error of the health model coefficient underestimated the true variation most when the estimated LUR model could select among 20 predictor variables, the model was developed with 20 measurement sites, and the proportion of explainable variability in the true exposure was 75%. Coverage of the 95% confidence intervals based on naive standard errors was low in all scenarios, but especially when variable selection was performed. Similar analyses were performed for logistic regression and the results were very similar to those for linear regression in terms of attenuation, but naive standard errors showed much less underestimation, leading to higher coverage values for confidence intervals. Szpiro et al. also performed simulations to investigate the effect of misspecification in a 3 predictor variable model (Szpiro et al., 2011a). Bias in the health effect estimates was very small, but in one situation the mean of the estimated health regression coefficient was higher than the true value. The main effect of exposure measurement error in spatial prediction was an increase in variance estimates, and this was also observed for another measurement error framework by Szpiro et al. (Szpiro et al., 2011b). The impact of measurement error is a complex issue, but it is clear that Sheppard et al. correctly identified that the precision of many reported health effect estimates may be overstated due to bias and/or incorrect standard errors (Sheppard et al., 2012).

As is often the case, the use of potentially inaccurate exposure estimates is justified on the basis that exposure misclassification is assumed to be non-differential, limiting the ability to detect associations. A recent review of the use of geographically modelled environmental exposure estimates concluded that 'exposure misclassification is too often dismissed based on the erroneous justification that an accurately classified exposure would have led to an even stronger estimate of relative risk' (Chang et al., 2014). The same review also noted that differential error can be postulated under a number of reasonable scenarios. Bias is not guaranteed to be toward the null even if the misclassification is approximately non-differential (Jurek et al., 2008).

3.4. TRUE RISK PER UNIT OF EXPOSURE

The results of epidemiological studies of air pollution are typically reported as an increase in the risk of an adverse health outcome such as lung cancer mortality that is associated with an increment of air pollution. These risk estimates are used to quantify the effects of current levels of air pollution i.e. the health or mortality burden on a population, or the effect of changes in the level of air pollution (Burnett et al., 2014). However, it has been shown in section 3.2.3 that estimates of some pollutants at the address of a subject may differ greatly from the level during the BREP for health endpoints with a long latent period. A good example is the latest update of the H6C study (Lepeule et al., 2012). For lung cancer, the investigators reported a RR of 1.37 (95% CI 1.07-1.75 per 10 $\mu g/m^3$ increase in $PM_{2.5}$). This is the largest of four estimates of the association between lung cancer and $PM_{2.5}$ exposure used by Burnett et al. to calculate burden, but it almost certainly overestimates

the effect of exposure. The analysis using Cox regression models by Lepeule et al. treated exposure as an annual time-dependent variable and used a 1- to 3-year moving average for lung-cancer mortality analyses (a 3-year time window). As noted in section 3.2.3, the falling levels of $PM_{2.5}$ mean that $PM_{2.5}$ levels during the BREP (1959-94) for a lung cancer death occurring at the end of the follow-up period in 2009 would have had a much wider range than the 1- to 3-year moving average in 2009, most likely at least 10 times wider. The difference is not as great earlier in the study follow-up period (1982-2008), but more lung cancer deaths will have occurred towards the end of the follow-up period and the RR estimated using 1- to 3-year moving average $PM_{2.5}$ levels will almost certainly considerably overestimate the true risk per unit of $PM_{2.5}$. One of the strongest indications of the importance of this effect is the very high RR of 2.84 (95% CI 1.06–7.59 per 10 $\mu g/m^3$ increase in $PM_{2.5}$) reported by Lepeule et al. for the follow-up period 2001-2009. This implausibly high RR is almost certainly due to the very small variation in exposure between cities during 2001-2009 compared to that during the BREP, assuming the exposure contrasts during the two periods were highly correlated. In section 3.2.3 it is shown the RR would reduce to at least 1.19 if exposures in the analysis were scaled upwards by a factor of six (the range of mean $PM_{2.5}$ exposures during the BREP for lung cancer is likely to be at least 6 times higher than the range of the 1- to 3-year moving average for $PM_{2.5}$ in 2005).

Few investigators acknowledge the problem, but it was recognised by Raaschou-Nielsen et al. who stated that “with decreasing air pollution concentrations and contrasts over time, risk estimates based on recent contrast might be too high” (Raaschou-Nielsen et al., 2013). The authors investigated this by back-extrapolating contrasts in two cohorts with long-term $PM_{2.5}$ monitoring measurements, and in seven cohorts with long-term PM_{10} monitoring measurements. However, the investigators only reported the effect on meta-HR when substituting original regression results with back-extrapolated regression results in the meta-regression. Unsurprisingly, the results for $PM_{2.5}$ were stated to be identical as only 2 of the 14 coefficients were changed. However, the meta-HR for PM_{10} fell from 1.13 to 1.09 when using the back-extrapolated contrasts. However, the investigators only extrapolated back to enrolment, and the absolute difference method used for particulate measurements is likely to have resulted in back-extrapolated estimates that were highly correlated with estimates for current exposures. This is because the adjustments were made using routine measurements from monitoring stations, but these were very few in number compared to the number of subjects. Hence, the HR would be unlikely to change. Given the low spatial contrast in exposure at most centres and the short period of extrapolation, it is likely that the adjustments made to individual exposure estimates had low variation compared to the within-centre variation at baseline.

4. CONFOUNDING

Confounding is a distortion in the estimated exposure effect that results from differences in risk between the exposed and unexposed that are not due to exposure, and a large portion of epidemiological methodology is concerned with avoiding or adjusting (controlling) for confounding (Rothman et al., 2008). Confounding is especially important in studies of air pollution and lung cancer because of the large effect of smoking; but in order for smoking to be a confounder, it also has to be associated with exposure to air pollution in the population from which the subjects arise. Smoking is not the only potential confounder. Samet et al. discussed other potential confounders in the context of a review of causes of cancer among never smokers (Samet et al., 2009). The authors noted that results on lung cancer risk from outdoor air pollution exposure among never smokers can be biased by residual confounding from occupational exposure to lung carcinogens and other social class-related factors. This section looks at how well study investigators have adjusted for smoking and other potential confounders with a focus on the reliability of the confounder measurements used, and the likelihood of residual confounding. It is noted that residual confounding by smoking is likely in many studies because the smoking information may be many years out of date, and residual confounding by smoking is also possible in studies of never smokers because of misclassification of current and ex-smokers as never smokers. It is also noted that many investigators claim to have controlled for occupational exposure to lung carcinogens, but the measures used are poor or not reported for some of these cohorts.

4.1. INDIVIDUAL VERSUS SPATIAL CONFOUNDING

Confounding has been described as a distortion of a true relationship between two factors (exposure and outcome) by the action of a third variable that is associated with exposure in the source population and, conditional on exposure, is also an independent risk factor for disease (Thomas, 2009). Confounding is a common and significant source of bias in environmental epidemiology, and the common refrain that this can be ameliorated by collecting a wider array of confounder data at the individual level may fall short of expectations.

When individuals with risk factors for disease or mortality are grouped together in certain areas, those areas will display the health effects stemming from the risk factors for those clustered individuals (i.e., “compositional clustering”) (Jerrett and Finkelstein, 2005). Contextual effects from such clustering occur when individual differences in health outcomes associate with the grouped variables that represent the social, economic, and environmental settings where the individuals live, work, or spend time (e.g., poverty in a neighbourhood) (Sheppard et al., 2012). Individuals living in the more polluted areas frequently have differing underlying health risks due to differences in a constellation of factors that influence health (e.g., smoking prevalence, diet, and socioeconomic status) compared to people living in less polluted areas (Dominici et al., 2014). Likewise, locational determinants of health (e.g., hospital or water quality) may also differ across places and are correlated with air pollution levels.

Statistically controlling for this form of confounding in the analysis is vital to generating less-biased results, as evidence suggests that the neighbourhood ecological/contextual effects associated with such clustering may influence health beyond individual risk factors for which data may have been collected (Sheppard et al., 2012). But, since many of these determinants of health remain unobserved (or at least unrecorded in research datasets), the conventional methods that rely on

statistical adjustment only for observed confounders is incomplete. Moreover, sufficient adjustments may not have been made even for the measurable differences. These unresolved challenges lead to an increased likelihood of biased effect estimates (Dominici et al., 2014).

The Cox proportional hazards modelling done in almost all of the major cohort studies of air pollution assumes that the survival experience among all subjects is statistically independent (Sheppard et al., 2012). Unfortunately, as discussed above, all but a few of those neighbourhood-level intrinsic risk factors can be included in any given model of any study due to data unavailability. An "incomplete" model may still appear to meet the independence assumption, but in reality it cannot. Further complicating matters is the possibility that study subjects living in clusters or different sub-clusters within the same cluster will share some lifestyle and environmental risk factors which, for individuals living in (sub)clusters living farther apart, are not as strongly shared. This suggests that large wide-area epidemiologic studies of the health effects of air pollution are inherently systematically biased since the mosaic of contextual correlations across a broad study area precludes the possibility that a single model is valid for all of the neighbourhood clusters within a study, and perhaps none of them. Whether or not these neighbourhood-level biases cancel each other out in the end is unknown.

An excellent example of compositional clustering is the Harvard Six Cities Study, a prospective cohort study of particular matter pollution and mortality, a "clean city vs. dirty city" comparison of mortality rates on which mortality RRs were calculated (Dockery et al., 1993). The results indicated a positive linear concentration-response, with the two cities with the highest RR (Steubenville, OH and Harriman, TN) also having the highest concentrations of particulate matter. As a result of their positioning on the scatterplot, those two data points significantly influenced the slope of the line towards positivity—that is, a statistically significant relative risk of 1.26 for the association between mortality and fine particulates. The linear model inferred that PM was an important causal agent as was smoking cigarettes. In their analysis, the authors adjusted for sex, smoking, BMI, educational attainment, and occupational exposure to fumes, dusts, and gases at the individual level. While this statistical adjustment removed some of the confounding, the study failed to control for compositional confounding factors endemic to the cities being compared. Both Steubenville and Harriman are located in areas of the U.S. where quality of life indicators are low—high prevalence of chronic disease such as heart disease, high drop-out rates from school, and poverty—and where pollution levels are high. The remaining four less-polluted cities with lower mortality rates fare better on those indicators. One would reasonably expect that Steubenville and Harriman would suffer from high mortality rates from a host of contextual factors even without the PM contribution, as their initial levels of health were relatively poor. Nevertheless, the statistical correlation between PM and mortality was apparent, albeit a potentially spurious association due to the presence of contextual confounding.

Finally, let us consider whether or not contextual factors are, in fact, confounders. Confounders must be associated with both the outcome and the exposure. Regarding the association with health outcomes, contextual factors acting together can result in "sick neighbourhoods", and individuals within those communities are therefore more likely to experience the health or mortality outcomes under study since they have initial reduced levels of health due to contextual factors (Dominici et al., 2014). For low educational attainment alone--one of many candidate contextual confounders-- several studies have observed positive statistical associations with air pollutants (Dockery et al., 1993, Hoek et al., 2013, Krewski et al., 2000, Lipsett et al., 2011, Miller et al., 2007, Pope et al., 2002). Regarding the exposure criterion, the greatest density of area air pollution monitors is in areas of

high population density. These urbanized areas also have a large number of individuals clustered in low SES areas with associated lifestyle-related health risks. Hajat et al. found statistically significant association of neighbourhood-level SES (NSES) measures with predicted air pollutant ($PM_{2.5}$ and NO_x) concentrations using cohort data from the Multi-Ethnic Study of Atherosclerosis in the U.S. (Hajat et al., 2013). NSES was measured by income, wealth, education, and occupation. A previous extensive review also found that, in general, ambient air pollution concentrations were higher in areas of lower SES. In summary, there exists a real potential for contextual confounding in observational epidemiologic studies of air pollutants, and current analytical methods (O'Neill et al., 2003).

4.2. SMOKING

Potential confounding by smoking is of particular importance because of the strength of effect. The strength of the smoking effect relative to the potential effect of air pollution is well illustrated by Figure 1 of Pope et al (Pope et al., 2011).

4.2.1. Quality of individual smoking data

The majority of studies of particulates and lung cancer possess individual level smoking information. Of the studies included in the meta-analysis by Hamra et al., only the studies by Cesaroni et al and Naess et al. did not possess individual level smoking information, so the results of the latter study could not be included in the meta-analysis (Cesaroni et al., 2013, Hamra et al., 2014, Naess et al., 2007). In addition, a study by Hart et al. only had individual smoking information for a small sample (approximately 5%), and the authors assessed the potential impact of unmeasured confounders using this small subset of subjects (Hart et al., 2011). Nevertheless, it has long been recognised that smoking history information collected in many epidemiological studies has reliability issues. For example, respondents in a U.S. national survey were questioned about their smoking habits in 2002 and 2003 and 11% who reported in 2002 being current or former smokers claimed in 2003 that they had never smoked (Soulakova et al., 2012). It is well known that a small proportion of subjects known to smoke currently or to have smoked in the past based bioassay results or repeated questionnaires will report as never having smoked (Wells et al., 1998). Misclassification of smokers as never smokers is of particular interest because of concern that the association between second hand tobacco smoke (SHS) and lung cancer could in whole or in part be ascribed to an upward bias caused by misclassification of smokers as never smokers. Such a bias is of less concern in air pollution studies, but it does mean that residual confounding by smoking may still be an issue in studies of never smokers.

Nevertheless, the individual smoking information used in some air pollution cohort studies has an important limitation resulting from the fact that it is usually only collected at baseline and not updated. For example, individual smoking histories collected in 1982 were used in analyses of mortality during the period 1982-2000 in ACS CPS-II study in both the full study (Krewski et al., 2009) and a subgroup of Californian subjects (Jerrett et al., 2013). Jerrett et al. noted that no follow up surveys were conducted in the full ACS CPS-II study, and key lifestyle characteristics may have changed during the follow up period. It was also noted that smoking rates declined precipitously across California between 1982 and 2000, and if the declines in smoking rates were spatially associated with the air pollution levels, then these would have had the capacity to confound the air pollution risk estimates. The problem is acknowledged by Turner et al. in a study published after the IARC evaluation that assessed possible joint effects of cigarette smoking and $PM_{2.5}$ exposure on lung cancer risk in ACS CPS-II subjects (Turner et al., 2014). For this analysis, follow-up was truncated at the first six years in order to classify

participants by cigarette smoking status, because updated information on cigarette smoking status was not collected after enrollment. The problem is even more pronounced for the H6C study. In the most recent update, Lepeule et al. used smoking information collected at enrolment (1974 to 1977) when analysing lung cancer mortality over a 36 year period from 1974 to 2009 (Lepeule et al., 2012). Information on smoking status (never, former, current) and cumulative smoking (pack years included separately for current and former smokers) was collected from subjects at enrolment. The subjects were aged 25-74 years at enrollment and clearly the information on cumulative smoking, and probably smoking status, of subjects who were young at enrollment is unlikely to be sufficiently accurate to permit adjustment for the smoking habits during the most recent period of follow up. However, the overall risk estimate for lung cancer and $PM_{2.5}$ exposure is strongly influenced by the risk estimate for the most recent follow-up period (2001-2009) of 2.84 (95% CI 1.06-7.59) for a 10 $\mu g/m^3$ increase in $PM_{2.5}$.

4.2.2. Value of proxy measures for smoking

The use of smoking histories collected at baseline is clearly a problem for most air pollution cohort studies except for those with a very short period of follow up (Hales et al., 2012). Errors in confounders compromise our ability to control for their effect, leaving residual confounding; if smoking is not associated with air pollution then there is no confounding or residual confounding even if the smoking data is unreliable. However, the likelihood of residual confounding by smoking is rarely acknowledged.

As noted in section 4.2.1, the majority of studies of particulates and lung cancer possess individual level smoking information. One notable exception is the large census-based study by Cesaroni et al. (Cesaroni et al., 2013). This is one of the most influential studies in the meta-analysis by Hamra et al., but the findings could be easily explained by the confounding effect of smoking (Hamra et al., 2014). If it is assumed that current smokers of 20 cigarettes per day have a 20 fold increased risk of lung cancer and the smoking prevalence among subjects in the lowest quintile of $PM_{2.5}$ exposure is 30%, then smoking prevalences of 31.4%, 33.2%, 32.5%, 32.8% in the higher quintiles would give RRs equal to the fully adjusted HRs for lung cancer reported in Table 2 of Cesaroni et al.

Cesaroni et al. adjusted for individual predictors of smoking prevalence such as education level and occupation, but also area-level SES factors which have been shown to be associated with smoking, after accounting for individual education and occupation (Roux et al., 2003). The area-level SES measure was a five-level small-area (census block; average population of 500 subjects per block) socioeconomic position index that was derived based on a factor analysis including education, occupation, house ownership, family composition, crowding, and immigrant status. It is unclear how well adjustment for individual factors and the socioeconomic position index used by Cesaroni et al. compares with adjustment using individual smoking data. Sanmartin et al. were able to predict smoking status reasonably well using statistical modelling techniques and census information including education and employment status, but the set of socio-economic and demographic characteristics that were predictive of smoking status varied by age and sex (Sanmartin et al., 2013).

Information on diet, alcohol consumption, and obesity were also not available in the RoLS and Cesaroni et al. further adjusted their models for pre-existing comorbidities related to smoking habits and diet [chronic obstructive pulmonary disease (COPD) hypertensive heart disease, and diabetes]. However, there is little information available to assess the value of adjusting for pre-existing comorbidities.

Cesaroni et al. cited a study by Gan et al. in support of their approach, although these authors used the same three conditions to define a pre-existing co-morbidity which was used as a proxy variable for common behavioural risk factors for coronary heart disease (Gan et al., 2011), and Gan et al. cited Pope et al. in support of their approach, but Pope et al. used age-standardized lung cancer and COPD death rates for U.S. counties as proxy indicators of cigarette smoking prevalence in a study which evaluated changes in life expectancy associated with differential changes in fine particulate air pollution (Pope et al., 2009).

It is well known that COPD is related to smoking. A prospective study by Lokke et al. with a 25 year period of follow up, found that approximately 25% of smokers developed COPD, and continuous smokers were 6.3 times more likely to develop COPD than never smokers (Lokke et al., 2006). In addition, COPD appears to be an independent predictor of lung cancer risk after adjusting for smoking, which might reflect in part, a shared genetic susceptibility to chronic smoking-induced inflammation (Young et al., 2009). However, the increased risk of lung cancer among those with COPD is of much lower in magnitude than the risk of lung cancer among smokers. Given that the prevalence of COPD in the RoLS was only 2.0%, adjusting for COPD was clearly not sufficient to adjust for the possible confounding effect of smoking. Cesaroni et al. did not report the lung cancer risk associated with pre-existing COPD, so it is not possible to assess what impact adjusting for COPD might have had (Cesaroni et al., 2013). The authors noted that Gan et al. asserted that "adjustment for pre-existing conditions might have led to an underestimation of the effect, because the comorbidities might act as intermediate variables" (Gan et al., 2011). However, it is not clear that COPD is an intermediary for particulates exposure and lung cancer. Dimakopolou et al. reported no association between air pollution exposure and non-malignant respiratory mortality among ESCAPE study subjects (Dimakopoulou et al., 2014). However, even if COPD is an intermediary for particulates exposure and lung cancer, the impact of adjusting for COPD will be greatly outweighed by the impact of not adjusting for smoking.

In conclusion, there is little evidence available to assess whether adjusting for individual predictors of smoking prevalence such as education level and occupation, and area-level SES factors is an adequate substitute for smoking history adjustment. COPD may reflect the cumulative effects of smoking, but it is clear that it is not a good proxy indicator for individual smoking history. However, when smoking history is available, there may be a good argument to adjust for COPD as well as smoking history.

4.2.3. Is evidence from never smokers sufficient to rule out residual confounding?

The ability to detect an effect of air pollution among never smokers would allay some concerns about residual confounding by smoking. However, if the same effect of air pollution was observed for current, former and never smokers, it could indicate residual confounding by other agents. Indeed, it seems unlikely that any association would be observed for current smokers as their intake of particulates from cigarettes dwarfs their intake from air pollution (see section 7.1), and the same is probably true for former smokers. Hamra et al. has recently conducted meta-analyses of 18 studies examining the relationship of exposure to $PM_{2.5}$ and PM_{10} with lung cancer incidence and mortality, including analyses by smoking status (Hamra et al., 2014). As discussed in Section 2.5, Hamra et al. noted that the meta-analyses originated with the IARC review and complement the qualitative classification of the evidence by the IARC Working Group, but include results from the NHS study reported by Puett et al. after the IARC evaluation (Hamra et al., 2014, Puett et al., 2014). However, only a few of the studies reported risk estimates

by smoking status. In the case of $PM_{2.5}$, Hamra et al. included risk estimates for current, former and never smokers from 4 studies: ESCAPE, H6C, ACS CPS-II and CNECSS (Hystad et al., 2013, Lepeule et al., 2012, Pope et al., 2002, Raaschou-Nielsen et al., 2013). Risk estimates for never smokers were also included in the meta-analysis from two other studies: the CTS study and the Nurse's Health Study (NHS) study (Lipsett et al., 2011, Puett et al., 2014). Turner et al. also reported results for never smokers in the ACS CPS-II study for a longer follow-up period than Pope et al (Pope et al., 2002, Turner et al., 2011). Katanoda et al. reported lung cancer risks associated with $PM_{2.5}$ for male current smokers, male former smokers and female never smokers included in TPCS (Katanoda et al., 2011). These were not included in the meta-analysis, although the results are based on 455 of the 518 lung cancer deaths that occurred during the follow-up period. There is even less information available for PM_{10}, and only ESCAPE reported risk estimates for current, former and never smokers, and the CTS and NHS reported results for never smokers (Lipsett et al., 2011, Puett et al., 2014, Raaschou-Nielsen et al., 2013). Puett et al. also reported results for $PM_{2.5}$ and PM_{10} for a group including people who had never smoked or had quit for at least 10 years, and a group including people who currently smoked or had smoked in last 10 years (Puett et al., 2014).

Hamra et al. reported that lung cancer risk associated with 10 $\mu g/m^3$ $PM_{2.5}$ was greatest for former smokers (meta-RR 1.44; 95% CI: 1.04-2.01), followed by never-smokers (meta-RR 1.18; 95% CI: 1.00-1.39), and then current smokers (meta-RR 1.06; 95% CI: 0.97-1.15) (Hamra et al., 2014). Puett et al. states that many population studies of associations between chronic exposures to ambient particulate matter (PM) and/or traffic related pollutants with lung cancer "have observed effect modification by smoking status, providing evidence for the link between PM exposure and lung cancer in the absence of the strong influence of smoking behaviour" (Puett et al., 2014). However, none of the studies included in the meta-analyses by subgroups of current, former, and never-smokers reported statistically significant effect modification/interaction by smoking status, and a test for homogeneity performed by Hamra et al. suggested no evidence of difference between subgroups (p=0.197). The meta-analysis provides some evidence of an effect among never smokers, but the information available is limited and the ACS CPS-II study has a weighting of approximately 75% in the meta-analysis for $PM_{2.5}$. Of the six studies included in the never smoking meta-analysis for $PM_{2.5}$, four studies reported higher RRs for never smokers and two studies reported lower RRs for never smokers. However, the meta-analysis did not provide evidence that the effect of air pollution was different for current, former and never smokers. There is even less difference between the meta-RR for the smoking subgroups if the risk estimates for TPCS are included in the meta-analysis (see Figures 5-7) (Katanoda et al., 2011). The lung cancer risk associated with $PM_{2.5}$ remains the highest for former smokers (meta-RR 1.33; 95% CI: 1.04-1.69), but the meta-RR for never-smokers (1.17; 95% CI: 1.05-1.30), and current smokers (1.19; 95% CI: 1.01-1.40) are now very similar. The meta-RR for former smokers is of a similar magnitude to the lung cancer risk associated with $PM_{2.5}$ reported by Puett et al. among a group of never smokers and smokers who had quit for at least 10 years (RR 1.37; 95% CI: 1.06-1.77). This group is dominated by former smokers who contributed 45% of person-years, but 79% of the lung cancer deaths. A recently published study of Canadian women was not included in the meta-analysis but only reported an increased risk of lung cancer for ever smokers (Tomczak et al., 2016). The authors reported a significantly elevated adjusted HR for lung cancer of 1.34 (95% CI 1.10- 1.65) in relation to an increase of 10 mg/m^3 increase in $PM_{2.5}$ exposure, but the increased risk was limited to those who smoked cigarettes (adjusted HR=1.40; 95% CI: 1.12-1.73) and there was no increased risk for never smokers (adjusted HR=1.01; 95% CI: 0.56-1.80) and (Tomczak et al., 2016). After the inclusion of risk estimates reported by Katanoda et al., only the risk estimates for never smokers show no evidence of heterogeneity

(I^2=0%, p=0.97). There is evidence of heterogeneity for current smokers (I^2=69.4%, p=0.011), and the test of heterogeneity for former smokers is close to statistical significance (I^2=57.2%, p=0.053). Hamra et al. could only perform a meta-analysis of the lung cancer risk associated with PM_{10} for never smokers (meta-RR 1.11; 95% CI: 0.94-1.33). Raaschou-Nielsen et al. reported a significantly elevated lung cancer risk associated with PM_{10} among current smokers (RR 1.27; 95% CI: 1.06-1.77), but Puett et al. reported no association among nurses who were either current smokers or had smoked in the last 10 years (RR 0.99; 95% CI: 0.88-1.12) (Raaschou-Nielsen et al., 2013). Overall, the available evidence goes some way to allay concerns about residual confounding by smoking, but questions about residual confounding by other factors are raised by the associations observed among current and former smokers.

Figure 5 Estimates of lung cancer risk associated a 10-µg/m^3 change in exposure to $PM_{2.5}$ for never smokers. Weights represent the contribution of each study effect estimate to the overall meta-estimate.

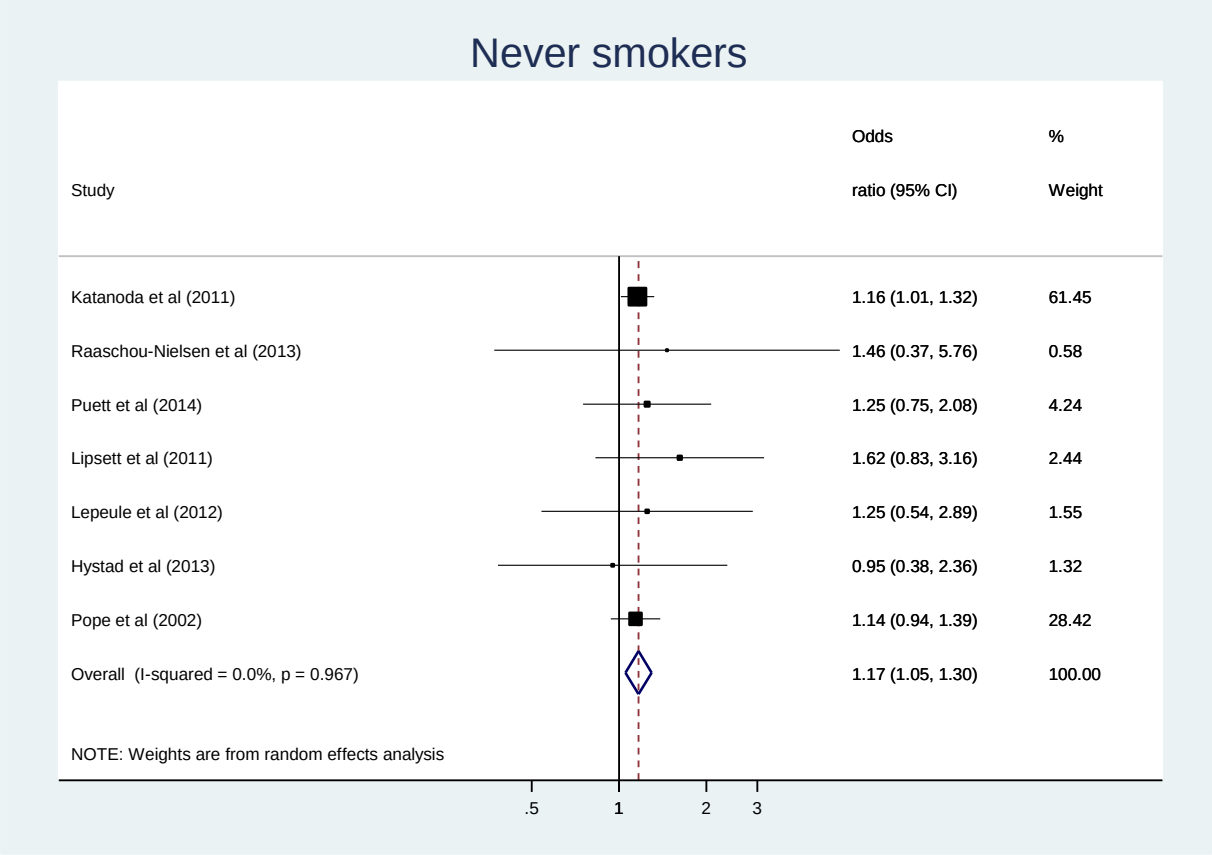


Figure 6 Estimates of lung cancer risk associated a 10-µg/m^3 change in exposure to $PM_{2.5}$ for former smokers. Weights represent the contribution of each study effect estimate to the overall meta-estimate

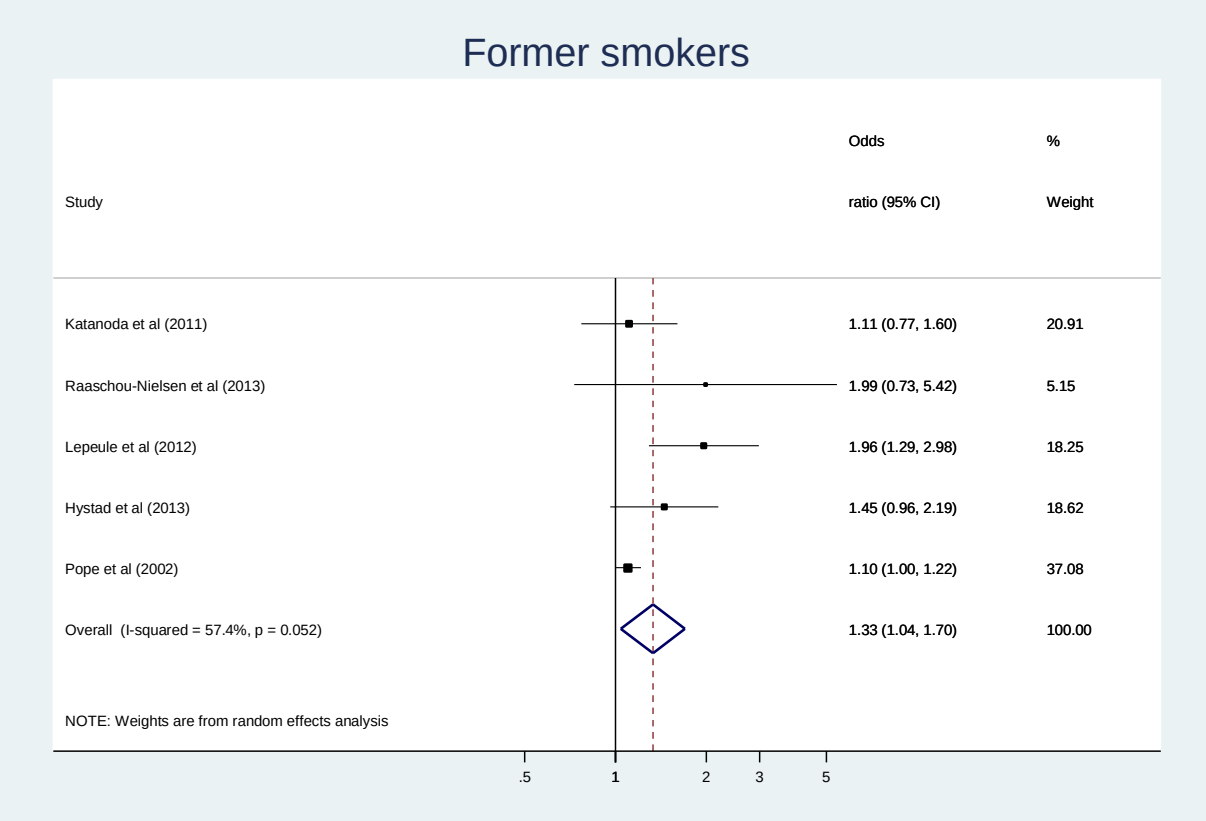
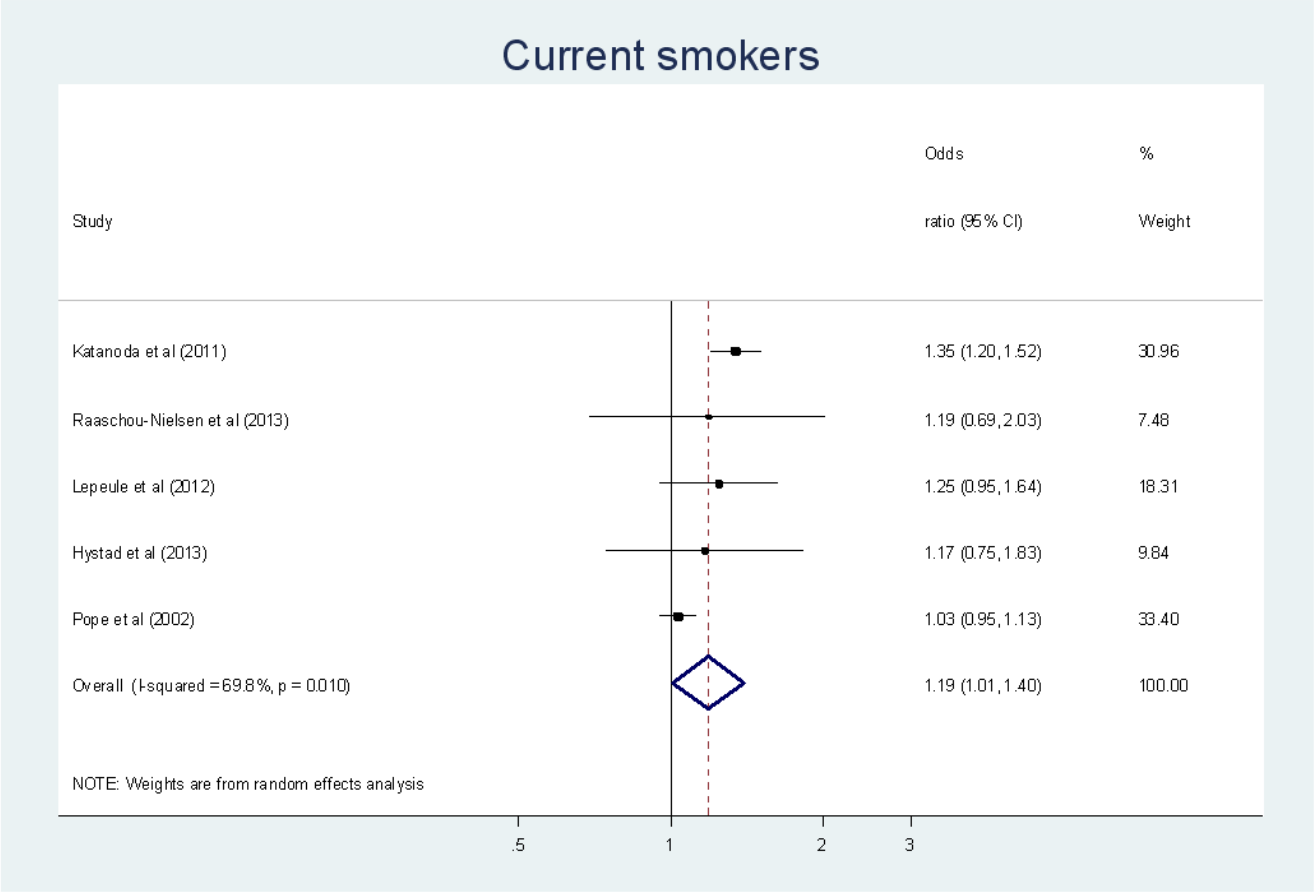


Figure 7 Estimates of lung cancer risk associated a 10-µg/m^3 change in exposure to $PM_{2.5}$ for current smokers. Weights represent the contribution of each study effect estimate to the overall meta-estimate



4.2.4. Does restricting analysis to adenocarcinoma reduce the impact of residual confounding?

Of the 3 major histological subtypes of lung cancer (squamous cell carcinoma, adenocarcinoma and small cell carcinoma), adenocarcinoma is the most common subtype in never smokers (Samet et al., 2007). Consequently, it has been suggested that restricting the analysis to adenocarcinoma may be useful to assess for causes other than smoking, compared with, for example, patients with squamous-cell carcinomas (Raaschou-Nielsen et al., 2013). However, residual confounding by smoking may still be a major problem if analyses are restricted to adenocarcinoma. In a pooled analysis of case-control studies which included 13,169 lung cancers, 45% of lung cancers occurring in never smoking men were adenocarcinomas and 56% of lung cancers in never smoking women (Pesch et al., 2012). However, it can also be deduced from Pesch et al. that the proportion of adenocarcinomas in men that occurred in never smokers was only 4.2% compared to 1.1% of squamous and small cell carcinomas, although the difference was much greater for women, and 33.9% of adenocarcinomas occurred in never smokers compared to 14.5% of squamous and small cell carcinomas. Overall, restricting the analysis to adenocarcinoma does not substantially increase the proportion of never smokers (13.0%) compared to analyses based on squamous and small cell carcinomas (2.9%), or all histology (6.3%). In addition, adenocarcinoma is also strongly associated with smoking. The strength of the association between cigarette smoking and lung cancer varies by cell type, with the odds ratios historically largest for squamous and small cell carcinomas and somewhat smaller for adenocarcinoma, but some studies suggest that the higher relative risk (or odds ratio) for small cell or squamous cell carcinoma simply reflects lower baseline risks in non-smokers of these two subtypes as compared with adenocarcinoma (Khuder, 2001, Yang et al., 2002).

Raaschou-Nielson et al. noted that they were not able to rule out potential residual confounding from smoking (because data for smoking were obtained at enrolment, and they did not account for changes in smoking habits during follow-up), but they considered that if residual confounding was present, then squamous-cell carcinomas should have been associated with air pollution, but in their study the association was mainly with adenocarcinoma (Raaschou-Nielsen et al., 2013). However, as noted above, adenocarcinoma is also strongly associated with smoking. For PM_{10}, Raaschou-Nielsen et al. did not observe an association with squamous cell carcinoma (HR=0.84 per 10 µg/m^3; 95% CI 0.50-1.40), but a strong association with adenocarcinoma was observed (HR=1.51 per 10 µg/m^3; 95% CI 1.10-2.08). However, for $PM_{2.5}$, the associations were similar for adenocarcinoma (HR=1.55 per 5 µg/m^3; 95% CI 1.05-2.29) and squamous cell carcinoma (HR=1.46 per 5 µg/m^3; 95% CI 0.43-4.90), suggesting that residual confounding could be present (Raaschou-Nielsen et al., 2013). Hystad et al. reported similar associations with $PM_{2.5}$ in the CNECSS study whether the analysis was restricted to adenocarcinoma (OR=1.27 per 10 µg/m^3; 95% CI 0.84-1.76), or all lung cancers (OR=1.29 per 10 µg/m^3; 95% CI 0.95-1.76) (Chen et al., 2007, Raaschou-Nielsen et al., 2013). Overall, analyses by histological subtype do not seem to be particularly useful to rule out the possibility of residual confounding by smoking.

4.3. CONFOUNDING BY OTHER FACTORS

There are a number of other potential confounders other than smoking; these include occupational exposure, socioeconomic status (SES) factors and radon.

4.3.1. Occupational Exposure

One of the most important potential confounders is occupational exposure. Siemiatycki et al. notes that occupational exposure is an important potential confounder in air pollution studies because it is plausible that individuals who live in highly polluted areas also work in more polluted environments (Siemiatycki et al., 2003). It has also long been recognised that a high percentage of lung cancer deaths/registrations are occupationally related, and Doll and Peto attributed 15% of male lung cancers and 5% of female lung cancers to occupational factors (Doll and Peto, 1981). A large proportion of these occupational lung cancers were attributed to asbestos exposure. However, the population attributable fraction varies considerably between studies and De Matteis et al. reported figures ranging from zero to 40% for males from 32 Italian and international studies (De Matteis et al., 2008). In the UK, Rushton et al. have estimated that 21.1% of male lung cancers and 5.3% of female lung cancers in 2004-2005 were attributable to occupation (14.5% overall): the main causes were estimated to be asbestos (40.8%), silica (16.7%), diesel engine exhaust emissions (12.8%) and mineral oils (8.6%) (Rushton et al., 2012).

Despite the potential importance of occupational confounding, few investigators have made adequate attempts to control for its influence. In some studies (e.g. the CTS study) occupational confounding is unlikely to be an important issue (Lipsett et al., 2011). The teachers in the CTS study share a relatively uniform occupational status, which precludes the need for statistical adjustment for potential exposures to lung carcinogens based on potentially problematic job exposure matrices; but there was no evidence of an association between lung cancer and particulate exposure in the CTS study. Some investigators such as Lepeule et al. acknowledge the potential for residual confounding by unmeasured factors such as occupational exposures if those factors co-vary with $PM_{2.5}$ (Lepeule et al., 2012). Other studies claim to have adjusted for occupation, but the information collected on occupational exposure is sometimes extremely basic and certainly insufficient to rule out confounding due to occupational exposure to lung carcinogens. For example, RoLS adjusted for occupation categorised as top qualified non-manually employed (i.e., managers, university and high school professors, researchers), other non-manually employed, manual labour employed, other employed (i.e., armed forces and retail sales), housewife, unemployed, retired or other (Cesaroni et al., 2013). Katanoda et al. adjusted for occupation which was the self-reported experience in occupations with potential exposure to gases, fumes, or dust (Katanoda et al., 2011). Raaschou-Nielsen et al. claim to have adjusted for occupation, but no information on occupational exposure to lung carcinogens was collected for 13 of the 17 subcohorts (Raaschou-Nielsen et al., 2013). Table S2 indicates that information on occupation was only collected for 7 subcohorts, but 3 of these subcohorts only have information about whether a subject was blue collar/manual or white collar/non-manual. In fact, information on occupational exposure was collected for only four of the 17 subcohorts, and this was very limited. Workers were categorised according to whether they had worked in an occupation

with a high risk of lung cancer, but three different definitions were used[1] and the percentages of workers having an occupation with a high risk of lung cancer ranged from 7% to 56%. The categorisation was extremely crude and a worker could have been employed in a job entailing exposure to asbestos for at least a year and not been categorised as having had an occupation with a high risk of lung cancer.

Considerable efforts were made to address possible confounding by occupational variables in a reanalysis of the ACS CPS-II and H6C studies (Krewski et al., 2000, Siemiatycki et al., 2003). The original ACS CPS-II study investigators only used self-reported exposure to six occupational dusts and fumes: asbestos, chemicals/ acids/solvents, coal or stone dusts, coal tar/pitch/asphalt, diesel engine exhaust, or formaldehyde and the H6C study investigators used self-reported occupational exposure to dusts, gases, and fumes. However, other occupational information was available: ACS CPS-II study participants were questioned at baseline about their current or last occupation, their occupation of longest duration and the time spent in these occupations, and H6C study participants listed their occupation and industry at the baseline interview. In the reanalysis, two new variables were developed: an indicator of the "dirtiness" of a subject's job and an indicator of possible exposure to occupational lung carcinogens. Possible lung carcinogen exposure was based on the job titles recorded by the original investigators at baseline and on the judgment of experts concerning typical exposure patterns in different occupations. A total of 2.7% of the ACS CPS-II study cohort were categorised as exposed to occupational lung carcinogens (6.0% of males and 0.3% of females), but the proportion rose to 7.5% among those who reported exposure to the one of the six occupational dusts and fumes (1.6% among those who did not report exposure). No breakdown by $PM_{2.5}$ exposure was given, but there was no association with sulphate particle exposure. The RR due to exposure to occupational lung carcinogens, as determined by the occupational lung carcinogens variable, was 1.23 (95% CI: 1.00-1.51) in the part of the cohort for which $PM_{2.5}$ exposure data were available. A total of 7.5% of the H6C cohort were categorised as exposed to occupational lung carcinogens (9.9% of males and 5.5% of females), and the proportion rose to 10.2% among those who reported exposure to dusts and fumes (5.3% among those who did not report exposure). There was some variability by town of residence in the prevalence of subjects exposed to occupational lung carcinogens, but prevalence was not clearly associated with the town's respective pollution level. In addition, subjects who had ever been occupationally exposed to known lung carcinogens did not exhibit an elevated risk of lung cancer.

The inclusion of the additional occupational variables did not materially change the results of any of the ACS CPS-II study analyses although there was no evidence of an association between lung cancer and $PM_{2.5}$ before the reanalysis: OR=1.01 (95% CI: 0.92-1.16) for a 10 $\mu g/m^3$ increase in $PM_{2.5}$. Nevertheless, Krewski et al. noted that even after the lung carcinogen index has been applied, the possibility of some residual confounding by occupation for mortality from lung cancer could not be ruled out (Krewski et al., 2000). In the H6C study, Krewski et al. reported that they found a insignificant excess in lung cancer risk related to fine particle air pollution, but this risk was attenuated considerably when the occupational confounders were included. However, much greater attenuation was achieved by controlling other covariates including extended smoking information, education level, and alcohol

[1] Latest or longest held job/industry was miner, rubber industry, leather tannery, shoe industry, metal worker, port worker, building construction worker, demolition, chimney sweeper, painting, truck- bus- taxi driver, china and pottery industry, butcher, car mechanic, waiter, chef, or electrician (SNACK). Same list but based on job/industry at enrolment (SALT). Ever employed for at least one year in the above job/industry list extended to include foundry, shipyard, glass industry, manufacture of asbestos or asbestos cement, asbestos insulation, cement article industry and welder (DCH, EPIC-San Sebastian).

consumption. Although, Krewski et al. made considerably more effort to address potential occupational confounding, they acknowledged that the occupational information collected from study subjects did not represent detailed lifetime work histories and that the validity of the occupation coding had not been established in relation to the actual jobs and occupations held, especially in the case of the ACS CPS-II study which coded job titles into only 68 categories, which indicated that the occupational lung carcinogens variable was relatively imprecise.

4.3.2. Socioeconomic Status/Position

SES/SEP has been shown to be a powerful driver of health status that could potentially confound environmental epidemiology studies if not adequately controlled. A recent study in the US demonstrated that ecological-level SES and behavioural factors associated with SES accounted for 74% of the variation in life expectancy among US counties (Dwyer-Lindgren et al., 2017). There is a large number of other potential confounders within the SES rubric (e.g. household income, educational attainment, census block GINI coefficients, income disparity, dirty vs clean occupations, diet/nutrition). Several of these have objectively confounded at least one study's results and were subsequently controlled for in that study's analysis. However, none have consistently held across all or even most studies. While SES may be difficult to operationalize in statistical models, it seems likely that some constituent of SES (or some "bundle" of factors) is consistently confounding the observational epidemiology studies. It is a reasonable assumption that every study's attempt at capturing the effects of SES is at best a partial success and hence residual confounding is a constant threat to the validity of the air pollution studies.

Epidemiologic studies that attempt to adjust for SES or SEP typically include 2-3 surrogate parameters in their statistical models; however, the list of candidate surrogates is in fact much longer. It is beyond the scope of this report to develop such a list and critique each of the proxy measures. But, some of the more commonly used measures will be briefly discussed.

Education is perhaps the most commonly used SES/SEP individual-lever surrogate in the literature despite some limitations that are often dismissed. This measure doesn't work well for the ends of the age spectrum. Seldom will young post-college adults (labelled high SES) have reached "cruising level" income levels directly after graduating, thus they may be living a low SES lifestyle over the short run, or even beyond depending on their chosen field. At the other end of the age spectrum, retirees from the previous generation were less dependent on a college (or even high school) education for prosperity and generous benefits in a manufacturing-based economy. On paper, they might be "low SES", but not in terms of their accumulated wealth.

Personal income is not always indicative of one's acquired wealth which is likely a better choice for an SES proxy. However, obtaining data on wealth is not practical. Home ownership is less frequently used as a marker for relatively higher SES, but fails to include the property's value or how the home was acquired (e.g., via inheritance).

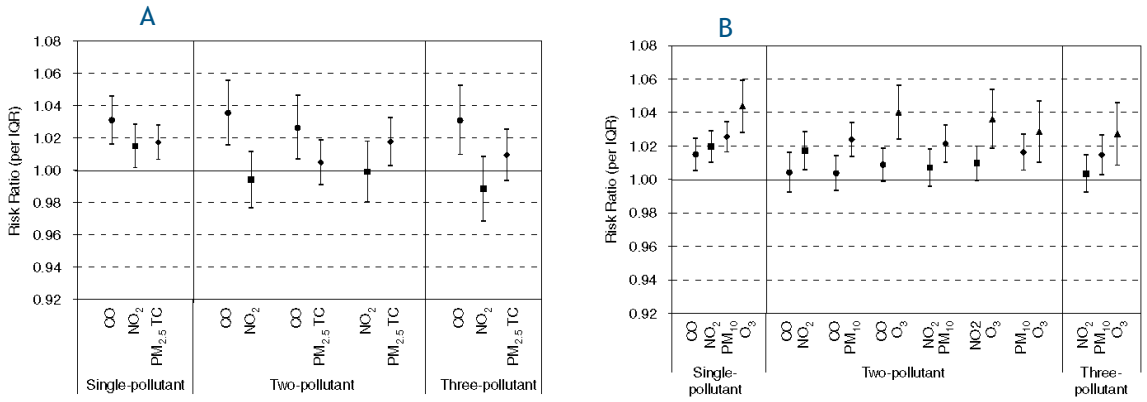
Several ecological-level SES variables and indices have been used. While some of these "neighbourhood effects" variables might be competent SES markers for some individuals, the ecological-individual correlation may be low for others. In short, capturing the full effects of SES/SEP on both levels of observation has remained elusive, thus one must assume that an unknown degree of residual confounding will be present in every study.

4.4. SINGLE OR MULTIPLE POLLUTANT MODELS

Multipollutant models have the ability of factoring collinearity issues into consideration, and for identifying those pollutants showing the strongest relationship with a particular outcome (Krall et al., 2015). When properly used, they are capable of determining whether the associations observed in single pollutant models are confounded by exposure to other pollutants found in ambient air. Multipollutant modelling has almost become a necessity in observational studies relying on measurements from fixed monitoring sites because of the complex relationships that can exist between commonly encountered air pollutants (Dionisio et al., 2014). The failure of these networks to adequately account for spatial and temporal variability in ambient air concentration can result in a substantial amount of error depending on the circumstances. Pollutants emitted by local sources may show considerable spatial heterogeneity that cannot be captured by centrally located monitoring sites e.g. carbon monoxide (CO), NO_x, and elemental carbon (EC). Oftentimes these pollutants will show a high degree of co-linearity so any associations observed using a single pollutant model may be confounded by the presence of other co-related pollutants in the ambient air. Under these circumstances multipollutant modelling will help disentangle the relative contribution of individual pollutants to the overall risk.

For instance, Tolbert et al. examined the association between emergency department visits for cardiovascular or respiratory illness and exposures to three pollutants both singly and in combination (Tolbert et al., 2007). As shown in Figure 8, statistically significant associations were observed between three-day moving average air exposures and the risk ratios for cardiovascular and pulmonary emergency department visits using a single-pollutant model. When two-pollutant or three pollutant models were applied using NO_2, CO, $PM_{2.5}$ EC, ozone or PM_{10} in some combination, the significant associations for NO_2 were no longer observed. Multi-pollutant modelling is necessary when individual pollutants are treated as independent risk factors but their exposure concentrations are strongly correlated, (Kim et al., 2007). This is often the case with NO_2, whose airborne concentrations are often significantly correlated with other pollutants such as $PM_{2.5}$, ultrafine particles (UFP), and CO. In the above example, the results of multi-pollutant modelling showed that the strongest associations with cardiovascular visits occurred with CO, whereas the respiratory visits were associated with ozone. This example demonstrates the importance of multi-pollutant modelling for NO_2 and highlights the erroneous associations that may arise when single-pollutant models are relied upon as the primary basis for establishing an association. Still, multi-pollutant modelling is not a panacea, and assumptions regarding co-linearity between pollutants, non-differential seasonal effects, and the absence of appreciable interactions with other physical or environmental factors cannot be guaranteed.

Figure 8 Results of selected multi-pollutant models for combined cardiovascular (A) and respiratory (B) diseases group, 1998–2004, Study of Particles and Health in Atlanta (SOPHIA) (Tolbert et al., 2007)



Although providing a methodological improvement, multi-pollutant models are not able to rule out that the pollutant with the highest adjusted risk estimate may be acting as a surrogate for other unmeasured exposures that are the true causative agent. This was well demonstrated in a study by Brook et al., who re-examined previous results of an association between NO_2 exposures and non-accidental mortality in 10 Canadian cities (Brook et al., 2007, Burnett et al., 2004). Although the original study indicated an NO_2-associated increase in mortality that was unaffected by adjustments for O_3, CO, or SO_2 in a two-pollutant model, the authors re-examined their results and concluded that NO_2 could be acting as a surrogate for a specific $PM_{2.5}$ component or possibly even volatile organic compounds (VOCs), polycyclic aromatic hydrocarbons (PAHs), or other oxidized nitrogen species. The authors based this opinion on the fact that high correlations were observed with a number of other traffic-related pollutants including VOCs and PAHs and that the strongest associations were observed in the summer months when photochemical reaction products are more frequently encountered in aged urban air masses.

4.5. CONFOUNDING INTRODUCED BY LAND USE REGRESSION MODELS

It is also worth noting that some of the predictors used for developing air pollution exposures with LUR models could introduce confounding when the exposure estimates are applied in epidemiological studies. Hoek et al. notes that LUR models including population density, for example, may be problematic as population density may also be associated with other adverse risk factors such as low socio-economic status or poor housing stock, which could influence some diseases of interest such as asthma (Hoek et al., 2008b). The authors note that one solution to this potential problem is the inclusion of area-level confounders that are more closely related to the disease of interest (e.g. percentage of low-income families in a neighbourhood) than the variable used in predicting air pollution (e.g. number of addresses in a 300 m buffer). Cefalu and Dominici note that the current literature treats the aspect of exposure prediction to overcome missing exposure e.g. a land use regression model, and the process of confounding adjustment in the health-effects regression model, as two distinct topics (Cefalu and Dominici, 2014). That is, methods that account for measurement error in the predicted exposure often do not consider whether predicting exposure with covariates that are correlated with the outcome might bias the health-effect estimates, while methods designed to control for confounding adjustment often do not acknowledge the possibility of confounding, while methods designed to control for confounding often fail to acknowledge that the exposure has been predicted rather than measured. They use

theoretical arguments and simulation studies to show that the bias of a health-effect estimate is influenced by the exposure prediction model, the type of confounding adjustment used in the health-effects regression model, and the relationship between these two. However, a full discussion of this problem is beyond the scope of this report, and it may not be particularly relevant to studies of lung cancer.

5. QUANTITATIVE RISK ASSESSMENT

A quantitative health risk assessment (QHRA) traditionally involves four stages: hazard identification, concentration-response analysis, exposure determination, and risk characterization. The following section examines several key considerations that are applicable to defining the magnitude of any lung cancer risk that may be associated with air pollutant exposures. Two aspects of the QHRA are focused upon since they provide insight into several unique characteristics that may affect the overall risk determination. The first concerns the results from laboratory animal studies and the implications for the species differences that have been observed; whereas the second focuses on the shape of the concentration-response curve identified in epidemiology studies.

Assuming that the concentration response function is known with some accuracy, a QHRA may be performed using the hazard information obtained from humans or laboratory animal studies. In actuality, however, the results from human epidemiology studies are often used in preference to the findings from chronic exposure studies in non-humans because of the uncertainties associated with extrapolating the results from animals to humans. This preference for human hazard data has been adopted by IARC in their carcinogen classification scheme (Hesterberg et al., 2005). A recent comparison of the lung cancer risk estimates obtained using human and rat hazard data for exposures to poorly-soluble particulates provide some insight into the differences that exist with the two approaches (Kuempel et al., 2009). Lung cancer risks were determined for employees exposed for a working lifetime to coal dust, carbon black, titanium dioxide, diesel exhaust particulate, or crystalline silica. Following adjustments for the physiological differences between rats and humans, the risk estimates obtained from the rat studies were generally lower than those obtained using the results from occupational epidemiology studies; however, the differences were not statistically different. The authors concluded that the low statistical power of the available human studies, the limited particle size distributions for the human exposures, and questions surrounding the animal to human extrapolation methods may have influenced the results.

A more robust approach to a QHRA would combine the hazard findings from humans and laboratory animals. Such a technique was recently proposed for evaluating the lung cancer risks from human and non-human studies with diesel exhaust (Pedeli et al., 2011). After adjusting for biases in the three occupational studies and the species differences in the three rat carcinogenicity studies, risk ratios were calculated and the results pooled in a random effects meta-analysis. The analysis showed that the risk ratio for the combined data set was 1.49 (95% CI 1.21-1.78), whereas the results for the animal only and the human only data sets were 1.37 (95% CI 1.08-1.65) and 1.59 (95% CI 1.09-2.10), respectively. The study also found that the failure to adjustment for uncertainties in the human epidemiology studies resulted in risk ratios that were biased higher (i.e. 1.59 95% CI 1.28-1.89). Although these are interesting from a methodological perspective, they need to be placed in context, since some organizations including the USEPA advocate against using the results from rat carcinogenicity studies in a QHRA because of the particle overloading that uniquely occurs in this species (Ris, 2007).

5.1. LABORATORY ANIMAL STUDIES

Studies in laboratory animals provide a separate line of evidence regarding the hazard potential for a substance of concern. Generally, the information from all animal studies is examined to generate a weight-of-evidence (WOE) evaluation that describes the overall harm posed a substance and its relevance for humans. This includes studies performed by atypical routes of administration including subcutaneous administration, intraperitoneal injection, and intratracheal installation. In contrast, a health risk determination uses concentration-response data generated in laboratory animals treated by the most relevant route of human exposure, which is typically the inhalation route. Although dozens of studies have been performed with different types of the particulate matter found in ambient air, many involved routes of exposure that are not relevant to a risk determination. There are, however, a large group of inhalation studies in both rats, mice, and hamsters exposed to either diesel exhaust or products from coal or wood combustion. Of these three particulate sources, studies with diesel exhaust have been the most extensively reported upon and they have yielded some interesting insights into the lung cancer hazard from particulate exposures.

Surprising few chronic inhalation studies have been performed using the particulate matter collected from urban air (IARC, 2016). Two of the five identified studies were nearly 50 years old and failed to describe the experimental approach in sufficient detail. The remaining three studies were all performed in mice exposed to Sao Paulo air for 2-8 months after intraperitoneal treatment with urethane, which was used as a tumour initiator (Cury et al., 2000, Pereira et al., 2011, Reymao et al., 1997). Lung tumour promotion was observed under some treatment conditions, but not in others. The studies generally suffer from an underreporting of the results and a failure to include a clean air control group. Although these studies contain some deficiencies, others using diesel exhaust are far more robust and have dominated past discussions of the link between air pollution and lung cancer.

A total of 16 chronic inhalation studies have been performed in rats with the diesel exhaust exposures lasting up to 7 days per week for 20 hours per day. Several thorough reviews have been published that explore both the experimental designs and research results from these studies (Hesterberg et al., 2005, IARC, 2014). Many of these studies were performed over a twenty-year span beginning in the early 1980s. All but one of the studies involved whole body exposures to unfiltered diesel exhaust at concentrations ranging as high as 7-8 mg/m^3. A single nose-only exposure has been performed at particulate concentrations of 2 and 10 mg/m^3 (Stinn et al., 2005). Tumor incidence was examined immediately following the last exposure session or after a post-exposure recovery period that typically lasted six months. All but three of the studies showed some evidence of lung tumour development that included the appearance of lung adenomas and squamous cell carcinomas. The studies with no statistically significant increase in lung cancer were generally performed for shorter exposure durations or included a smaller number of test animals.

The positive results generally found in diesel exhaust exposed rats contrast with the findings from other animal species, including mice and hamsters (IARC, 2014). Five chronic inhalation studies have been performed in mice and three in hamsters. The mouse studies were performed at diesel particulate concentrations ranging from 0.35 to 7.0 mg/m^3 and lasted from 23 to 30 months. The three hamster studies employed particulate exposures ranging from 0.7 to 6.6 mg/m^3 for a period of 24 to 30 months. The exposure regimen in both species involved treatments that lasted 4 to 19 hours per day for 4 to 5 days per week. The results showed that diesel exhaust

was not tumorigenic in 7 of the 8 studies. A statistically significant increase in lung adenocarcinomas was observed in a single mouse study; however, an exposure-response relationship was not observed. These laboratory animal studies clearly showed that species differences existed in the susceptibility to lung cancer following diesel exhaust exposure with rats being far more sensitive than other animal species.

The observed species differences in diesel exhaust induced lung cancer was traced to the inability of rats to effectively clear the particulate from their lungs when exposed to high concentrations (Hesterberg et al., 2012, Oberdorster, 1995). This particle overload phenomenon occurs when the rate of particle deposition in the lung exceeds the rate of mucociliary clearance that is mediate by alveolar macrophages. This results in excessive particle accumulation in the lung and the initiation of an inflammatory response that can lead to the development of chronic health effects. This particle overload phenomenon is unique to the rat and does not take place in other animal species including humans. Consequently, many risk assessors, including those at the USEPA and the California EPA, have stated that the results observed in particulate inhalation studies in rats should not be used to determine human lung cancer risk (CALEPA, 1998, USEPA, 2002, Warheit et al., 2016). Surprisingly, although the recent IARC evaluation of air pollution and lung cancer described many of the laboratory animals studies that have been performed with diesel exhaust, there was no discussion of the rat lung overload observations and the relevance of these findings to the human hazard evaluation.

5.2. CONCENTRATION-RESPONSE

Concentration-response is an important issue which merits greater discussion than can be covered in this report. Most cohort and case-control studies have estimated relative risks assuming a log-linear relationship between a continuous exposure variable and lung cancer rates. The linear model assumption implies that there is no exposure level at which the risk to human health is not incrementally increased from zero upwards. In the case of particulates, there are a variety of mechanisms that may be acting and which have different implications for the shape of the response curve. Raaschou-Nielsen et al. noted that particulate matter with absorbed polycyclic aromatic hydrocarbons, transition metals, and other substances is capable of causing oxidative stress, inflammation, and direct and indirect genotoxicity (Raaschou-Nielsen et al., 2013). Pulmonary overload may also be important since toxicology studies of particulates have shown that adverse lung pathology is often displayed only after lung clearance mechanisms have been surpassed. Saturation of the clearance mechanism will result in appreciable lung inflammation as particle build-up occurs in the lungs, although the relevance of this mechanism is in dispute for humans (Borm et al., 2015). However, a more important factor may be the involvement of ultrafine particulates. The potential impact of nanoparticulates in the aetiology of lung cancer from environment exposures has not been adequately investigated and may explain many of the discrepancies that have been observed in competing epidemiology studies. Nevertheless, the linearity assumption may be realistic within the range of collected data.

Hamra et al. lists six studies of particulates and lung cancer that have considered alternatives to a linear exposure–response model: ACS CPS-II, H6C, CNECSS, RoLS, NH, and ESCAPE (Cesaroni et al., 2013, Hamra et al., 2014, Hystad et al., 2013, Lepeule et al., 2012, Pope et al., 2002, Puett et al., 2014, Raaschou-Nielsen et al., 2013, Turner et al., 2011). The methods used included categorical modelling and application of smoothing functions. Hamra et al. notes that all of these analyses concluded that there is no evidence of marked deviation from linearity. However, the studies that reported results of categorical modelling did not provide much

support for a linear concentration response, and some of the studies had limited power to detect a departure from linearity. The Canadian study of Hystad et al. reported ORs for linear models and a categorical analysis based on quintiles of the distribution for $PM_{2.5}$. Table 3 shows ORs for the association between lung cancer incidence and $PM_{2.5}$ exposures derived using a national spatiotemporal model. The categorical model results for $PM_{2.5}$ do not suggest linearity with the top four quintiles showing flat response curve. ORs for quintiles of NO_2 (not shown) also show a flat response across the top four quintiles (highly significant ORs between 1.49 and 1.66). Cesaroni et al. also reported the results of a categorical analysis based on quintiles of the distribution for $PM_{2.5}$ for their study of subjects from Rome, Italy (see Table 4). It is notable that the maximum $PM_{2.5}$ value for the Canadian study of 19.6 mg/m^3 was only slightly higher than the lowest quartile of the Italian study (19.4 mg/m^3). The HRs for the top 4 quintiles reported by Cesaroni et al. also do not suggest that log HR is linearly related to $PM_{2.5}$ concentration, especially the flat response for the top three quintiles. Turner et al. also reported results of categorical modelling for non-smokers in the ACS CPS-II study. $PM_{2.5}$ concentrations were categorised using cut-points based on the three quartiles (11.8, 14.3 and 16 µg/m^3), and the 90th percentile (17.9 µg/m^3). Fully adjusted HRs for lung cancer mortality relative to the reference category (less than 11.8 µg/m^3) were presented graphically in Figure 1 of Turner et al. As in the two previous studies, the response is relatively flat across the four categories.

Table 3 Adjusted ORs for the association between lung cancer incidence and $PM_{2.5}$ exposure, as derived from national spatiotemporal models for Canada (Hystad et al., 2013)

Exposure	Cases	Controls	Adjusted OR[1]
All lung (per 10 µg/m^3)	2,154	3,264	1.29 (0.95-1.76)
Q1 (Hamra et al., 2014)	378	718	1.00
Q2 [9.0-10.9]	470	598	1.26 (0.99-1.59)
Q3 [11.0-12.8]	462	619	1.35 (1.06-1.71)
Q4 [12.9-14.7]	445	646	1.39 (1.08-1.79)
Q5 [>14.7]	399	683	1.19 (0.90-1.57)

[1] Unconditional logistic regression model with random effect for census division lived in the longest, adjusted for all individual variables, study province, ecological radon exposure, years living in the lowest quintile of neighbourhood median household income, percent without a high school diploma and percent of households greater than 30 years old dwellings.

Table 4 Adjusted HRs (95% CI) of lung cancer mortality according to $PM_{2.5}$ exposure, Rome 2001–2010 (Cesaroni et al., 2013).

Exposure	Cases	Adjusted HR[1]
All lung (per 10 µg/m^3)	12,208	1.05 (1.01, 1.10)
Q1 [<19.4]	2,090	1.00
Q2 [19.4–22.5]	2,268	1.04 (0.98, 1.10)
Q3 [22.5–24.8]	2,397	1.09 (1.02, 1.15)
Q4 [24.8–26.8]	2,611	1.07 (1.01, 1.13)
Q5 [>26.8]	2,842	1.08 (1.02, 1.15)

[1] Adjusted for sex, marital status, place of birth, education, occupation, and area-based socioeconomic position.

The other studies used smoothing functions or splines to test for departures from non-linearity. In addition to categorical modelling, Cesaroni et al. also estimated natural splines (two degrees of freedom) concentration–response curves for non-accidental mortality, cardiovascular, IHD, and lung-cancer mortality for $PM_{2.5}$ and NO_2 based on a 20% random sample of the study population (Cesaroni et al., 2013). There was little evidence of an association between lung cancer and $PM_{2.5}$ for the 20% sample and it is not surprising that the lung cancer mortality spline response curve for $PM_{2.5}$ also showed no evidence of an association and hence was linear. Raaschou-Nielsen et al. also provided weak evidence that there was no marked deviation from linearity. The investigators tested the linear assumption in the relationship between each air pollutant and lung cancer in each of the 14 sub-studies (Raaschou-Nielsen et al., 2013). The test was performed by replacing the linear term with a natural cubic spline with three equally spaced inner knots, and comparing the model fit of the linear and the spline models by the likelihood-ratio test. However, the ranges of $PM_{2.5}$ and PM_{10} concentrations were very narrow for many centres, and the numbers of lung cancer cases were small for most centres. Consequently, the confidence intervals for HRs for some centres were extremely wide. It is not surprising that departures from linearity were not detected as tests for most centres probably had very limited power to detect a deviation from linearity. Indeed, none of the centres had sufficient power to reject the hypothesis that there was no association between lung cancer and $PM_{2.5}$, and only one centre would have rejected the hypothesis that the OR was 0.67. Figure 2 of Pope et al. shows a nonparametric smoothed exposure response relationship between log RR and lung cancer mortality among ACS CPS-II subjects (Pope et al., 2002). The response curve did not deviate markedly from linearity, but the response curve was much steeper below 12-13 µg/m^3. Lepeule et al. reported that model fit was better without the spline term for all-cause mortality and specific causes of death (p-values between 0.24 and 0.43), indicating a linear relationship with $PM_{2.5}$ (Lepeule et al., 2012). However, the study could not detect an interaction during the follow-up period for lung cancer, even though the RR for a 10 µg/m^3 increase in $PM_{2.5}$ was 2.84 (95% CI 1.06-7.59) for 2001-2009 compared to 1.37 (95% CI 1.07-1.75) over the full follow-up period from 1974-2009. Puett et al. noted that there were no statistically significant deviations from linearity (Puett et al., 2014).

Concentration-response is clearly a complex issue and merits further study, particularly at the low concentrations now observed in industrialized Western European and North American countries. The Health Effects Institute has funded three PM studies in these areas to assess the low-level concentration-response relationship. The few studies have considered alternatives to a linear exposure-

response model have concluded that there is no evidence of a marked deviation from linearity (Hamra et al., 2014), but some had limited information to reach such a conclusion. Even if the linearity assumption may be realistic within the range of data, it doesn't appear to be a realistic assumption for some studies. Likewise, it would be wrong to conclude that the linear model can be used to extrapolate beyond the range of the data. However, it is not clear that "integrated exposure-response" models, which integrate exposures to $PM_{2.5}$ from different combustion types (ambient air pollution, SHS, household solid cooking fuel, and active smoking) are better for extrapolation (Burnett et al., 2014, Pope et al., 2011). In the case of lung cancer risk, the response curve is dominated by risk estimates from epidemiology studies of smokers.

6. CONSISTENCY AND HETEROGENEITY

This section looks at the consistency of the evidence linking exposure to outdoor pollution with lung cancer. Consistency of findings from different studies is very important when judging causality, and the IARC Working Group noted that:

> *"both cohort and case-control studies with exposures assessed in the population setting, involving millions of subjects and many thousands of lung cancer cases in different parts of the world, consistently showed an association between exposure to outdoor pollution and the risk of lung cancer, in both sexes and after adjustment for the main potential confounders".*

A causal interpretation is generally considered to be strengthened when studies of dissimilar populations, exposure characteristics or research methods yield similar measures of effect. However, Phillips and Goodman noted that there are situations where inconsistency can make us more comfortable with a causal interpretation (Phillips and Goodman, 2006). They note that most investigators find it reassuring when the association with an exposure is stronger for histologically-confirmed cancers than it is for a less reliable definition of disease status. This is because we would expect to see a stronger association when there is less measurement error (independent, non-differential). In addition, there are many reasons why one would not expect the evidence to be consistent.

Three characteristics of studies which are often examined in heterogeneity analyses are study populations, research methods and exposure characteristics (assessment methodology, agent and exposure contrast). These aspects of heterogeneity are explored in sections 6.1 to 6.3 for air pollution studies. Population air pollution studies are especially heterogeneous in terms of exposure characteristics, and even the agent may be heterogeneous. Consequently, consistency might not be expected even if there is causality, and explanations for consistency other than causality may be required such as residual confounding. Despite the many reasons for expecting substantial heterogeneity, meta-analyses such as that by Hamra et al. provide relatively little evidence of heterogeneity and section 6.4 looks at the detectability of heterogeneity using standard statistical tests (Hamra et al., 2014). Finally, section 6.5 looks in detail at the heterogeneity of response between studies in Europe and the U.S, and the evidence that some regions show no effect of particulates exposure.

6.1. HETEROGENEITY OF STUDY POPULATION

Heterogeneity of study population is one factor that can contribute to the variations in observed associations between an air pollutant and a human health effect. Although air pollution studies typically control for potential confounding from conditions such as gender, age, education, marital status, and socioeconomic status there are several overlooked characteristics that can dramatically influence lung cancer risk. Chief among these are diet, genetic polymorphisms, and physical activity. There has been considerable interest in the role of diets high in fruit and vegetable as a preventative factor in lung cancer development. In one of many studies, the consumption of 500 g/day of fruits and vegetables was predicted to decrease lung cancer incidence by 0.2% in citizens of Europe (Soerjomataram et al., 2010). In another examination of prospective cohort studies, a meta-analysis revealed that the consumption of fruits and vegetables was inversely related to lung cancer development with the optimal portion being about 2 servings/day (Wang et al., 2015). The mechanism for this protective is thought to involve flavonoids found

in foods and beverages of plant origin. When dietary consumption to these powerful antioxidants was restricted, an increased risk of lung squamous cell carcinoma was shown to exist (Christensen et al., 2012). The relationship between diet and lung cancer susceptibility continues to be an area of great interest and curiosity with recent findings indicating that the consumption of green or black teas has a protective effect on lung cancer development, while the consumption of coffee increases the risk of lung cancer (Wang et al., 2014a, Xie et al., 2016).

Another cause of heterogeneous responses seen in observational studies examining the relationship between air pollutants and lung cancer is the genetic polymorphisms that exist in the genes coding for key metabolic enzymes found in the lung. Although the relationship of these polymorphisms to lung cancer risk has been studied for a large number genes, several have been investigated in great detail (Gresner et al., 2007). These include glutathione-S-transferase (GSTM1) and cytochrome P450 1A1 (CYP1A1) (Hung et al., 2003). GSTM1 is responsible for detoxifying hydrophilic electrophiles including those derived from polycyclic aromatic hydrocarbon epoxides by conjugating them with glutathione. CYP1A1 is a microsomal phase 1 activating enzyme that is capable of converting procarcinogens such as polycyclic aromatic hydrocarbons into active electrophiles. Individuals possessing the null genotype for glutathione-S-transferase (GSTM1) have been associated with an increase in the incidence of lung cancer (Shi et al., 2008). The study however was restricted to Chinese individuals who display a higher prevalence of this genotype (50-60%) compared to Caucasians (10-20%) (Raimondi et al., 2006). Similarly, observational studies with Il2462Val polymorphism for CYP1A1 has been shown to increase microsomal oxidation capacity and lung cancer risk, but the relationship was restricted to meta-analyses performed with Asian populations (Wang et al., 2011). No association was observed in Caucasians, Africans or other ethnic groups. Together, these studies suggest that ethnicity may be an important contributor to response heterogeneity observed in environmental epidemiology studies focusing on lung cancer.

Finally, there is relatively consistent evidence that those individuals who exercise on a regular basis are more resistant to lung cancer. A risk reduction was observed in both Canadian men and women who engaged in moderate to vigorous exercise over a four year period (Mao et al., 2003). The risk reduction was more profound among smokers and those with low and medium body masses. Other studies suggest that 6-8 hours/week of moderate intensity physical activity could appreciably decrease lung cancer incidence in males (Lee et al., 1999). These findings were duplicated in a meta-analysis that pooled the results from 11 observational studies (Tardon et al., 2005). Both men and women were advantaged by the increase in physical activity, with associations being somewhat higher in women.

The preceding examples provide a glimpse at behavioural and genetic factors that can result in considerable heterogeneity when not properly controlled for in a cohort or case-control study. If left unmanaged, substantial heterogeneity and confounding can occur.

6.2. HETEROGENEITY OF RESEARCH METHODS

It was noted in section 2, that there is a wide range of study designs used in air pollution studies. The generic problems with the two main designs of cohort and case-control including bias have been discussed in section 2 and won't be discussed further here. In addition, section 4 has discussed problems with confounding and highlighted the variation between studies in terms of how they controlled confounding by smoking and other potential confounders. Data analysis methods can further contribute to heterogeneity.

6.3. HETEROGENEITY IN EXPOSURE CHARACTERISTICS

As noted in section 3, there is considerable variability in the way that exposure is estimated in different studies. There is variability not only in the methodology used, but also the time period for which the exposure contrast is estimated and its relevance to the BREP. In addition, there are large differences in the level of exposure in different populations and variability across subjects. Even the reporting of effect as an increase in risk per unit of exposure has a huge variation in meaning between studies. For instance, it might refer to exposure at start of follow up at the address where a subject lived then (ACS-CPII), exposure in a short window before lung cancer occurred at the baseline address (H6C), or exposure at the baseline address at end of follow up (ESCAPE). Given that particulate levels fell considerably over the follow up period of many studies, there is clearly a wide variation even in the definition of a unit of particulate exposure.

A less appreciated factor is the variation in pollutant composition. Airborne particulate is composed of a heterogenous mixture of solids and liquids that are continually undergoing change as the air mass ages. This change chemical composition is the result atmospheric photochemical reactions that are mediated by hydroxyl radicals (Robinson et al., 2006). When exposed to sunlight, particle absorbed PAHs with 2 or 3 rings were shown to disappear rapidly at a bi-exponential rate (Kim et al., 2009). The first phase of this loss process typically had a half-life of less than 5 hours, whereas the half-life of second phase was generally 14–50 hours. The degradation of PAHs with 4 to 6 rings was, in contrast, mono-exponential and occurred at a far slower rate. Likewise, other studies have shown that the bulk chemical composition of particulates can vary greatly for different urban areas around the world. These variations are related in part to climatic conditions, local emission sources, and the oxidant status of the regional atmosphere.

Particulate matter contains varying quantities organic carbon, elemental carbon, sulfate, nitrate, ammonium, chloride, crustal minerals, and biological materials (Harrison and Yin, 2000). The sulfate content of particulate matter is directly impacted by local SO_2 emissions, whereas the ammonium levels are more influenced by regional agricultural practices. $PM_{2.5}$ collected from western regions of the United States have been shown to contain far lower sulfate levels and higher amounts of elemental and organic carbon than $PM_{2.5}$ from eastern locales. Perhaps of greatest interest are variations in the trace metal content of particulate matter since a recent study has shown an association of PM_{10} and $PM_{2.5}$ metal content with the risk of lung cancer. Using the information from 14 cohort studies performed in eight European countries, statistical significant associations were observed with $PM_{2.5}$ copper (HR 1.25, 95% CI: 1.01-1.53 per 5 ng/m^3), PM_{10} zinc (HR 1.28, 95% CI: 1.02-1.59 per 20 ng/m^3), PM_{10} sulfur (HR 1.58, 95% CI: 1.03-2.44 per 200 ng/m^3), PM_{10} nickel (HR 1.59, 95% CI: 1.12-2.26 per 2 ng/m^3), and PM_{10} potassium (HR 1.25, 95% CI: 1.02-1.33 per 100 ng/m^3) (Raaschou-Nielsen et al., 2016). These results however, were only obtained when the analysis was restricted to residents who did not change residence during the follow-up. When all members of the cohort were included in the analysis, the associations were no longer significant. Land use regression models were used to estimate exposure concentration at each subject's residential address. Confounding because of gender, calendar year, age, smoking status, smoking intensity, smoking duration, environmental tobacco smoke, occupation, fruit intake, marital status, education level, employment status, and socio-economic status were all taken into consideration.

Whereas the study by Raaschou-Nielsen et al. is the only examination of the associations between particle matter constituents and lung cancer development, other studies of the relationship to pulmonary mortality yield somewhat conflicting

results. For instance, Heo et al., failed to see an association between short-term exposure to PM-related copper, zinc, and bromine and pulmonary mortality using a multi-pollutant model (Heo et al., 2014). A three day lagged exposure to the lead in particulate matter was, however, associated with an increase in pulmonary mortality in a multi-pollutant model. Likewise Ostro et al. noted an association between the silicon from particulate matter and pulmonary and chronic pulmonary mortality, but the association was not significant for iron, potassium, and zinc using two-pollutant models (Ostro et al., 2010). Some of the heterogeneity in these study results is likely attributable to the variable trace metal content in particulate matter. A recent study of the variance in trace metal content of PM_{10} and $PM_{2.5}$ for twenty study locations in Europe found that the concentration of copper, iron, potassium nickel, sulphur, silicon, vanadium, and zinc could vary greatly both within and between study locations (Tsai et al., 2015). In fact, the variance in PM_{10} copper, iron, and zinc content far exceeded the variation observed PM_{10}, $PM_{2.5}$, and NO_2.

Clearly particulate matter varies considerably in chemical speciation between locations and hence we might not expect to see the same effects per unit of particulate exposure in studies from different locations. In addition, the effect of exposure to $PM_{2.5}$ might be expected to be more variable than exposure to PM_{10} as it includes a higher proportion of mutagenic species, many of which are products of combustion, and the smaller particles penetrate more deeply into the lung and are more likely to be retained, whilst the coarse fraction of PM_{10} consists mainly of minerals and biological materials (Hamra et al., 2014). Less heterogeneity might also be expected to be seen in the results of studies where speciation is not an issue e.g. studies of ozone, SO_2 and NO_2. However, there is still likely to be considerable variation in other factors such as the study population and exposure methodology, and heterogeneity may also have resulted from residual confounding from factors such as smoking, occupation, and SES (see section 4). In addition, NO_2 is usually regarded as reflecting traffic and local combustion sources, while SO_2 reflects power plant and industrial source emissions; so these pollutants are not treated as risk factors in their own right. Section 8 considers further whether the associations with lung cancer seen in air pollution studies for $PM_{2.5}$, PM_{10} and NO_2 can be considered as indicating a causal agent, or whether the agents are acting as a surrogate for the outdoor air pollution mixture.

6.4. IS HETEROGENEITY DETECTABLE

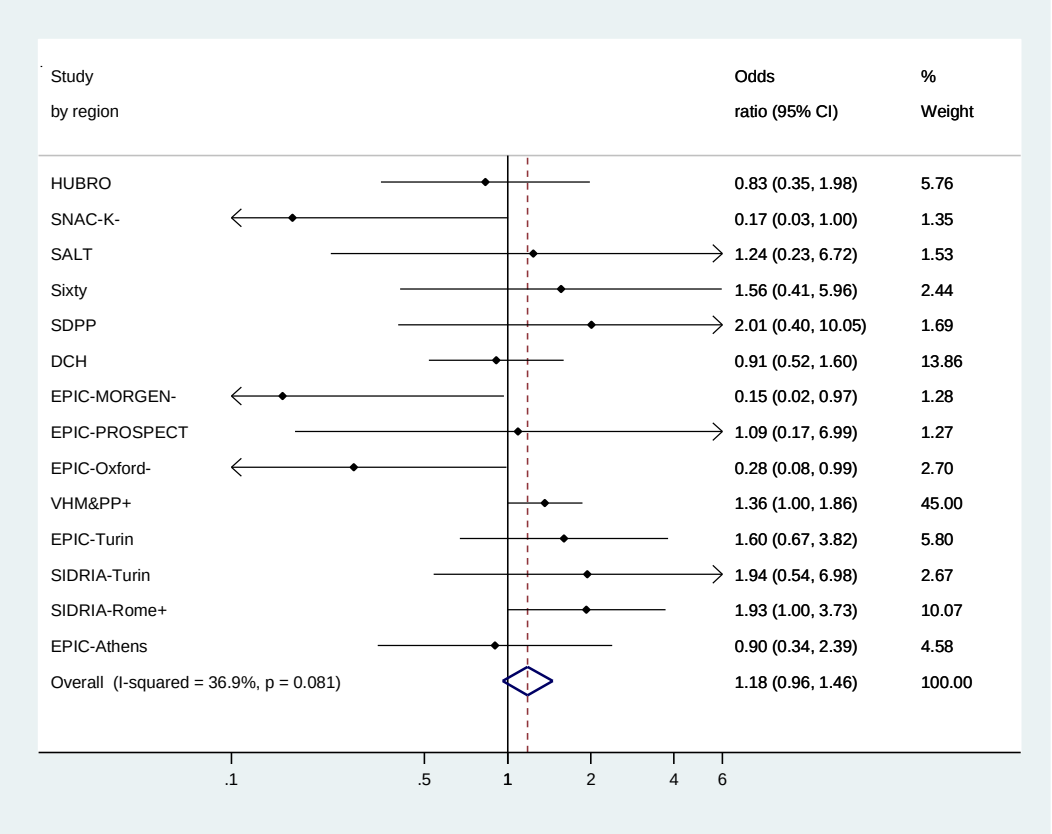
As noted earlier, consistency of findings from different studies is very important when judging causality. Consistency is one of the widely used Bradford Hill criteria for causation, Hill notes that consistent findings observed by different persons in different places circumstances, and times strengthens the likelihood of an effect (Hill, 1965). Consistency or the absence of significant heterogeneity is also important when judging how widely the results from a meta-analysis can be generalised to other populations. For instance, Raaschou-Nielsen et al. concluded that the absence of significant heterogeneity in their study (effectively a meta-analysis of 17 sub studies) meant that the results could be generalised to all European populations. In addition, testing for heterogeneity is important when assessing an effect modification (Raaschou-Nielsen et al., 2013).

Raaschou-Nielsen et al. and Hamra et al. used random-effects models to pool the results for different studies/cohorts. I^2 statistics and p values for the Chi-squared test from Cochran's Q were calculated to investigate the heterogeneity among cohort-specific effect estimates (Hamra et al., 2014, Raaschou-Nielsen et al., 2013). Inconsistencies are difficult to identify as statistically significant because tests of heterogeneity such as Cochran's Q are usually insensitive (Higgins et al., 2003, Rothman et al., 2008), and a non-significant result must not be taken as

evidence of no heterogeneity (Higgins and Green, 2011). For this reason, it is common practice to use a p value of 0.01, but the test still has low power. I^2 describes the percentage of total variation across a study that is due to heterogeneity rather than sampling error. It measures inconsistency (overlap of confidence intervals), not the amount of heterogeneity, and thresholds for the interpretation of I^2 can be misleading since the importance of inconsistency depends on several factors (Higgins and Green, 2011, Rucker et al., 2008). I^2 will increase with increasing numbers of subjects per study even if the heterogeneity variance is kept fixed.

The ESCAPE study illustrates the difficulty of identifying heterogeneity when the risk estimates have large sampling variability (Raaschou-Nielsen et al., 2013). Cochran's Q test had very low values for both $PM_{2.5}$ and PM_{10}, but especially for $PM_{2.5}$. For $PM_{2.5}$, the Q test value was 6.6 (based on the reported p value), and needed to exceed 13 for I^2 to be non-zero and 22.36 to be statistically significant at a 0.05 level. The value is surprisingly low given the obvious sources of heterogeneity listed elsewhere in this report. A considerable amount of heterogeneity can be introduced without obtaining a significant test result for heterogeneity. Figure 9 shows a modified version of the $PM_{2.5}$ results for the ESCAPE study but with the same meta-HR and 95% CI. However, the HRs of two centres (VHM&PP, SIDRIA-Rome) have been increased until they became statistically significant, and the HRs of three centres (SNAC-K, EPIC-MORGEN, EPIC-Oxford) have been reduced until they became statistically significant. The widths of the confidence intervals on the log scale for the HRs of the five centres i.e. the sampling variability, were not changed. Despite the obvious heterogeneity, the I^2 value was only 36.9% and Cochran's Q test had a non-significant p-value of 0.081.

Figure 9 Modified version of $PM_{2.5}$ meta-analysis shown in Figure 3 of Raaschou-Nielsen et al (2013) with increased heterogeneity.



\+ Centres with increased HR
\- Centres with decreased HR

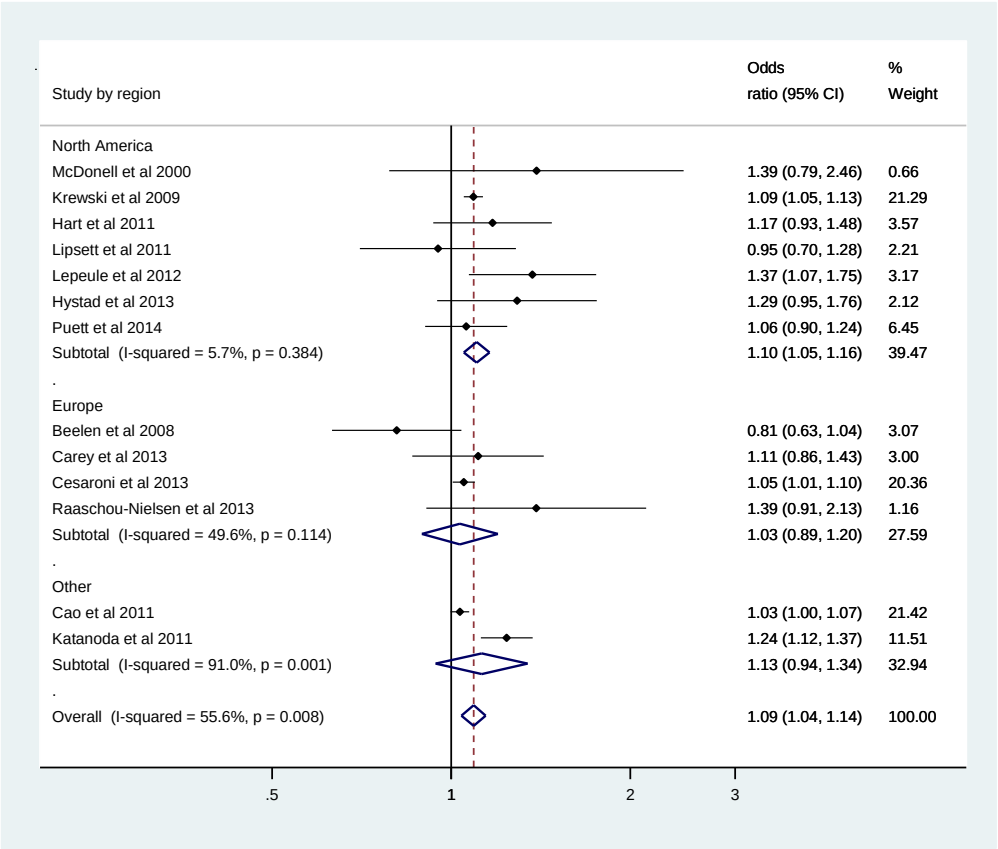
6.5. WHAT EVIDENCE IS THERE OF REGIONAL HETEROGENEITY

Hamra et al. calculated region-specific (Europe, North America, and other continents) meta-estimates for the association between lung cancer and $PM_{2.5}/PM_{10}$ (Hamra et al., 2014). The $PM_{2.5}$ meta-estimates for North America, Europe, and other continents were 1.11 (95% CI: 1.05-1.16), 1.03 (95% CI: 0.89-1.20), and 1.13 (95% CI: 0.94-1.34) per 10 µg/m^3 $PM_{2.5}$, respectively. Hamra et al. noted that the confidence intervals were largely overlapping, and homogeneity tests suggested no evidence of differences across regions, but there was some evidence of heterogeneity across all centres (I^2=53.0%, p=0.01), and across centres from Europe (I^2=50.0%, p=0.112) and other continents (I^2=91.0%, p=0.001). However, Table 2 of Hamra et al. reported different I^2 and p values for North America (I^2=6.5%, p=0.378) and the overall analysis (I^2=56.4%, p=0.007). The meta-analysis included risk ratios for $PM_{2.5}$ and PM_{10} from the ESCAPE study by Raaschou-Nielsen et al. which are themselves meta-HR based on risk estimates from what are effectively 14 separate studies (Raaschou-Nielsen et al., 2013). Consequently, it can be argued that the HRs from the 14 ESCAPE cohorts should have been included in the meta-analysis. The meta-analysis shown in Figure 1 of Hamra et al. was repeated including the 14 separate HRs from the ESCAPE cohorts instead of the meta-HR result (giving 26 study results for $PM_{2.5}$, and 22 results for PM_{10}). The meta-analysis of Hamra et al. was also repeated as it is not possible to replicate it exactly because it is not known what assumptions were made to deal with non-symmetric confidence intervals when estimating the standard errors of risk estimates. The results of the original and

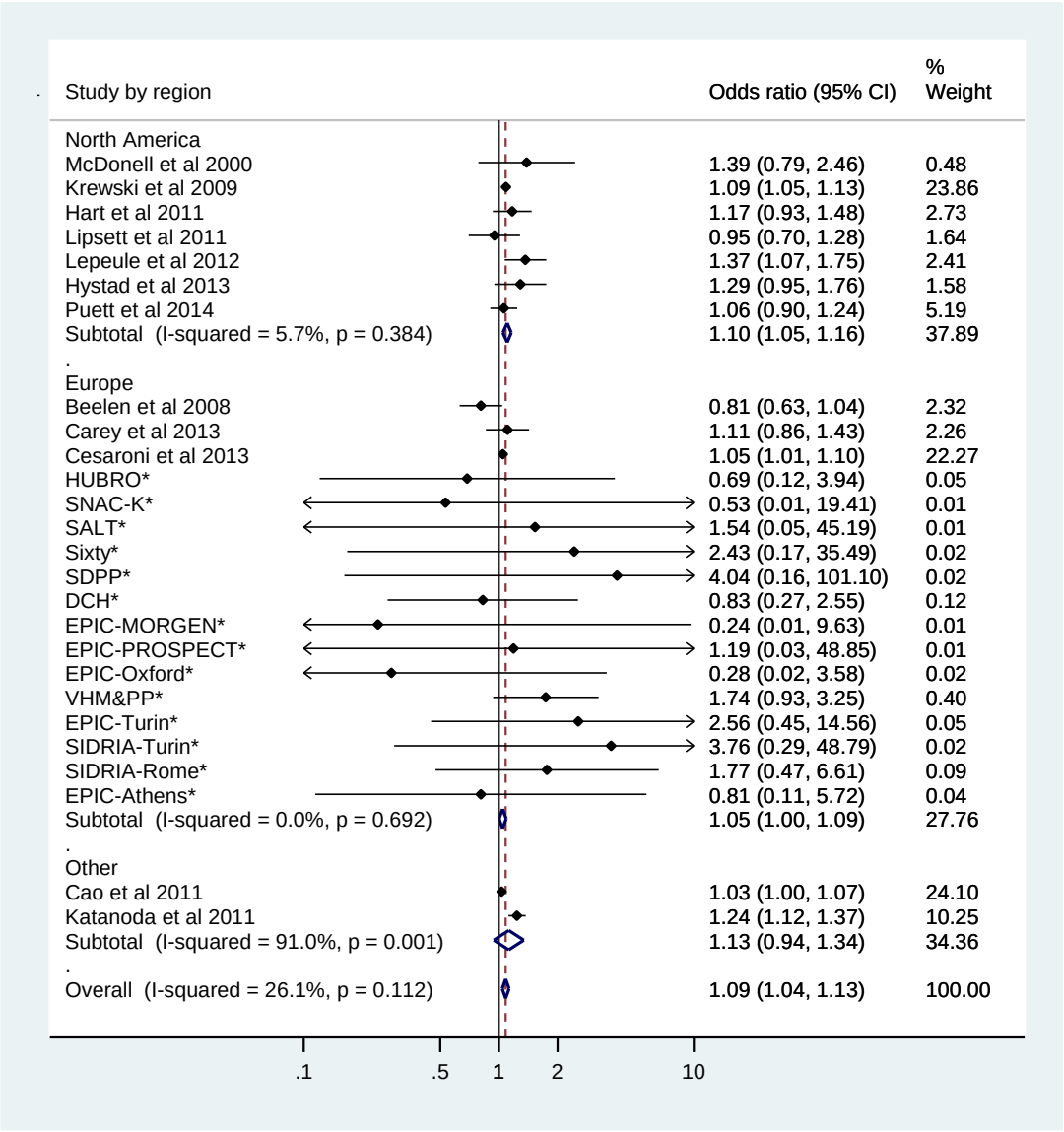
extended meta-analysis are shown in Figure 10. The individual ESCAPE centres have very little weight in the meta-analysis, but their low Cochran's Q test result dilutes the test for heterogeneity among European centres reducing I^2 to zero. This effectively turns the random effects meta-analysis for European centres into a fixed effects analysis and results in a much tighter confidence interval for the meta-risk estimate for Europe. The European analysis is dominated by Cesaroni et al., although this study has no information on the key confounder (i.e. smoking habits) (Cesaroni et al., 2013). The overall risk estimate and confidence interval were not changed substantially.

Figure 10 Estimates of lung cancer risk associated with a 10 µg/m^3 change in exposure to $PM_{2.5}$ including meta-HR result from ESCAPE study (A) or 14 individual HRs from ESCAPE study (B)

A.



B.

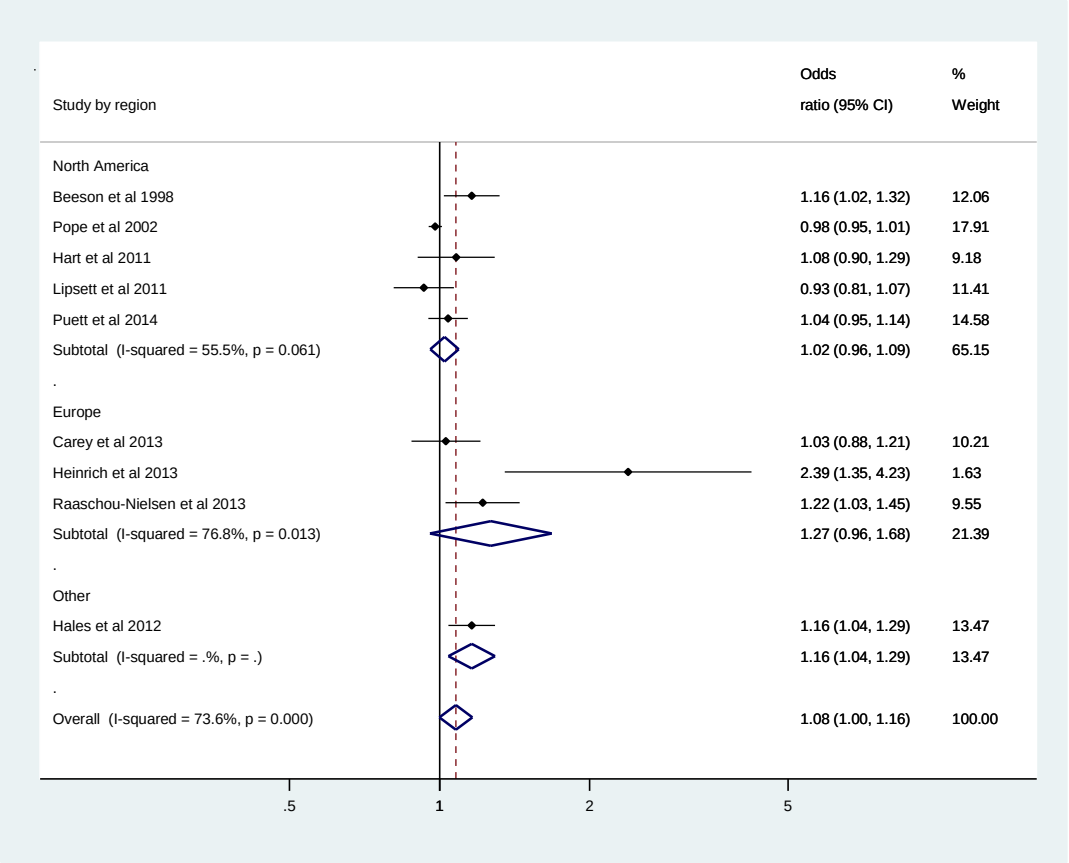


* ESCAPE cohorts (Raaschou-Nielsen et al., 2013)

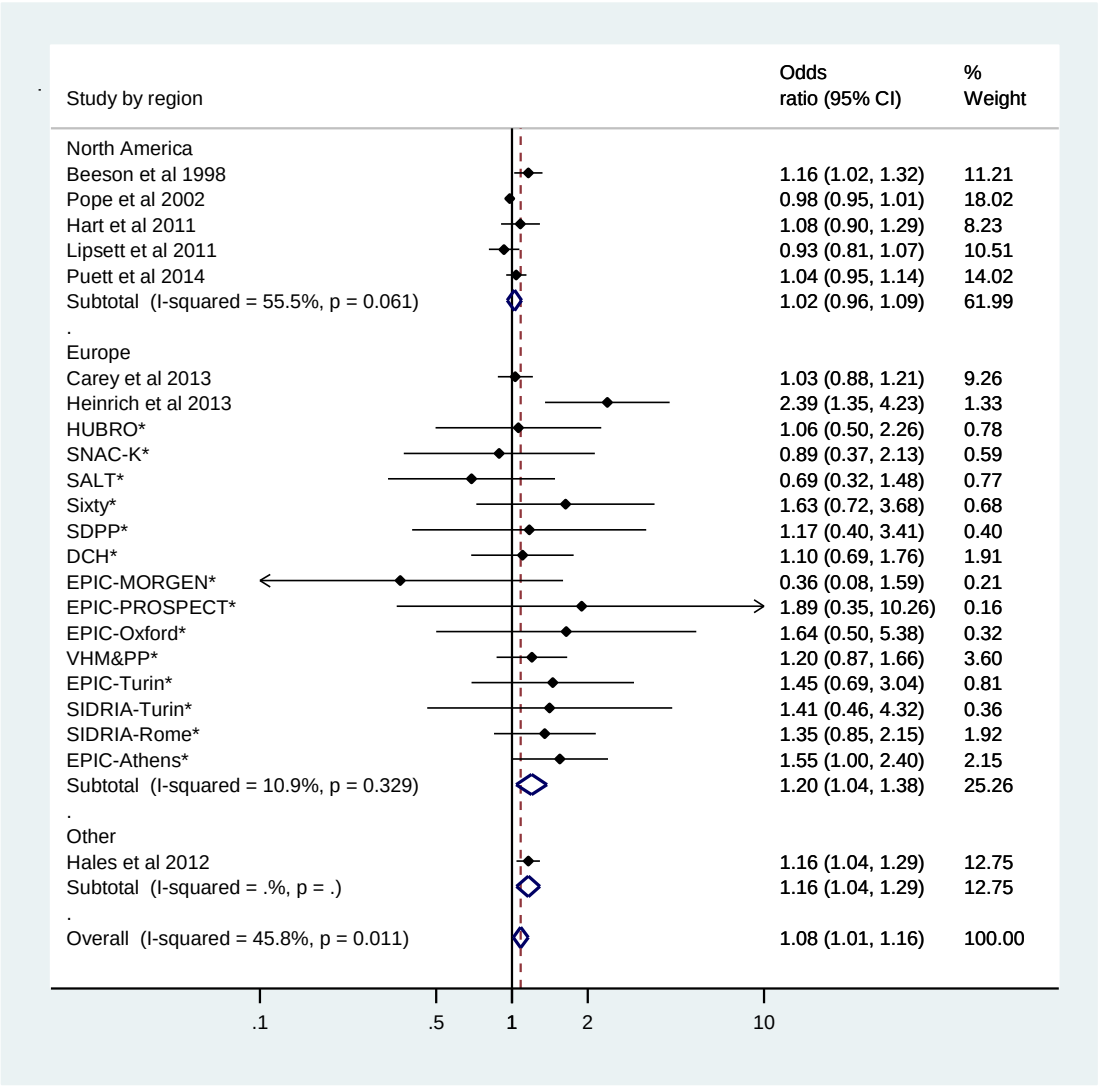
For PM_{10}, Hamra et al. reported meta-estimates for Europe and North America of 1.27 (95% CI: 0.96-1.68) and 1.02 (95% CI: 0.96-1.09) per 10 µg/m^3 PM_{10}, respectively. However, there was only one study result available from outside of Europe and North America (Hamra et al., 2014). There was evidence of heterogeneity across all centres (I^2=74.6%, p=0.000), and across centres from Europe (I^2=57.7%, p=0.051) and North America (I^2=76.5%, p=0.014). Including all 14 ESCAPE HRs in the analysis again considerably reduced the I^2 value for Europe from 76.8% to 10.9% and considerably tightened the confidence interval of the meta-risk estimate for Europe (see Figure 11).

Figure 11 Estimates of lung cancer risk associated with a 10 µg/m^3 change in exposure to PM_{10} including meta-HR result from ESCAPE study (A) or 14 individual HRs from ESCAPE study (B)

A.



B.



* ESCAPE cohorts (Raaschou-Nielsen et al., 2013)

The ESCAPE study also provides other indications of heterogeneity of response across European centres (Raaschou-Nielsen et al., 2013). Although the formal tests of heterogeneity were not statistically significant for both $PM_{2.5}$ and PM_{10}, Morfeld et al. argued that there was only evidence of an association between lung cancer and particulates exposure for the five centres from the most southerly countries (Austria, Italy, and Greece) (Morfeld et al., 2013). For these centres, Morfeld et al. calculated a meta-hazard ratio (meta-HR) of 1.33 (95% CI 1.03–1.71, p=0.028) per 5 $\mu g/m^3$ $PM_{2.5}$, and 1.34 (95% CI 1.08–1.65, p=0.008) per 10 $\mu g/m^3$ PM_{10}. In contrast, meta-HR of 0.92 (95% CI 0.63–1.33, p=0.65) per 5 $\mu g/m^3$ $PM_{2.5}$ and 1.06 (95% CI 0.80–1.39, p=0.69) per 10 $\mu g/m^3$ PM_{10} were calculated for the centres from the most northerly countries (Sweden, Norway, Denmark, The Netherlands, UK). A meta-regression estimated the relative risks (north vs. south) as 0.79, p = 0.21, for PM_{10} and 0.69, p=0.13, for $PM_{2.5}$.

In the US, the ACS CPS-II study provides some evidence of an east/west difference in response (Krewski et al., 2009). The nationwide analysis gave an HR of 1.09 (95% CI: 1.03-1.15) per 10 $\mu g/m^3$ $PM_{2.5}$ (random effects model with 44 individual covariates and seven ecological covariates). However, intra-urban analyses

performed for New York City and Los Angeles gave contrasting results. In Los Angeles the HR for lung cancer was 1.39 (95% CI: 0.96-2.01) per 10 µg/m^3 $PM_{2.5}$ (LUR exposure, 44 individual covariates model). In contrast, there was an almost opposite effect in New York City with an HR of 0.74 (95% CI: 0.30-1.79) per 10 µg/m^3 $PM_{2.5}$ (LUR model for average 1999-2001 exposure with 44 individual covariates). However, the range of $PM_{2.5}$ concentrations for New York City was very narrow with an interdecile range of 1.5 µg/m^3. Jerrett et al. also analysed a subset of data from California residents in the ACS CPS-II study and reported an HR of 1.12 (95% CI: 0.92-1.37) per 10 µg/m^3 $PM_{2.5}$. The interdecile range of $PM_{2.5}$ concentrations for Californian residents was 8.97 µg/m^3 (Jerrett et al., 2013). However, other regional studies of lung cancer and $PM_{2.5}$ from the U.S. do not support this east/west difference in response. A non-significantly elevated risk ratio was reported by one Californian study, the Adventist Health Study on Smog (AHSMOG), but not by another Californian study, CTS, (Lipsett et al., 2011, McDonnell et al., 2000). Nevertheless, there is evidence of heterogeneity of response between studies in Europe and the U.S. and the reasons for these differences require more investigation as suggested by Morfeld et al. (Morfeld et al., 2013).

7. OBSERVABLE CONSEQUENCES AND BIOLOGICAL COHERENCE

IARC evaluations of carcinogenic risk typically involve an examination of mechanistic data. This process includes a thorough inspection of any and all information on the physiological, functional, molecular, and kinetic changes that accompany the onset of an adverse health effect. This compilation of data helps establish biological plausibility, which is one of the nine Bradford Hill criteria used to establish causality. Weed and Gorelic noted over twenty years ago that coherence, another of Hill's causal perspectives, was rarely considered in assessments of causality: most likely because reviewers may have equated it with biological plausibility, thus making it redundant (Hill, 1965, Weed and Gorelic, 1996). However, Weed and Gorelic noted that Hill had distinguished these considerations on logical grounds with an association being plausible if it is consistent with current biological knowledge, and coherent if it does "not seriously conflict with the generally known facts of the natural history and biology of the disease" (Hill, 1965, Weed and Gorelic, 1996). More recently it has been noted that the distinction between biological plausibility and biological coherence is a fine one, even for the examples listed by Hill (Rothman et al., 2008) . Nevertheless, it is usually not difficult to distinguish between plausibility and coherence considerations as coherence is based on an assumption that there is causality and hence not dependent on biological plausibility. The distinction between plausibility and coherence is well illustrated by the difference in lung cancer incidence between men and women which is one of the examples listed by Hill in reference to the association between smoking and lung cancer. In this case, the consideration is if smoking causes lung cancer, does the pattern of lung cancer in men and women conflict with that theory? In a recent survey of frameworks for best practices in weight-of-evidence analyses, Rhomberg et al. essentially identified coherence as one of six key general guidelines for integrating evidence (Rhomberg et al., 2013). Their recommendations included asking the question, if the proposed causative processes and modes of action were true, what other observable consequences should be expected, and then checking if these other manifestations of the hypothesized process are in fact seen in available data and do they act as expected (Rhomberg et al., 2013).

This section looks at some observable consequences that might be expected to be observed in studies of outdoor air pollution. For instance, would one expect to observe a concentration response relationship between estimates of exposure to particulates and lung cancer for smokers when the intake of particulates from outdoor air pollution is tiny compared to the dose from cigarette smoking? Other coherence issues that are considered relate to the findings for adenocarcinoma versus those for squamous and small cell lung cancer, and the strength of associations observed for men versus women. Another aspect considered in this section is whether the associations observed between chronic obstructive pulmonary disease (COPD) or non-malignant respiratory disease (NMRD) and air pollution are coherent with those for lung cancer. Finally, subsection 7.5 looks at whether the results seen in occupational studies are coherent with those from population studies.

7.1. ARE THE EFFECTS SEEN IN CURRENT, FORMER SMOKERS AND NEVER SMOKERS COHERENT

Pope et al. conducted an evaluation of the shape of the $PM_{2.5}$-mortality exposure-response relationship for lung cancer (Pope et al., 2011). The investigators derived estimates of adjusted relative risks (RRs) over different increments of active cigarette smoking for the primary analytic cohort of the ACS CPS-II study of almost

800,000 subjects (540,000 had $PM_{2.5}$ exposure estimates for at least one-time period). Comparative estimates of excess risk of lung cancer from long-term exposure to $PM_{2.5}$ were taken from the ACS CPS-II and H6C studies, and comparative estimates of excess risk of lung cancer from SHS at home and work were also included. For active smoking, the average inhaled dose was assumed to be 12 mg $PM_{2.5}$ per cigarette. Risk and dose estimates (average daily inhaled dose of $PM_{2.5}$) for different increments of active smoking, SHS, and ambient $PM_{2.5}$ were used to fit a simple power function of the form [$RR = 1 + \alpha(dose)^{\beta}$]. The active smoking and SHS RRs were well fitted by the model $RR = 1 + 0.3195(dose)^{0.7433}$, where dose is the daily exposure of $PM_{2.5}$ (mg). However, the exposure response curve suggests that among smokers, the additional exposure to particulates in air pollution would have a minute effect. The mean $PM_{2.5}$ level at the start (1979-1983) of the ACS CPS-II study was 21.1 $\mu g/m^3$, and had fallen to 14.0 $\mu g/m^3$ by 1999-2000 (Pope et al., 2002). However, 21.1 $\mu g/m^3$ only equates to an average daily dose of inhaled $PM_{2.5}$ of 0.38 mg (assuming the same inhalation rate of 18 m^3/day as used by the authors of the study), whereas a 20 cigarette a day smoker inhales 240 mg of particulates each day. The model derived by Pope et al. predicts that the lung cancer RR for this smoking rate will only increase from 19.78 to 19.80 when additionally exposed to 21.1 $\mu g/m^3$ $PM_{2.5}$, assuming that the particulates in air pollution have the same toxicity as those in cigarette smoke (Pope et al., 2011). By contrast, the RR of a never smoker exposed to the same level of $PM_{2.5}$ is estimated by the model to be 1.15.

Although the model of Pope et al. predicts that the risk of lung cancer in a smoker will increase almost imperceptibly when exposed to particulates in air pollution, there is evidence of an association between lung cancer and air pollution in most studies (Pope et al., 2011). It was noted in Section 4.2.3 that Hamra et al. conducted meta-analyses of 18 studies examining the relationship of exposure to $PM_{2.5}$ and PM_{10} with lung cancer incidence and mortality, including an analysis by smoking status (Hamra et al., 2014). Hamra et al. reported that lung cancer risk associated with a 10 $\mu g/m^3$ change in $PM_{2.5}$ was greatest for former smokers (meta-RR 1.44; 95% CI: 1.04-2.01), followed by never-smokers (meta-RR 1.18; 95% CI: 1.00-1.39), and then current smokers (meta-RR 1.06; 95% CI: 0.97-1.15). However, it was also noted in Section 4.2.3. that relevant information was also available from TPCS and if these results are included, then the lung cancer risk associated with $PM_{2.5}$ remains the highest for former smokers (meta-RR 1.33; 95% CI: 1.04-1.69), the meta-RR for never-smokers (1.17; 95% CI: 1.05-1.30) is little changed, but the meta-RR for current smokers (1.19; 95% CI: 1.01-1.40) increases to a similar value to that of never-smokers. The meta-RR of 1.19 for current smokers is clearly not coherent with the lung cancer risk predicted by the model of Pope et al. which is only 1.005 for a 10 $\mu g/m^3$ change in $PM_{2.5}$ for smokers consuming an average of 1.5 cigarettes a day (the lowest daily consumption rate in the study) and much lower for a typical daily consumption rate (Pope et al., 2011). In addition, a study of Canadian women published after the IARC evaluation has reported a significantly elevated risk of lung cancer per 10 $\mu g/m^3$ change in $PM_{2.5}$ for ever smokers (adjusted HR=1.40; 95% CI: 1.12-1.73), but no increased risk for never smokers (adjusted HR=1.01; 95% CI: 0.56-1.80) (Tomczak et al., 2016). For never smokers, the model of Pope et al. predicts an RR of 1.09 for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ level, assuming that the particulates in air pollution have the same toxicity as those in cigarette smoke (Pope et al., 2011). Consequently, the meta-RR of 1.17 for never smokers seems surprisingly high given, the many different forms of error involved in estimating exposure and the attenuation that would occur. The model doesn't predict what will happen for former smokers and much more complex models may be needed (Vlaanderen et al., 2014). Nevertheless, it is difficult to imagine a plausible scenario in which the strongest association between lung cancer and $PM_{2.5}$ exposure would be observed for ex-smokers.

7.2. ARE THE EFFECTS SEEN FOR ADENOCARCINOMA AND SQUAMOUS/SMALL CELL CARCINOMA COHERENT

In section 4.2.4 we have noted that adenocarcinoma is the most common lung cancer histological subtype in never smokers, but there is also plenty of evidence that adenocarcinoma is strongly linked to smoking. It has also been noted that a higher proportion of people with adenocarcinoma than squamous or small cell carcinoma will be never smokers. For example, the proportion of adenocarcinoma cases that were never smokers in one large case-control study was 13.0% versus 2.9% of squamous and small cell carcinoma cases (Pesch et al., 2012). Consequently, some improved ability to detect an association with air pollution might be expected if analysis is restricted to adenocarcinoma.

Substantial increases in the incidence rate of lung adenocarcinoma have been reported during the last several decades, and Devesa et al. reported that through 1997, the incidence of lung adenocarcinoma increased in virtually all areas of the world, with the increases among men exceeding 50% in many parts of Europe (Devesa et al., 2005). The shift to low-tar filter cigarettes has been hypothesised as a cause of the relative increase in incidence rates of adenocarcinomas and decrease in squamous-cell carcinomas of the lung in the USA because the smoke has a lower content of polycyclic aromatic hydrocarbons, which are thought to be associated with squamous-cell carcinoma, and a higher content of nitrates and toxic agents formed from NO_x such as nitrosamines, which are associated with adenocarcinomas. In addition, the deeper inhalation of the smoke results in the transport more distally toward the bronchoalveolar junction where adenocarcinomas often arise (Chen et al., 2007, Raaschou-Nielsen et al., 2013). Air pollution may also be a cause, and studies of time trends and geographical correlations have suggested that adenocarcinoma of the lung may be associated with NO_x emissions (Chen et al., 2007, Chen et al., 2009). However, the evidence of an association with NO_x is inconsistent and Raaschou-Nielsen et al. reported significant associations between NO_x emissions and the incidence of both small cell carcinoma and squamous cell carcinoma, but not adenocarcinoma of the lung (Raaschou-Nielsen et al., 2010).

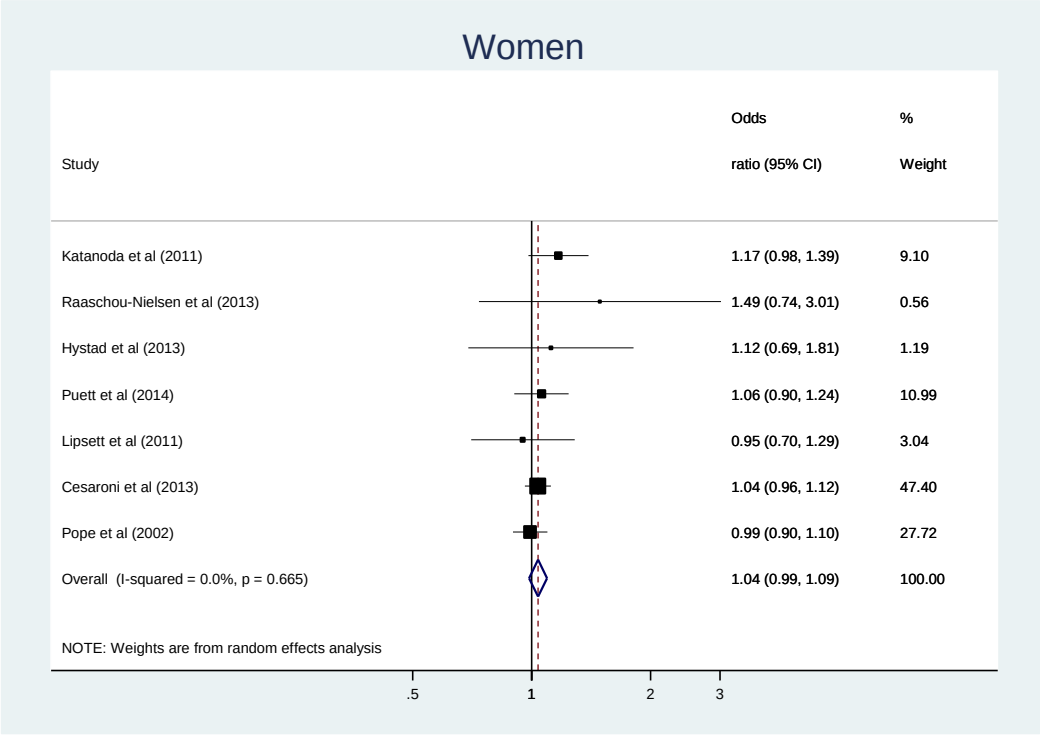
There is little evidence in the literature to assess whether adenocarcinoma is more strongly associated with particulate exposure. At the time of the IARC evaluation, results were available from two studies. The meta-analysis by Hamra et al. also includes results form a later study by Puett et al., but isn't very informative (Hamra et al., 2014, Puett et al., 2014). In the ESCAPE study, Raaschou-Nielsen et al. reported a much stronger association for adenocarcinoma (HR=1.51 per 10 µg/m^3; 95% CI 1.10-2.08) than squamous cell carcinoma (HR=0.84 per 10 µg/m^3; 95% CI 0.50-1.40) for PM_{10}, but for $PM_{2.5}$, the associations were similar for adenocarcinoma (HR=1.55 per 5 µg/m^3; 95% CI 1.05-2.29) and squamous cell carcinoma (HR=1.46 per 5 µg/m^3; 95% CI 0.43-4.90) (Chen et al., 2007, Raaschou-Nielsen et al., 2013). However, Hystad et al. reported similar associations with $PM_{2.5}$ in the CNECSS study when the analysis was restricted to adenocarcinoma (OR=1.27 per 10 µg/m^3; 95% CI 0.84-1.76) as for all lung cancers (OR=1.29 per 10 µg/m^3; 95% CI 0.95-1.76) (Chen et al., 2007, Raaschou-Nielsen et al., 2013). More recently, Puett et al. reported stronger associations with both $PM_{2.5}$ and PM_{10} in the NHS study when analysis was restricted to adenocarcinoma, and Tomczak et al. reported an increase in the HR for $PM_{2.5}$ exposure when analysis was restricted to adenocarcinoma (Hamra et al., 2014, Puett et al., 2014). Overall, the limited evidence suggests that associations are stronger for adenocarcinoma than for other types of lung cancer, and that the findings for adenocarcinoma and other types of lung cancer are coherent.

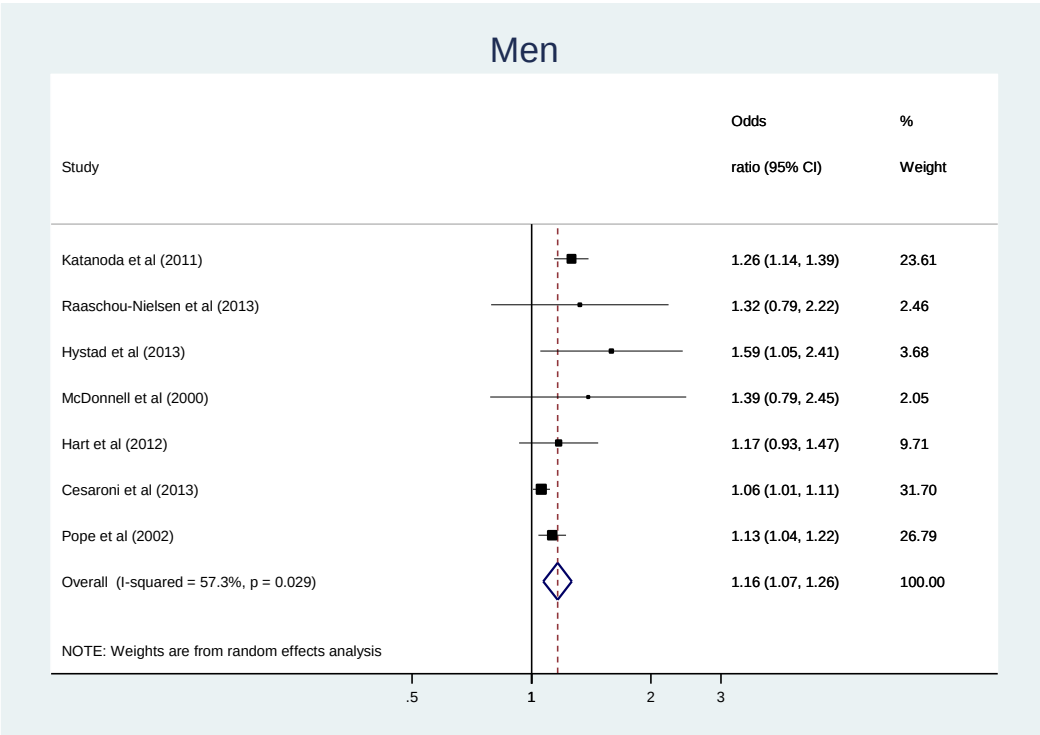
7.3. ARE THE EFFECTS SEEN IN MEN AND WOMEN COHERENT?

There are a number of reasons why the association between lung cancer and particulates exposure might be expected to be different for men and women. Women are less likely to smoke than men and Pesch et al. reported that 24.2% of lung cancers in women occur in never smokers compared to 2.1% for men (Pesch et al., 2012). In addition, women are less likely to have been exposed to lung carcinogens or dust and fumes at work as was observed by Krewski et al. in a reanalysis of the H6C and ACS CPS-II studies (see section 4.3) (Krewski et al., 2000). The time spent by women at the address where exposure has been assessed is also likely to be higher, especially historically. Hence, it seems more likely that if air pollution causes lung cancer, then an association between lung cancer and particulates exposure will be observed for women rather than men.

Hamra et al. recently conducted meta-analyses of 18 studies examining the relationship of exposure to $PM_{2.5}$ and PM_{10} with lung cancer incidence and mortality, but did not include an analysis by gender (Hamra et al., 2014). However, risk estimates available from the studies identified by Hamra et al. have been used to derive meta-RR for men and women using the same meta-analysis methodology (see Figure 12). Five studies reported associations between lung cancer and $PM_{2.5}$ separately for men and women, two studies of male subjects and two studies of females also reported the association. The findings do not follow the pattern expected as the summary lung cancer risk associated with 10 $\mu g/m^3$ $PM_{2.5}$ is not significantly elevated for women (meta-RR 1.04; 95% CI: 0.99-1.09), but is higher and significantly elevated for males (meta-RR 1.16; 95% CI: 1.07-1.26). The risk estimates for women show little evidence of heterogeneity (I^2 = 0%, p=0.67), but there is evidence of heterogeneity for men (I^2 = 57.3%, p=0.029). However, the composition of the groups of males probably varies more due to variation in the prevalence of smoking: the group studied by McDonnell et al. are Seventh Day Adventists who would be unlikely to smoke (Katanoda et al., 2011, McDonnell et al., 2000). Overall, the observed lung cancer risks for men and women are not coherent with our understanding of exposures among men and women.

Figure 12 Estimates of lung cancer risk associated a 10-µg/m^3 change in exposure to $PM_{2.5}$ for women and men. Weights represent the contribution of each study effect estimate to the overall meta-estimate.





7.4. ARE EFFECTS COHERENT FOR DIFFERENT ENDPOINTS

Burnett et al. developed an integrated exposure–response model to estimate the burden of disease attributable to long-term exposure to $PM_{2.5}$ by integrating available relative risk information from studies of ambient air pollution, SHS, household solid cooking fuel, and active smoking (Burnett et al., 2014). This approach was adopted because direct evidence to identify the shape of the mortality exposure response functions was not available at the high ambient concentrations of $PM_{2.5}$ observed in many places in the world. Burnett et al. made the assumption that the observed RRs from epidemiology studies of ambient air pollution, SHS, household solid cooking fuel, and active smoking are a function of $PM_{2.5}$ mass inhaled concentration across all combustion particle sources. The toxicity of $PM_{2.5}$ with respect to lung cancer, ischemic heart disease, stroke, and COPD was assumed to differ only with regard to inhaled mass and not with $PM_{2.5}$ composition. However, the authors acknowledged that the toxicity of emissions from different combustion sources may differ, but current knowledge did not allow definitive and quantifiable conclusions to be made regarding their relative toxicity. A similar approach was suggested by Pope et al. to provide insight into the shape of the cardiovascular and lung cancer exposure response relationship over a much wider range of exposures (Pope et al., 2011). Pope et al. integrated epidemiologic evidence from ambient air pollution, SHS, and active smoking in the absence of empirical epidemiologic evidence on the magnitude of the association with mortality at high exposures of $PM_{2.5}$ in ambient environments.

Much stronger elevations in disease risk among smokers are seen for lung cancer and COPD than ischaemic heart disease and stroke. Consequently, more consistency might be expected between effects seen for COPD and lung cancer in air pollution studies, especially if there is some truth in the toxicity assumption made by Burnett et al. (Burnett et al., 2014). Consequently, this section will focus on COPD and NMRD as many studies report results for NMRD but not COPD, and COPD deaths make up a large proportion of NMRD deaths. COPD, like lung cancer, is strongly associated with tobacco smoking which is recognised as the most important risk factor for the development and the progression of COPD. Schikowski et al. noted that although tobacco smoke and combustion-related air pollution emit a range of pollutants in common, the role of ambient air pollution on the underlying chronic disease processes that ultimately lead to COPD are not well investigated (Schikowski et al., 2014a). Schikowski et al. reviewed evidence from eight morbidity and six mortality studies (Schikowski et al., 2014b). It was noted that neither mortality nor hospitalisation studies could unambiguously distinguish acute from long-term effects on the development of the underlying pathophysiological changes, and that the evidence of chronic effects of air pollution on the prevalence and incidence of COPD among adults was suggestive but not conclusive. Schikowski et al. also reported little evidence of association between $PM_{2.5}$ exposure and incident and prevalent COPD among subjects in the ESCAPE study (Schikowski et al., 2014a). The CPRD study conducted in the UK reported a significant association between $PM_{2.5}$ and COPD with an HR of 1.43 (95% CI, 1.00-1.99) per 10 $\mu g/m^3$ $PM_{2.5}$, but limited, inconclusive evidence for associations between air pollution and COPD incidence (Atkinson et al., 2015).

If the assumption made by Burnett et al. that the toxicity of ambient air pollution and tobacco smoke can be related to $PM_{2.5}$, then one would expect mortality due to NMRD to be elevated in population studies of air pollution (Burnett et al., 2014). However, this evidence is also inconclusive. Hoek et al. reviewed the literature and reported that the random effect pooled estimate per 10 $\mu g/m^3$ for $PM_{2.5}$ was 1.029 (95%CI, 0.941-1.126 (Hoek et al., 2013). The heterogeneity across studies was statistically significant with an I^2 statistic of 59%. Studies published after this review

continued to be inconsistent. The CPRD study reported a highly significant association between $PM_{2.5}$ and NMRD with an HR of 1.54 (95% CI, 1.27-1.86) per 10 µg/m^3 $PM_{2.5}$, but a much weaker association with lung cancer (HR=1.11; 95% CI, 0.85-1.43) (Carey et al., 2013). The association with NMRD was even stronger for never smokers (HR=1.99; 95% CI, 1.50-2.61), and significant associations with $PM_{2.5}$ exposure were also observed for current and ex-smokers. In contrast, Dimakopoulou et al. reported a protective effect of $PM_{2.5}$ exposure for NMRD among ESCAPE study participants with a meta-HR of 0.79 (95%CI, 0.44-1.25) per 10 µg/m^3 $PM_{2.5}$, with a statistically significant reduction among never smokers (meta-HR=0.22; 95% CI, 0.00-0.77) (Dimakopoulou et al., 2014). However, Raaschou-Nielsen et al. reported a non-significantly elevated risk of lung cancer among ESCAPE study participants (meta-HR=1.39; 95% CI, 0.91-2.13) (Raaschou-Nielsen et al., 2013). Clearly the assumption made by Burnett et al (2014) about the toxicity of particulate matter contained in ambient air pollution and tobacco smoke is unrealistic, as they themselves acknowledge. Nevertheless, much more consistency might have been expected between the results for lung cancer, COPD and NMRD in air pollution studies. In addition, making extrapolations from air pollution studies using findings from studies of tobacco smokers to extend the dose range does not appear to be supported in the case of respiratory disease.

7.5. ARE THE EFFECTS SEEN IN POPULATION STUDIES COHERENT WITH THE EFFECTS SEEN IN OCCUPATIONAL STUDIES?

Risk assessors are compelled to use the entire body of evidence when evaluating the strength of any arguments regarding causal associations between an air pollutant and particular human health effect. This includes an examination of environmental and occupational epidemiological evidence. In some cases, such as diesel exhaust, the conclusions regarding causality are derived solely from occupational studies because of the difficulty collecting reliable measurements in the ambient environment (Vermeulen et al., 2014). Since air pollutants represent a highly complex mixture of thousands agents that may exist as solids, liquids, or gases, there is always a concern that the pollutant of interest may simply be serving as a surrogate for the actual toxicant (Mayoralas-Alises and Diaz-Lobato, 2012). To partly circumvent this problem, epidemiologists often look to identify workplace cohorts to understand any putative relationships that may exist. Although a number of lung carcinogens have been identified through workplace studies, these have largely involved substances that are not primary air pollutants (Steenland et al., 1996). Consequently, there is little to no concordance between the lung carcinogens identified in environmental studies and those identified in occupational studies. An exception to this statement is diesel exhaust particulate, which can be found in both occupational and community environments. The diesel exhaust particulates found in ambient air cannot, however, be effectively separated from the particulates released from other combustion sources. Consequently, the workplace provides a good surrogate environment where secondary emission sources can be identified and accounted for in a health effects investigation. The results from these occupational studies can then provide a basis for making a hazard determination for the general public.

In fact, these were the circumstances surrounding IARCs recently completed evaluation diesel exhaust and lung cancer (IARC, 2014). This expert review used new and updated information from several occupational epidemiology studies to conclude the existence of a causal relationship that warranted the classification of diesel exhaust as a group I human carcinogen. Although dozens of occupational epidemiology studies have been performed with diesel exhaust over the years, the opinion was largely based on the results from three studies where the historical reconstruction of past exposures was purported to be vastly improved (Benbrahim-

Tallaa et al., 2012). In their analysis, IARC acknowledged the availability of environmental epidemiology data on the association between lung cancer and particulate matter; however, the information was not included in the review because it did not make a material contribution to the insight gained from the occupational studies. The key studies that were weighed heavily in the IARC analysis included investigations with non-metal miners, railroad workers, and those in the trucking industry.

The Diesel Exhaust in Miners Study (DEMS) attracted the most attention by the IARC review committee, since it included a retrospective cohort and a nested case-control evaluation of over 12,315 workers employed at eight U.S. non-metal mines where the use of diesel-powered equipment was common underground (Attfield et al., 2012, Silverman et al., 2012). Both studies used measurements of respirable elemental carbon (REC) as the basis for reconstructing historical exposures to diesel exhaust. For those time-periods where REC measurements were not directly available (approx. 30-50 years), the values were estimated based on past knowledge of carbon monoxide (CO) levels in the mines. The trend in these measurements was then compared to the REC exposure levels available for a three-year period to determine the mathematical relationship between CO and REC. To compensate for the fact that CO measurements were unavailable for periods earlier than about 1976, the authors used the total horsepower output of diesel equipment being used in each mine, the length of use of each piece of equipment, and the mine exhaust ventilation rate to devise a surrogate for the CO measurements that were unavailable. Both studies controlled for smoking and other confounders including silica, asbestos, non-diesels PAHs, radon, and respirable dust. The cohort study revealed that the lung cancer standardized mortality ratio (SMR) was 1.26 (95% CI 1.09-1.44) for the complete cohort, which included both above ground and underground worker locations. This association was obscured when the results were stratified by location. The nested case-control study included 198 individuals with lung cancer that were matched with 582 controls on the basis of location, gender, ethnicity, and birth year. Exposure cut-points (quartile and tertile) were established using cumulative exposure, exposure intensity, and exposure duration in years. Confounding from smoking, mine location, and non-malignant respiratory disease were taken into consideration. The results for all workers above and below ground showed a statistically significant increase in the risk for lung cancer for cumulative exposures lagged for 15-years.

Although the results from these two NIOSH-sponsored studies have been widely cited and used in support of listing diesel exhaust as a lung carcinogen, they continue to be controversial. The most commonly levelled criticism centred on the methods used to determine historical REC exposures on the basis of CO levels that were calculated rather than measured (Crump and Van Landingham, 2012, McClellan, 2012). Other critiques included an over adjustment for the healthy worker survival effect, a selection bias in the identification of matched controls, and discrepancies seen in above and below ground workers (Boffetta, 2012, Mohner and Wendt, 2017). Despite these views, an independent review of the findings from DEMS found that it was designed and constructed in a logical fashion with good use of available information for the exposure reconstructions (HEI, 2015). The panel concluded further that the results were valid for use in a quantitative risk assessment.

The IARC review of lung cancer and diesel exhaust also examined the results from a retrospective cohort study of workers in the trucking industry (Garshick et al., 2008, Garshick et al., 2012). Unionized male workers from four national trucking companies were recruited to participate in the study based on the availability of historical records. Exposure histories were established for 31,135 workers using the

coefficient of haze (COH) as a surrogate for elemental carbon (EC) measurement which were unavailable for the period from 1971 to 2000. The EC measurements collected from 2001 to 2006 were used to validate the COH exposure model. Monthly cumulative EC exposures were used to determine any associations with lung cancer using a proportional hazard model that included terms for race and residential location. After adjusting for employment duration which was inversely associated with lung cancer risk, the analysis with a 5-year lag found a linear exposure-response relationship with a hazard ratio of 1.07 (95% CI 0.99-1.15) per 1000 $\mu g/m^3$ months of cumulative exposure. A major drawback of this study was the failure to adjust the finding for cigarette smoking, which the authors felt was not a serious problem since the risks were not affected when the results were adjusted for the smoking rates obtained in a survey of workers in the industry.

The last occupational cohort study explicitly cited by IARC in their review of diesel exhaust focused on the associations observed in a group of 54,973 U.S. railroad workers (Garshick et al., 2004, Laden et al., 2006a). The 38-year study from 1959 to 1996 began at a time when diesel engines began replacing coal-fired locomotives. Using historical records, cumulative diesel particulate exposures were estimated from the annual average emission adjustment factor (EAF) for each of the locomotives in use during the study period. This measure of exposure was used along with a diesel fraction metric that represented the probability of a diesel particulate exposure for a given year. The study did not include an adjustment for smoking or any other potential confounders other than age and the healthy worker effect. Increasing exposure-response trends were observed for lung cancer in those workers hired from 1959 to 1966 when the conversion from steam engines to diesel was complete. In contrast, regardless of the exposure level, a relative risk of 1.77 (95% CI 1.50-2.09) was observed. There was no evidence of an exposure-response relationship in those workers hired before 1995 or when exposures were measured as cumulative intensity years. Consequently, some have argued that the study does not support the existence of a causal link between diesel exhaust and lung cancer (Gamble et al., 2012, Mohner and Wendt, 2017).

At first glance, the results from these recent workplace investigations with diesel exhaust would seem to support the belief that a causal relationship exists between lung cancer and outdoor air pollution and particulate matter (Loomis et al., 2013). In fact, despite the continued controversy regarding the suitability of these studies in ongoing causality discussions, the results from these three cohort studies with diesel exhaust provide suggestive evidence that a relationship may exist. This remark, however, must be tempered with the knowledge that a well-conducted retrospective cohort study in potash miners did not find an association between diesel exhaust exposures, represented as REC levels, and lung cancer (Mohner et al., 2013). This lack of concordance between studies, the acknowledged absence of objective exposure measurements, and methodological discrepancies has led some to conclude that sufficient information is not yet available regarding causality and that a clear exposure-response relationship cannot be identified for diesel exhaust (Moolgavkar et al., 2015, Sun et al., 2015). This ongoing debate may provide a reason why the aforementioned studies were not included in the IARC monograph on air pollution and lung cancer (IARC, 2016). Instead, the evaluation relied on results from occupational epidemiology studies involving professional drivers, police officers, mail carriers, and filling station attendants exposed to particulate matter from a variety of sources. But these studies are all compromised by the fact that indirect measures of exposure (period of employment, years of employment, type of driving) were used as a surrogate for the diesel exhaust levels (Tsoi and Tse, 2012). These problem areas with the occupational epidemiology studies with diesel exhaust need to be resolved before they provide meaningful support for the findings from environmental epidemiology studies.

8. INDICATOR OF AIR POLLUTION OR CAUSAL AGENT

Epidemiological studies of lung cancer and air pollution typically include measures of regulated pollutants such as $PM_{2.5}$, PM_{10}, NO_2/NO_x, SO_2 and ozone, and some have included measures of traffic intensity or distance from heavy traffic roads as surrogate measures of traffic-related air pollution. The most comprehensive information is available for $PM_{2.5}$, PM_{10} and NO_2. Particulate matter is often treated as both an indicator of air pollution and as a potential causal agent whereas NO_2 is generally treated as an indicator of air pollution, although there is some evidence that trends in the incidence of adenocarcinoma of the lung may be explained by NO_x emissions. If particulate matter does play a causal role, then it is more plausible that it is due to $PM_{2.5}$ rather than the coarse component of PM_{10}. Hamra et al. focuses attention on $PM_{2.5}$ because it includes a higher proportion of mutagenic species, many of which are products of combustion (Hamra et al., 2014). In addition, Hamra et al. noted that smaller particles penetrate more deeply into the lung and are more likely to be retained whilst the coarse fraction of PM_{10} consists mainly of minerals and biological materials.

Nitrogen dioxide is generated by some of the same sources that generate $PM_{2.5}$ (e.g. traffic) and levels are correlated with those of $PM_{2.5}$, but as an indicator of air pollution, NO_2 does not reflect power plant emissions (as SO_2 does), which are another source of $PM_{2.5}$. In health studies, estimated $PM_{2.5}$ and NO_2 exposures of subjects are often highly correlated e.g. Cesaroni et al. reported a correlation of 0.79 in an Italian study of over a million subjects (Cesaroni et al., 2013). However, the pattern of exposure is very different as NO_2 arises from local sources and is known to vary over smaller areas in proximity to traffic, whereas $PM_{2.5}$ is a mixture generated by primary and secondary sources that is more regionally dispersed (Gilliland et al., 2005). Consequently, levels of NO_2 recorded at central monitoring stations do not reflect the fine scale on which these variations occur. The localised pattern for NO_2 can be clearly seen in Figure 1 of Cesaroni et al. which shows maps of estimated concentrations of $PM_{2.5}$ and NO_2 levels for residents of Rome (the NO_2 pattern is described as having higher “resolution”) (Cesaroni et al., 2013). LUR models might be expected to estimate intra-urban NO_2 better than $PM_{2.5}$ and there is some evidence of this difference; although the disparity was fairly small in the ESCAPE study (Krewski et al., 2009, Wang et al., 2014b). However, Montagne et al. reported that correlations between modelled outdoor estimates and outdoor and personal measurements within three cities were much higher for NO_2 than $PM_{2.5}$ (Montagne et al., 2013).

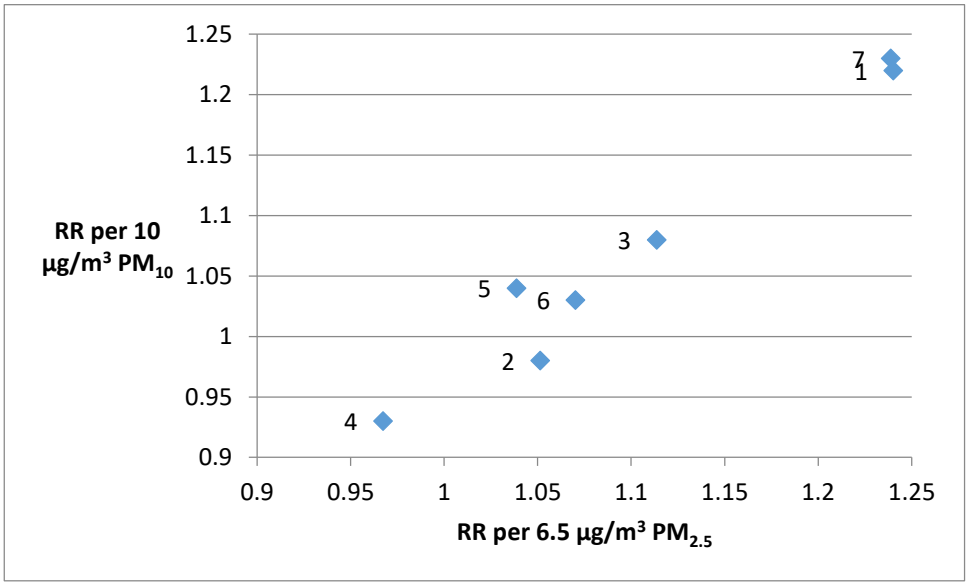
If PM is a causal agent, then one would expect that $PM_{2.5}$ would be associated more strongly with lung cancer than PM_{10}, and a weaker association will be observed for NO_2. Hamra et al. reported similar meta-estimates for the association between lung cancer and $PM_{2.5}$ and lung cancer and PM_{10}, although the latter meta-estimate was less precise (Hamra et al., 2014). At a regional level there were differences, and the North American meta-estimate for $PM_{2.5}$ (based on eight studies) was significantly elevated, whereas the corresponding meta-estimate for PM_{10} (based on five studies) was not suggestive of a relationship between PM_{10} and lung cancer. However, the $PM_{2.5}$ meta-estimate for North America is heavily influenced by the result from the ACS CPS-II study which was taken from the extended analysis by Krewski et al.; whereas the PM_{10} result for the ACS CPS-II study was taken from the earlier report by Pope et al. which had less refined exposure data and a slightly shorter follow-up period (two years fewer), and had less impact on the meta-estimate (Krewski et al., 2009, Pope et al., 2002). In contrast, the PM_{10} meta-estimate for Europe is much more suggestive of an association than that for $PM_{2.5}$, but this is largely due to a single study in the meta-analysis for $PM_{2.5}$ (Beelen et al.,

2008b, Heinrich et al., 2013). However, risk ratios for $PM_{2.5}$ and PM_{10} taken from the seven studies included in the meta-analysis of Hamra et al. are highly correlated. $PM_{2.5}$ risk estimates were standardised to a change of 6.5 µg/m^3, (i.e. assuming a $PM_{2.5}$ to PM_{10} ratio of 0.65) (Hamra et al., 2014). In addition, estimates for the AHSMOG study were both taken from the report on lung cancer mortality by McDonnell et al., instead of using the PM_{10} estimate from the report on lung cancer incidence by Beeson et al. which only studied PM_{10} (Beeson et al., 1998, McDonnell et al., 2000). In addition, both estimates for the ACS CPS-II study were taken from Pope et al. as Krewski et al. did not report a risk estimate for PM_{10}. The correlation between the risk ratios for $PM_{2.5}$ and PM_{10} is 0.98 (see Figure 13) (Krewski et al., 2009, Pope et al., 2002).

In order to compare the quantitative evidence of association for PM and NO_2, a meta-analysis was performed of the lung cancer risk associated with exposure to NO_2. This was done in two ways in a comparable way to the meta-analyses for PM in section 6.4. One analysis included the meta-HR from the ESCAPE study reported by Raaschou-Nielsen et al. and the other included all 17 individual HRs from the ESCAPE centres that had estimates for NO_2 (see Figure 14A and B) (Hamra et al., 2014, Raaschou-Nielsen et al., 2013). Hamra et al. performed a similar meta-analysis for NO_2 to that shown in Figure 14A, but included a non-significant result from the study by Filleul et al. that its authors did not consider to be the most valid result (Filleul et al., 2005, Hamra et al., 2014). In addition, Hamra et al. did not include the result reported by Vineis et al. for the European Prospective Investigation into Cancer and Nutrition (EPIC) study because of overlap with the ESCAPE study, but less than a quarter of the lung cancer cases studied by Vineis et al. were included in the ESCAPE study, the exposure methodologies and designs of the two studies were completely different (Raaschou-Nielsen et al., 2013, Vineis et al., 2006). Estimates by sex for the AHSMOG study were taken from a mortality analysis by Abbey et al., and were combined using fixed-effects estimation in the analysis reported here and by Hamra et al. (Abbey et al., 1999, Hamra et al., 2015). The significantly elevated RR for NO_2 in the AHSMOG mortality analysis was not confirmed in a corresponding analysis of lung cancer incidence by Beeson et al. which included more cases (36 versus 29), but the finding could not be included in the meta-analysis as the RR was not reported for women (Abbey et al., 1999, Beeson et al., 1998). However, the AHSMOG study has little weight in the meta-analysis. The meta-relative risk for lung cancer was 1.06 (95% CI: 1.02-1.14) per 10 µg/m^3 increase of NO_2 (Figure 14B), compared to meta-relative risk of 1.04 (95% CI: 1.01-1.08) reported by Hamra et al. (Hamra et al., 2015).

The meta-analysis for NO_2 is probably more convincing than that for $PM_{2.5}$ when one takes into consideration the fact that the range of NO_2 estimates for subjects in most studies is much larger than the range of $PM_{2.5}$ estimates. For example, Cesaroni et al. reported an IQR for NO_2 of 10.7 µg/m^3 and an IQR for $PM_{2.5}$ of 5.8 µg/m^3, but the difference is much greater in most studies: the IQR for NO_2 in the CNECSS study is six times greater than that for $PM_{2.5}$;.the ratios of the SD of estimates of NO_2 to $PM_{2.5}$ for 14 centres in the ESCAPE study ranged from 1.4 to 11.7 with a median of 5.1; and the ratios of the IQRs of estimates of NO_2 to $PM_{2.5}$ in the ACS-CPII study were 3.2 for $PM_{2.5}$ results from 1979-83 and 4.7 for $PM_{2.5}$ results from 1999-2000 (the NO_2 measurements were taken in 1980) (Cesaroni et al., 2013, Hystad et al., 2013, Krewski et al., 2009, Raaschou-Nielsen et al., 2013). Hamra et al. reported a meta-relative risk for lung cancer of 1.09 (95% CI: 1.04-1.14) per 10 µg/m^3 increase of $PM_{2.5}$ and 1.04 (95% CI: 1.01-1.08) per 10 µg/m^3 increase of NO_2 (Hamra et al., 2015). Even a 2-fold bigger exposure contrast for NO_2 than $PM_{2.5}$ would result in the relative risk for NO_2 increasing to 1.08 (95% CI: 1.02-1.16), and this further increases to 1.12 (95% CI: 1.02-1.16), if the results of the meta-analysis shown in Figure 14B are used.

Figure 13 Comparison of risk ratios for PM_{10} (10 µg/m^3 change) and $PM_{2.5}$ (6.5 µg/m^3 change) for studies which reported both estimates.

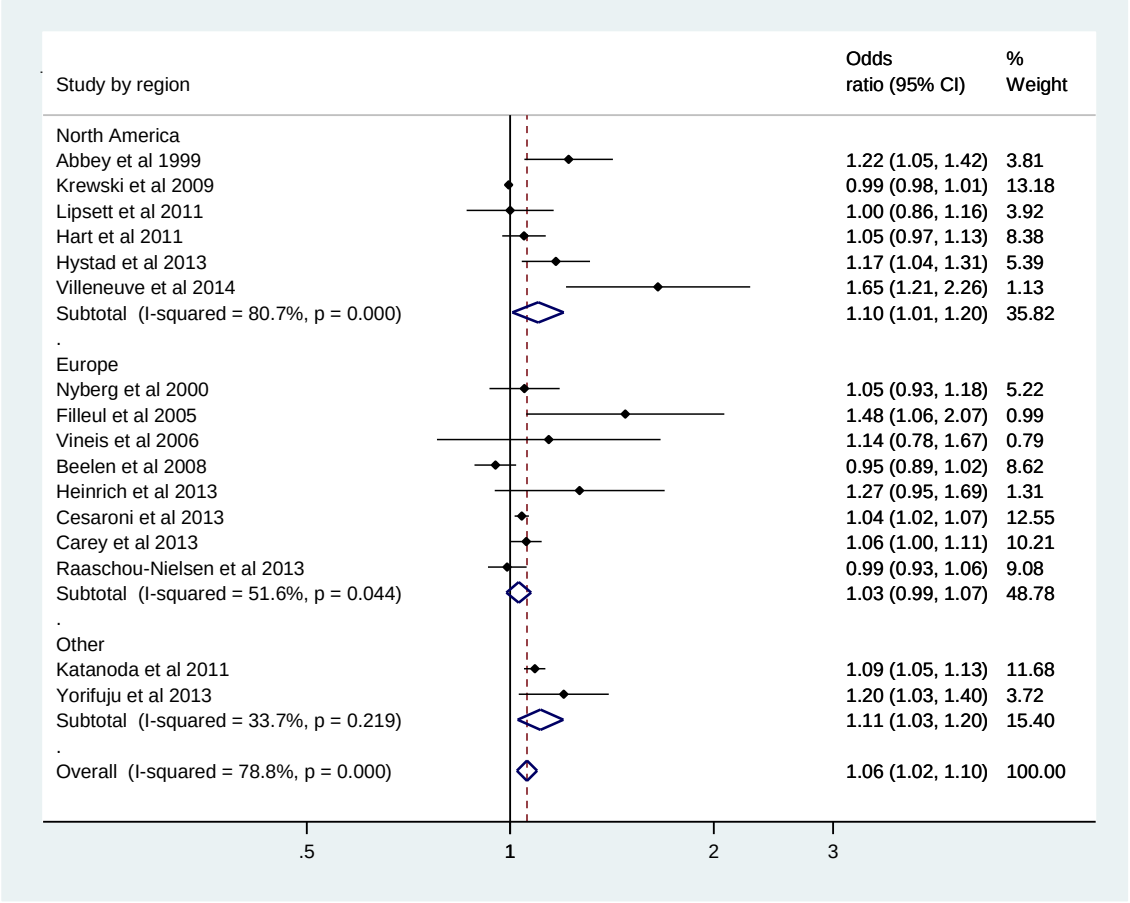


1 = Raaschou-Nielsen et al (2013), 2 = Pope et al (2002), 3 = Hart et al (2011), 4 = Lipsett et al (2011), 5 = Puett et al (2014), 6 = Carey et al (2013), 7 = McDonnell et al (2000)

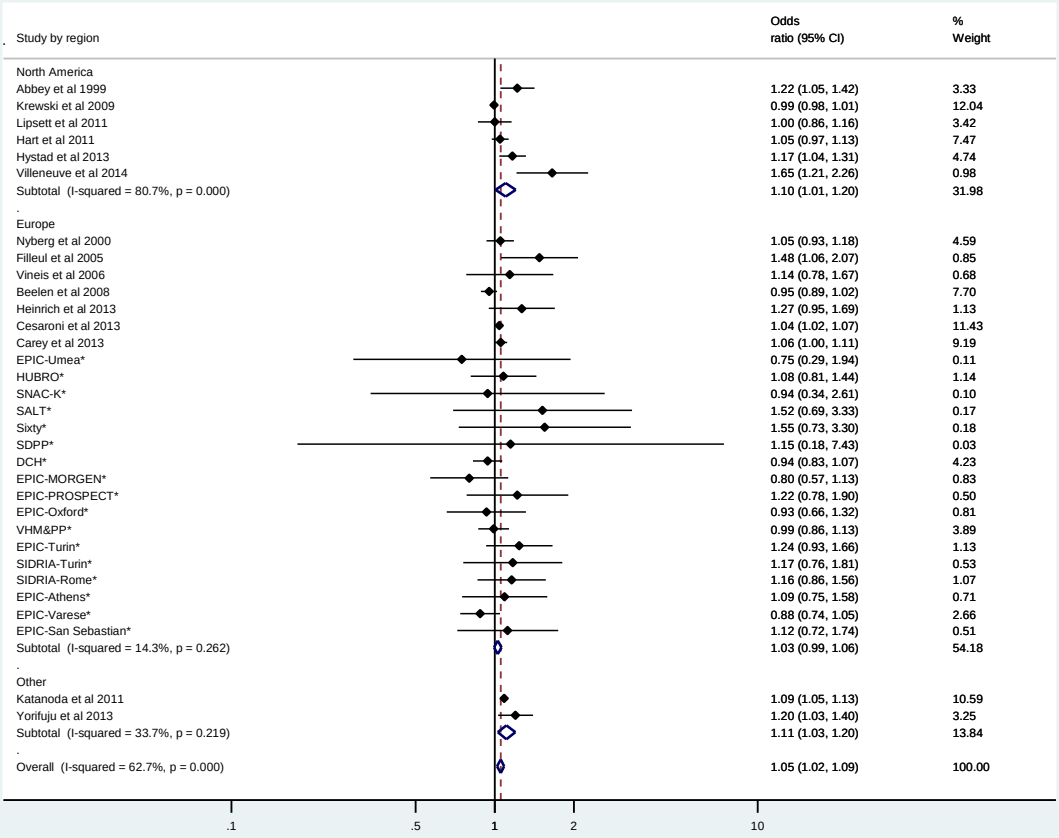
.

Figure 14 **Estimates of lung cancer risk associated with a 10 µg/m^3 change in exposure to NO_2 including meta-HR result from ESCAPE study (A) or 17 individual HRs from ESCAPE study (B)**

A.



B.



* ESCAPE cohorts (Raaschou-Nielsen et al., 2013)

9. CONCLUSIONS

Three lines of evidence may be employed in an evaluation of health hazards and risks; these include traditional toxicology studies in laboratory animals, human experimental clinical studies, and environmental epidemiology studies. The results from all three need to be expertly examined and carefully scrutinized to ensure that all necessary precautions have been taken to guard against spurious conclusions. Each of these approaches also possess particular advantages and disadvantages that need to be understood and appreciated when compiling the information into a weight of evidence evaluation. Human clinical studies and toxicology investigations are experimental in nature and allow for the control of independent variables that can impact the results. Environmental epidemiology studies on the other hand are observational in nature, so safeguards need to be taken to minimize the bias and confounding that can impact the findings. Environmental epidemiology studies have assumed a preeminent role in hazard and risk determinations because they involve direct observations in humans, do not require extrapolation from high exposure levels, can target sensitive subgroups, can involve large sample sizes, and are able to focus on both acute and chronic disease states (Brunekreef, 2008, Paddle and Harrington, 2000). Although studies in laboratory animals can also be designed to evaluate short-term and long-term health effects, they are encumbered by potential species differences that can limit their relevance for humans. Human experimental clinical investigations are also somewhat constrained since they are restricted to investigating acute reversible health effects. As such, the results from expertly designed and well performed epidemiology studies can yield health effects information that is both directly relevant to humans and usable in a quantitative risk assessment. To be of value, however, risk assessors need to be certain that the study protocol has taken into consideration various pitfalls that can prejudice the final observations.

The preceding report examines those factors and conditions that may affect the outcome of chronic epidemiology studies, especially those focusing on the lung cancer associated with exposures to ambient air particulates. The goal was to highlight specific issues that should be considered when designing or interpreting a cohort or case-control study seeking to examine the relationship between human exposure and carcinogenicity. Chief among these are the methods and techniques used to establish the cumulative exposure levels. Every attempt needs to be made to minimize exposure misclassification by incorporating newer models and estimation techniques that are capable of describing residential or personal levels with a high degree of spatial and temporal resolution (HEI, 2010, Smith et al., 2016). Since the science of exposure estimation is evolving at a very rapid rate with the development of new hybrid models capable of taking time-activity patterns into consideration, it is essential that outdated proximity-based approaches be abandoned in favour of those that provide a more realistic description of the exposures people actually experience. The increased accuracy and sophistication afforded by new exposure modelling techniques is accompanied by the need to carefully validate their performance for the time periods of interest. Ideally, this would include a sensitivity analysis and some measure of model performance relative to actual personal exposure levels.

Another important exposure-related issue that requires careful consideration is the inclusion of measurements during the biologically-relevant exposure period. All too often, investigators begin estimating exposures at the time of recruitment and do not reconstruct exposures during the time period when the disease process actually begins. This requires the implementation of suitable back extrapolation techniques that can relate past exposures to an emission source or an emission factor that has

been accurately tracked during the time periods of interest. The application of these and other techniques help ensure that differential and non-differential errors in the exposure estimates are minimized and that any identified risks are valid for the population of interest.

Confounding is a systemic problem in environmental epidemiology that needs to be assiduously and relentlessly tracked because of its deleterious impact on study quality. Although confounding can occur in many forms in a chronic health effects investigation, some types are particularly difficult to spot and may seem atypical. Examples include studies where there is only a partial adjustment for a participant's smoking history. This includes investigations where an individual's smoking history is determined at recruitment with no adjustment during the follow-up period as well as studies where the smoking history cannot be determined so surrogate measures such as socio-economic status are used as a substitute. Both of these circumstances may impact the results of a study because the impact of smoking was not adequately adjusted for in the hazard model. The resulting residual confounding may be appreciable under some circumstances and must be carefully evaluated when weighing the significance of any purported associations. This report has shown that smoking was not well controlled for in any of the population air pollution epidemiological studies. Indeed, no smoking information at all was available for the RoLS study which had considerable weight in the meta-analyses for $PM_{2.5}$ and NO_2 (Cesaroni et al., 2013). Also, the individual smoking information used in some air pollution cohort studies has an important limitation resulting from the fact that it was usually only collected at baseline and not updated. The magnitude of the association between smoking and lung cancer is so great that even a small degree of residual confounding could have a large effect on risk estimates. Consequently, IARC was correct to attach considerable weight to results for never smokers where confounding by smoking is less of an issue. However, the strength of this body of evidence is considerably weakened by the unexplained associations of a comparable or greater magnitude observed in smokers and former smokers, and further studies are needed to understand the reasons for this discrepancy.

Attention must also be given to other potential confounders; these include factors such as the compositional clustering of individuals living in the same residential area, secondary occupational exposures, and the collinearities that may exist with unmeasured or unmodelled ambient air pollutants. It is essential that these and other sources of confounding are appropriately addressed during the design and implementation stages of a study to ensure that any potential bias is kept to a minimum.

The results from an environmental epidemiology study may be used to conduct a quantitative health risk assessment provided that the concentration-response relationship has been well characterized. Even when reliable data is available from a chronic study using laboratory animals, preference may be given to the results from a cohort or case-control study, because of concerns surrounding potential species differences, which have been shown to exist for the chronic inhalation of particulates. A log-linear non-threshold model has generally been assumed to exist for those particulates capable of causing lung cancer. Although competing response functions have been identified, evidence for the linear concentration-response relationship seems to apply to many, but not all, investigations showing an association between lung cancer and particulate exposures. These factors need to be carefully considered in a quantitative risk assessment and steps should be taken to evaluate how alternative concentration-response functions can impact the risk determination.

An often overlooked characteristic of environmental epidemiology studies surrounds the heterogeneity or inconsistency that can be seen between different studies. Although observed health outcome differences may be related variations in the composition of particulate matter at separate locales or to genetic differences in different study populations, other factors such as diet and physical activity should not be overlooked. In many cases, the heterogeneity between studies is directly observable but in others, where there is considerable variability across the different studies, the application of appropriate statistical tests is necessary. These statistical techniques are not a panacea however, and careful consideration needs to be given to the ability of these tests to reliably detect response heterogeneity when it exists. Their application can cause some to conclude that there is no evidence of heterogeneity when the situation is in fact far more complicated. Oftentimes the issue cannot be reasonably evaluated without a detailed analysis of data subsets to determine if the results remain consistent or inconsistent.

A final consideration that is particularly relevant to those studies examining the relationship between air pollution and lung cancer concerns the plausibility of any findings showing an association. This is particularly relevant for investigations where comparisons can be made of the lung cancer risk for particulate exposures among current smokers, past smokers, and never smokers. The results from these different subgroups need to be carefully examined and interpreted to be certain that any differences are consistent with expectations. Another aspect that needs to be closely examined to ensure credible findings focuses on the types of tumours observed and their incidence in smokers and non-smokers exposed to ambient air particulates. Similarly, the incidence rate for pulmonary disease other than lung cancer needs to be examined to assess whether these health outcomes parallel the risk findings for lung cancer. Studies that segregate men and women should show the expected sex difference with higher risks tending to exist for male members of the cohort. A gradation in lung cancer risk would also be expected in studies examining the association with pollutants other than $PM_{2.5}$, which would be expected to show the strongest relationship. Finally, confidence is increased when the results from environmental epidemiology agree with those from occupational studies focusing on same types of particulates found in ambient air. These issues highlight some, but not all, of the plausibility concerns that need to be investigated to assess the believability of the study findings. If the plausibility litmus test is satisfied then there will be greater confidence in the validity of the findings whether positive or negative.

In October of 2013, IARC issued a statement declaring that outdoor air pollution was carcinogenic to humans (Group 1). This overall evaluation was based on evaluations of the strength of evidence arising from humans and experimental animals, and the strength of mechanistic evidence. In humans, IARC considered that there was sufficient evidence that outdoor air pollution caused cancer of the lung and sufficient evidence that particulate matter in outdoor air pollution caused lung cancer. They further concluded that a positive relationship had been observed between exposure and lung cancer in which chance, bias and confounding could be ruled out with sufficient confidence. The basis for this finding was a series of epidemiology studies conducted in many different parts of the world using different study designs and a range of approaches to quantitatively or qualitatively estimate exposure levels. The results consistently showed a positive exposure-response relationship between lung cancer and outdoor air pollution concentrations, most often assessed as mass concentrations of particulate matter (Loomis et al., 2014). The available studies were examined both individually and in a meta-analysis reported by Hamra et al. which was not included in the IARC monograph (Hamra et al., 2014). Hamra et al. noted that the meta-analysis originated with the IARC review, and that their quantitative analyses complemented the qualitative

classification of the evidence by the IARC Working Group (Hamra et al., 2014). After pooling the 18 available studies using a random effects model, a risk ratio of 1.09 (95% CI 1.04-1.14) per 10 µg/m^3 of $PM_{2.5}$ was obtained. Further support for this evaluation was provided by the observation that the increased risk associated with outdoor air pollution was also seen in studies restricted to never smokers which were not subject to confounding by tobacco smoke. It was recognised by IARC that all the studies were subject to error in estimating exposure, but it was considered that the most likely effect of such error would be an attenuation of the risk estimates.

The evaluation of PM as a causal agent was based on the same epidemiological evidence. IARC noted that most evidence for an association between outdoor air pollution and lung cancer came from the results for PM, hence the evidence for lung cancer and PM is generally similar to that for outdoor air pollution as a whole. IARC noted that PM could be acting as a surrogate for the outdoor air pollution mixture or other individual components, but concluded that there was a causal effect of PM exposure. However, IARC only discussed one hypothetical alternative to PM being the causal agent in the outdoor air pollution mixture. The alternative hypothesis discussed was the existence of gas-phase carcinogens highly correlated with the PM concentrations, but this was considered unrealistic and contradicted by the known presence of multiple carcinogens in airborne PM. It was also reiterated that associations have been observed in multiple locations with different pollution mixtures, and lung cancer risk increased with increasing concentrations of mass-based PM indicators. However, such associations and exposure response relationships would presumably also be observed if gas-phase carcinogens existed that were highly correlated with the PM concentrations.

PM as the causal agent may be the most plausible hypothesis, but the IARC evaluation that PM is carcinogenic to humans isn't very helpful for understanding risk. Hamra et al. focused attention on $PM_{2.5}$ because it included a higher proportion of mutagenic species, many of which are products of combustion. In addition, it was noted that smaller particles penetrated more deeply into the lung and were more likely to be retained (Hamra et al., 2014, Hamra et al., 2015). Borm et al. noted that combustion-derived nanoparticles, the dominant particle type by number in urban air, represented a key component of the particulate matter mix because they contained a large surface area, transition metals, and organic species (Borm et al., 2007). In contrast, much of the coarse fraction of PM_{10} consisted of low-toxicity components such as ammonium sulphates and nitrates, sea salt (sodium chloride), crustal dust, and road dust (Borm et al., 2007). Nevertheless, as discussed in section 8, the epidemiological evidence isn't sufficient to distinguish between $PM_{2.5}$ and PM_{10}, or even NO_2 which is presumably only a surrogate for exposure. Furthermore, it should be clear after reading this report that the epidemiological database is simply not good enough for a quantitative risk assessment. At best, the exposure contrast defined by the exposure estimates may be a reasonable surrogate for the true exposure contrast. At worst, the few studies looking at actual exposure versus predicted exposure at the gatepost of a subject suggest that there may be very little relationship with the true exposure contrast during the relevant period of exposure. Certainly, the estimates of risk per unit of exposure are not good enough to be used in global burden calculations. Hopefully, studies with better exposure assessment methodologies will be used in future that are better able to identify the agent that drives the adverse effect. Borm et al. postulated that measurement of the oxidant activity of particulate matter might be a better metric than mass concentration, and a recent study has examined whether $PM_{2.5}$ oxidative burden (i.e. the ability of $PM_{2.5}$ to cause oxidative stress) is more strongly associated with lung cancer mortality than $PM_{2.5}$ mass concentration (Borm et al., 2007, Weichenthal et al., 2016).

There are many who advocate a more systematic approach to establishing causality that includes a more rigorous examination of residual confounding in any observational studies being evaluated (Goodman et al., 2013, Rooney et al., 2014, Swaen and van Amelsvoort, 2009). This includes an evaluation of study quality, internal validity, and consistency. The preceding evaluation shows that there are many issues and attributes that need to be examined when assessing the validity of a study evaluating the relationship between particulate exposures and lung cancer, or in designing and analysing future studies. These include but are not limited to those summarized in Table 5 below.

Table 5 Issues and safeguards affecting the strength of observed relationships between air pollutant exposures and lung cancer

Source of Error	Safeguard
exposure misclassification	use of advanced exposure estimation models along with some measure of validation
measurements during the biologically-relevant exposure period	implementation of suitable back extrapolation techniques
confounding from tobacco use	avoid the use of surrogate measures and verify full adjustment during the follow-up period
compositional clustering from contextual variables	implement suitable statistical controls using datasets that are appropriate for this source of confounding
pollutant collinearities	use of multi-pollutant models
heterogeneity across studies	careful analysis of the results from the compiled studies and the application of appropriate statistical tests
implausible outcomes	detailed evaluation of the consistency in the findings for different sexes, tumour types, and tobacco use habits

A thorough and complete investigation of these areas of concern will help ensure that the findings from a particular group of investigations are valid and relevant for use in a risk assessment.

10. GLOSSARY

Acronym	Definition
ACS	American Cancer Society
AHSMOG	Adventist Health Study on Smog
APEX	Air Pollutant Exposure model
BREP	biologically relevant exposure period
CMAQ	Community Multi-scale Air Quality
CPRD	Clinical Practice Research Datalink
CNECSS	Canadian National Enhanced Cancer Surveillance System
CO	carbon monoxide
COPD	chronic obstructive pulmonary disease
CPS-II	US Cancer Prevention Study II
CTS	California Teachers Study
DM	dispersion modelling
EC	elemental carbon
ESCAPE	European Study of Cohorts for Air Pollution Effects
GIS	geographic information systems
GPS	global positioning system
H6C	Harvard Six Cities
HOV	hold-out validation
IDW	inverse distance weighting
LOOCV	leave-one-out cross-validation
LUR	land use regression
NHS	Nurses Health Study
NMRD	non-malignant respiratory disease
NO_2	nitrogen dioxide
NO_x	nitrogen oxides
PAHs	polycyclic aromatic hydrocarbons
PM	particulate matter
$PM_{2.5}$	particulate matter with diameter of 2.5 µm or less
$PM_{2.5}$ absorbance	measurement of the blackness of $PM_{2.5}$
PM_{10}	particulate matter with diameter of 10 µm or less
PM_{coarse}	coarse particulate matter = PM_{10} – $PM_{2.5}$
QHRA	quantitative health risk assessment
r^2	percentage of variance explained
RoLS	Rome Longitudinal Study
RR	relative risk
SHS	second hand tobacco smoke
SO_2	sulphur dioxide
TPCS	Three-Prefecture Cohort Study
SHEDS	Stochastic Human Exposure and Simulation
SPM	suspended particulate matter
TSP	total suspended particulates
TrIPS	Trucking Industry Particle Study
TSP	total suspended particles
UFP	ultrafine particles
VOCs	volatile organic compounds
WOE	weight-of-evidence

11. REFERENCES

1. Abbey, D. E., Nishino, N., McDonnell, W. F., Burchette, R. J., Knutsen, S. F., Beeson, W. L., and Yang, J. X. 1999. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. American Journal of Respiratory and Critical Care Medicine 159:373-382.

2. Akita, Y., Baldasano, J. M., Beelen, R., Cirach, M., de Hoogh, K., Hoek, G., Nieuwenhuijsen, M., Serre, M. L., and de Nazelle, A. 2014. Large scale air pollution estimation method combining land use regression and chemical transport modeling in a geostatistical framework. *Environmental Science & Technology* 48:4452-4459.

3. Atkinson, R. W., Carey, I. M., Kent, A. J., van Staa, T. P., Anderson, H. R., and Cook, D. G. 2015. Long-term exposure to outdoor air pollution and the incidence of chronic obstructive pulmonary disease in a national English cohort. *Occupational and Environmental Medicine* 72:42-48.

4. Attfield, M. D., Schleiff, P. L., Lubin, J. H., Blair, A., Stewart, P. A., Vermeulen, R., Coble, J. B., and Silverman, D. T. 2012. The Diesel Exhaust in Miners Study: A cohort mortality study with emphasis on lung cancer. *Jnci-Journal of the National Cancer Institute* 104:869-883.

5. Barnett, A. G., Huang, C. R., and Turner, L. 2012. Benefits of publicly available data. *Epidemiology* 23:500-501.

6. Basagana, X., Aguilera, I., Rivera, M., Agis, D., Foraster, M., Marrugat, J., Elosua, R., and Kunzli, N. 2013. Measurement error in epidemiologic studies of air pollution based on land-use regression models. *American Journal of Epidemiology* 178:1342-1346.

7. Basagana, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., and Kunzli, N. 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmospheric Environment* 54:634-642.

8. Batterman, S., Burke, J., Isakov, V., Lewis, T., Mukherjee, B., and Robins, T. 2014. A comparison of exposure metrics for traffic-related air pollutants: Application to epidemiology studies in Detroit, Michigan. *International Journal of Environmental Research and Public Health* 11:9553-9577.

9. Baxter, L. K., Dionisio, K. L., Burke, J., Sarnat, S. E., Sarnat, J. A., Hodas, N., Rich, D. Q., Turpin, B. J., Jones, R. R., Mannshardt, E., Kumar, N., Beevers, S. D., and Ozkaynak, H. 2013. Exposure prediction approaches used in air pollution epidemiology studies: Key findings and future recommendations. *Journal of Exposure Science and Environmental Epidemiology* 23:654-659.

10. Baxter, L. K., Wright, R. J., Paciorek, C. J., Laden, F., Suh, H. H., and Levy, J. I. 2010. Effects of exposure measurement error in the analysis of health effects from traffic-related air pollution. *Journal of Exposure Science and Environmental Epidemiology* 20:101-111.

11. Beelen, R., Hoek, G., Fischer, P., van den Brandt, P. A., and Brunekreef, B. 2007. Estimated long-term outdoor air pollution concentrations in a cohort study. *Atmospheric Environment* 41:1343-1358.

12. Beelen, R., Hoek, G., van den Brandt, P. A., Goldbohm, R. A., Fischer, P., Schouten, L. J., Armstrong, B., and Brunekreef, B. 2008a. Long-term exposure to traffic-related air pollution and lung cancer risk. *Epidemiology* 19:702-710.

13. Beelen, R., Hoek, G., van den Brandt, P. A., Goldbohm, R. A., Fischer, P., Schouten, L. J., Jerrett, M., Hughes, E., Armstrong, B., and Brunekreef, B. 2008b. Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). *Environmental Health Perspectives* 116:196-202.

14. Beelen, R., Voogt, M., Duyzer, J., Zandveld, P., and Hoek, G. 2010. Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a Dutch urban area. *Atmospheric Environment* 44:4614-4621.

15. Beeson, W. L., Abbey, D. E., and Knutsen, S. F. 1998. Long-term concentrations of ambient air pollutants and incident lung cancer in California adults: Results from the AHSMOG study. *Environmental Health Perspectives* 106:813-822.

16. Beevers, S. D., Kitwiroon, N., Williams, M. L., Kelly, F. J., Anderson, H. R., and Carslaw, D. C. 2013. Air pollution dispersion models for human exposure predictions in London. *Journal of Exposure Science and Environmental Epidemiology* 23:647-653.

17. Bellander, T., Berglind, N., Gustavsson, P., Jonson, T., Nyberg, F., Pershagen, G., and Jarup, L. 2001. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environmental Health Perspectives* 109:633-639.

18. Benbrahim-Tallaa, L., Baan, R. A., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., Guha, N., Loomis, D., and Straif, K. 2012. Carcinogenicity of diesel-engine and gasoline-engine exhausts and some nitroarenes. *Lancet Oncology* 13:663-664.

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

20. Beyea, J., and Hatch, M. 1999. Geographic exposure modeling: A valuable extension of geographic information systems for use in environmental epidemiology. *Environmental Health Perspectives* 107:181-190.

21. Beyea, J., Stellman, S. D., Hatch, M., and Gammon, M. D. 2008. Airborne emissions from 1961 to 2004 of Benzo[a]pyrene from US vehicles per km of travel based on tunnel studies. *Environmental Science & Technology* 42:7315-7320.

22. Boffetta, P. 2012. Re: The Diesel Exhaust in Miners Study: A nested case-control study of lung cancer and diesel exhaust and a cohort mortality study with emphasis on lung cancer. *Journal of the National Cancer Institute* 104:1842-1843.

23. Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., and Freudenheim, A. L. 2003. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology* 14:408-412.

24. Borm, P., Cassee, F. R., and Oberdorster, G. 2015. Lung particle overload: Old school - new insights? *Particle and Fibre Toxicology* 12:1-5.

25. Borm, P. J. A., Kelly, F., Kunzli, N., Schins, R. P. F., and Donaldson, K. 2007. Oxidant generation by particulate matter: from biologically effective dose to a promising, novel metric. *Occupational and Environmental Medicine* 64:73-74.

26. Brauer, M., Ainslie, B., Buzzelli, M., Henderson, S., Larson, T., Marshall, J., Nethery, E., Steyn, D., and Su, J. 2008. Models of Exposure for Use in Epidemiology Studies of Air Pollution Health Impacts. In *Air Pollution Modeling and Its Applications XIX*, 589-604. Borrego, C., and Niranda, A. I., eds. New York, NY.

27. Bravo, M. A., Fuentes, M., Zhang, Y., Burr, M. J., and Bell, M. L. 2012. Comparison of exposure estimation methods for air pollutants: Ambient monitoring data and regional air quality simulation. *Environmental Research* 116:1-10.

28. Briggs, D. J. 2007. The use of GIS to evaluate traffic-related pollution. *Occupational and Environmental Medicine* 64:1-2.

29. Brook, J. R., Burnett, R. T., Dann, T. F., Cakmak, S., Goldberg, M. S., Fan, X. H., and Wheeler, A. J. 2007. Further interpretation of the acute effect of nitrogen dioxide observed in Canadian time-series studies. *Journal of Exposure Science and Environmental Epidemiology* 17:S36-S44.

30. Brunekreef, B. 2008. Environmental epidemiology and risk assessment. *Toxicology Letters* 180:118-122.

31. Brunekreef, B., Baker, D., and Nieuwenhuijsen, M. J. 2008. What is Environmental Epidemiology? In *Environmental Epidemiology: Study Methods and Application*, 1-14. Baker, D., and Nieuwenhuijsen, M. J., eds. Oxford, UK: Oxford University Press.

32. Burnett, R. T., Pope, C. A., Ezzati, M., Olives, C., Lim, S. S., Mehta, S., Shin, H. H., Singh, G., Hubbell, B., Brauer, M., Anderson, H. R., Smith, K. R., Balmes, J. R., Bruce, N. G., Kan, H. D., Laden, F., Pruss-Ustun, A., Michelle, C. T., Gapstur, S. M., Diver, W. R., and Cohen, A. 2014. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environmental Health Perspectives* 122:397-403.

33. Burnett, R. T., Stieb, D., Brook, J. R., Cakmak, S., Dales, R., Raizenne, M., Vincent, R., and Dann, T. 2004. Associations between short-term changes in nitrogen dioxide and mortality in Canadian cities. *Archives of Environmental Health* 59:228-236.

34. Buzas, J. S., Stefanski, L. A., and Tosteson, T. D. 2014. Measurement Error. In *Handbook of Epidemiology*, 1241-1282. Ahrens, W. E., and Pigeot, I. R., eds. New York, NY: Springer-Verlag.

35. CALEPA. 1998. Part B: Health Risk Assessment for Diesel Exhaust. Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento, CA.

36. Carey, I. M., Atkinson, R. W., Kent, A. J., van Staa, T., Cook, D. G., and Anderson, H. R. 2013. Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. *American Journal of Respiratory and Critical Care Medicine* 187:1226-1233.

37. Cayo, M. R., and Talbot, T. O. 2003. Positional error in automated geocoding of residential addresses. *International Journal of Health Geographics* 2:1-12.

38. Cefalu, M., and Dominici, F. 2014. Does exposure prediction bias health effect estimation? The relationship between confounding adjustment and exposure prediction. *Epidemiology* 25:583-590.

39. Cesaroni, G., Badaloni, C., Gariazzo, C., Stafoggia, M., Sozzi, R., Davoli, M., and Forastiere, F. 2013. Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. *Environmental Health Perspectives* 121:324-331.

40. Cesaroni, G., Badaloni, C., Porta, D., Forastiere, F., and Perucci, C. A. 2008. Comparison between various indices of exposure to traffic-related air pollution and their impact on respiratory health in adults. *Occupational and Environmental Medicine* 65:683-690.

41. Cesaroni, G., Porta, D., Badaloni, C., Stafoggia, M., Eeftens, M., Meliefste, K., and Forastiere, F. 2012. Nitrogen dioxide levels estimated from land use regression models several years apart and association with mortality in a large cohort study. *Environmental Health* 11:1-10.

42. Chang, E. T., Adami, H. O., Bailey, W. H., Boffetta, P., Krieger, R. I., Moolgavkar, S. H., and Mandel, J. S. 2014. Validity of geographically modeled environmental exposure estimates. *Critical Reviews in Toxicology* 44:450-466.

43. Checkoway, H., Pearce, N., and Kriebel, D. 2004. Characterizing the workplace environment. In *Research Methods in Occupational Epidemiology*. New York, NY: Oxford University Press.

44. Checkoway, H., Pearce, N., and Kriebel, D. 2007. Selecting appropriate study designs to address specific research questions in occupational epidemiology. *Occupational and Environmental Medicine* 64:633-638.

45. Chen, F., Cole, P., and Bina, W. F. 2007. Time trend and geographic patterns of lung adenocarcinoma in the United States, 1973-2002. *Cancer Epidemiology Biomarkers & Prevention* 16:2724-2729.

46. Chen, F., Jackson, H., and Bina, W. F. 2009. Lung adenocarcinoma incidence rates and their relation to motor vehicle density. *Cancer Epidemiology Biomarkers & Prevention* 18:760-764.

47. Chinn, S., Florey, C. D., Baldwin, I. G., and Gorgol, M. 1981. The relation of mortality in England and Wales 1969-73 to measurements of air pollution. *Journal of Epidemiology and Community Health* 35:174-179.

48. Christensen, K. Y., Naidu, A., Parent, M. E., Pintos, J., Abrahamowicz, M., Siemiatycki, J., and Koushik, A. 2012. The risk of lung cancer related to dietary intake of flavonoids. *Nutrition and Cancer-An International Journal* 64:964-974.

49. Cordioli, M., Ranzi, A., Sterrantino, A. F., Erspamer, L., Razzini, G., Ferrari, U., Gatti, M. G., Bonora, K., Artioli, F., Goldoni, C. A., and Lauriola, P. 2014. A comparison between self-reported and GIS-based proxies of residential exposure to environmental pollution in a case-control study on lung cancer. *Spatial and Spatio-temporal Epidemiology* 9:37-45.

50. Crump, K., and Van Landingham, C. 2012. Evaluation of an exposure assessment used in epidemiological studies of diesel exhaust and lung cancer in underground mines. *Critical Reviews in Toxicology* 42:599-612.

51. Cury, P. M., Lichtenfels, A. J. F. C., Reymao, M. S. F., Conceicao, G. M. S., Capelozzi, V. L., and Saldiva, P. H. N. 2000. Urban levels of air pollution modifies the progression of urethane-induced lung tumours in mice. *Pathology Research and Practice* 196:627-633.

52. Cyrys, J., Hochadel, M., Gehring, U., Hoek, G., Diegmann, V., Brunekreef, B., and Heinrich, J. 2005. GIS-based estimation of exposure to particulate matter and NO2 in an urban area: Stochastic versus dispersion modeling. *Environmental Health Perspectives* 113:987-992.

53. Dal-Re, R., Ioannidis, J. P., Bracken, M. B., Buffler, P. A., Chan, A. W., Franco, E. L., La Vecchia, C., and Weiderpass, E. 2014. Making prospective registration of observational research a reality. *Science Translational Medicine* 6:1-6.

54. de Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M. J., Badaloni, C., Beelen, R., Bolignano, A., Cesaroni, G., Pradas, M. C., Cyrys, J., Douros, J., Eeftens, M., Forastiere, F., Forsberg, B., Fuks, K., Gehring, U., Gryparis, A., Gulliver, J., Hansell, A. L., Hoffmann, B., Johansson, C., Jonkers, S., Kangas, L., Katsouyanni, K., Kunzli, N., Lanki, T., Memmesheimer, M., Moussiopoulos, N., Modig, L., Pershagen, G., Probst-Hensch, N., Schindler, C., Schikowski, T., Sugiri, D., Teixido, O., Tsai, M. Y., Yli-Tuomi, T., Brunekreef, B., Hoek, G., and Bellander, T. 2014. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. *Environment International* 73:382-392.

55. De Matteis, S., Consonni, D., and Bertazzi, P. A. 2008. Exposure to occupational carcinogens and lung cancer risk: Evolution of epidemiological estimates of attributable fraction. *Acta Biomedica* 79:34-42.

56. de Mesnard, L. 2013. Pollution models and inverse distance weighting: Some critical remarks. *Computers & Geosciences* 52:459-469.

57. Devesa, S. S., Bray, F., Vizcaino, A. P., and Parkin, D. M. 2005. International lung cancer trends by histologic type: male : female differences diminishing and adenocarcinoma rates rising. *International Journal of Cancer* 117:294-299.

58. Dimakopoulou, K., Samoli, E., Beelen, R., Stafoggia, M., Andersen, Z. J., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M., Vineis, P., Xun, W., Hoek, G., Raaschou-Nielsen, O., Oudin, A., Forsberg, B., Modig, L., Jousilahti, P., Lanki, T., Turunen, A., Oftedal, B., Nafstad, P., Schwarze, P. E., Penell, J., Fratiglioni, L., Andersson, N., Pedersen, N., Korek, M., De Faire, U., Eriksen, K. T., Tjonneland, A., Becker, T., Wang, M., Bueno-de-Mesquita, B., Tsai, M. Y., Eeftens, M., Peeters, P. H., Meliefste, K., Marcon, A., Kramer, U., Kuhlbusch, T. A. J., Vossoughi, M., Key, T., de Hoogh, K., Hampel, R., Peters, A., Heinrich, J., Weinmayr, G., Concin, H., Nagel, G., Ineichen, A., Jacquemin, B., Stempfelet, M., Vilier, A., Ricceri, F., Sacerdote, C., Pedeli, X., Katsoulis, M., Trichopoulou, A., Brunekreef, B., and Katsouyanni, K. 2014. Air pollution and nonmalignant respiratory mortality in 16 cohorts within the ESCAPE project. *American Journal of Respiratory and Critical Care Medicine* 189:684-696.

59. Dionisio, K. L., Baxter, L. K., and Chang, H. H. 2014. An empirical assessment of exposure measurement error and effect attenuation in bipollutant epidemiologic models. *Environmental Health Perspectives* 122:1216-1224.

60. Dionisio, K. L., Isakov, V., Baxter, L. K., Sarnat, J. A., Sarnat, S. E., Burke, J., Rosenbaum, A., Graham, S. E., Cook, R., Mulholland, J., and Ozkaynak, H. 2013.

Development and evaluation of alternative approaches for exposure assessment of multiple air pollutants in Atlanta, Georgia. *Journal of Exposure Science and Environmental Epidemiology* 23:581-592.

61. Dockery, D. W., Pope, C. A., Xu, X. P., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris, B. G., and Speizer, F. E. 1993. An association between air pollution and mortality in 6 United States cities. *New England Journal of Medicine* 329:1753-1759.

62. Doll, R., and Peto, R. 1981. *The Cause of Cancer*. New York, NY: Oxford University Press.

63. Dominici, F., Greenstone, M., and Sunstein, C. R. 2014. Particulate matter matters. *Science* 344:257-259.

64. Dosemeci, M., Stewart, P. A., and Blair, A. 1990. Three proposals for retrospective, semiquantitative exposure assessments and their comparison with the other assessment methods. *Applied Occupational and Environmental Hygiene* 5:52-59.

65. Dwyer-Lindgren, L., Bertozzzi-Villa, A., and Stubbs, R. W. 2017. Inequalities in Life Expectancy Among US Counties, 1980 to 2014. Temporal Trends and Key Drivers. *JAMA Internal Medicine* doi: 10.1001/jamainternmed.2017.0918. [Epub ahead of print].

66. Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dedele, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Grazuleviciene, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Molter, A., Nador, G., Nieuwenhuijsen, M., Nonnemacher, M., Pedeli, X., Raaschou-Nielsen, O., Patelarou, E., Quass, U., Ranzi, A., Schindler, C., Stempfelet, M., Stephanou, E., Sugiri, D., Tsai, M. Y., Yli-Tuomi, T., Varro, M. J., Vienneau, D., von Klot, S., Wolf, K., Brunekreef, B., and Hoek, G. 2012. Development of land use regression models for PM2.5, PM2.5 absorbance, PM10 and PMcoarse in 20 European study areas; Results of the ESCAPE project. *Environmental Science & Technology* 46:11195-11205.

67. Eeftens, M., Beelen, R., Fischer, P., Brunekreef, B., Meliefste, K., and Hoek, G. 2011. Stability of measured and modelled spatial contrasts in NO2 over time. *Occupational and Environmental Medicine* 68:765-770.

68. Filleul, L., Rondeau, V., Vandentorren, S., Le Moual, N., Cantagrel, A., Annesi-Maesano, I., Charpin, D., Declercq, C., Neukirch, F., Paris, C., Vervloet, D., Brochard, P., Tessier, J. F., Kauffmann, F., and Baldi, I. 2005. Twenty five year mortality and air pollution: results from the French PAARC survey. *Occupational and Environmental Medicine* 62:453-460.

69. Fryer, M., Collins, C. D., Ferrier, H., Colvile, R. N., and Nieuwenhuijsen, M. J. 2006. Human exposure modelling for chemical risk assessment: a review of current approaches and research and policy implications. *Environmental Science & Policy* 9:261-274.

70. Gamble, J. F., Nicolich, M. J., and Boffetta, P. 2012. Lung cancer and diesel exhaust: An updated critical review of the occupational epidemiology literature. *Critical Reviews in Toxicology* 42:549-598.

71. Gan, W. Q., Koehoorn, M., Davies, H. W., Demers, P. A., Tamburic, L., and Brauer, M. 2011. Long-term exposure to traffic-related air pollution and the risk of coronary

heart disease hospitalization and mortality. *Environmental Health Perspectives* 119:501-507.

72. Garshick, E., Laden, F., Hart, J. E., Davis, M. E., Eisen, E. A., and Smith, T. J. 2012. Lung cancer and elemental carbon exposure in trucking industry workers. *Environmental Health Perspectives* 120:1301-1306.

73. Garshick, E., Laden, F., Hart, J. E., Rosner, B., Davis, M. E., Eisen, E. A., and Smith, T. J. 2008. Lung cancer and vehicle exhaust in trucking industry workers. *Environmental Health Perspectives* 116:1327-1332.

74. Garshick, E., Laden, F., Hart, J. E., Rosner, B., Smith, T. J., Dockery, D. W., and Speizer, F. E. 2004. Lung cancer in railroad workers exposed to diesel exhaust. *Environmental Health Perspectives* 112:1539-1543.

75. Gilliland, F., Avol, E., Kinney, P., Jerrett, M., Dvonch, T., Lurmann, F., Buckley, T., Breysse, P., Keeler, G., de Villiers, T., and McConnell, R. 2005. Air pollution exposure assessment for epidemiologic studies of pregnant women and children: Lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. *Environmental Health Perspectives* 113:1447-1454.

76. Goldberg, M., Kromhout, H., Guenel, P., Fletcher, A. C., Gerin, M., Glass, D. C., Heederik, D., Kauppinen, T., and Ponti, A. 1993. Job exposure matrices in industry. *International Journal of Epidemiology* 22:S10-S15.

77. Goodman, J. E., Prueitt, R. L., Sax, S. N., Bailey, L. A., and Rhomberg, L. R. 2013. Evaluation of the causal framework used for setting National Ambient Air Quality Standards. *Critical Reviews in Toxicology* 43:829-849.

78. Greenland, S., and Robbins, J. 1994. Invited Commentary - Ecologic studies - biases, misconceptions, and counter examples. *American Journal of Epidemiology* 139:747-760.

79. Gresner, P., Gromadzinska, J., and Wasowicz, W. 2007. Polymorphism of selected enzymes involved in detoxification and biotransformation in relation to lung cancer. *Lung Cancer* 57:1-25.

80. Grimes, D. A., and Schulz, K. F. 2012. False alarms and pseudo-epidemics The limitations of observational epidemiology. *Obstetrics and Gynecology* 120:920-927.

81. Gulliver, J., de Hoogh, K., Fecht, D., Vienneau, D., and Briggs, D. 2011a. Comparative assessment of GIS-based methods and metrics for estimating long-term exposures to air pollution. *Atmospheric Environment* 45:7072-7080.

82. Gulliver, J., Morris, C., Lee, K., Vienneau, D., Briggs, D., and Hansell, A. 2011b. Land use regression modeling to estimate historic (1962-1991) concentrations of black smoke and sulfur dioxide for Great Britain. *Environmental Science & Technology* 45:3526-3532.

83. Hajat, A., Diez-Roux, A. V., Adar, S. D., Auchincloss, A. H., Lovasi, G. S., O'Neill, M. S., Sheppard, L., and Kaufman, J. D. 2013. Air pollution and individual and neighborhood socioeconomic status: Evidence from the multi-ethnic study of atherosclerosis (MESA). *Environmental Health Perspectives* 121:1325-1333.

84. Hales, S., Blakely, T., and Woodward, A. 2012. Air pollution and mortality in New Zealand: Cohort study. *Journal of Epidemiology and Community Health* 66:468-473.

85. Hamra, G. B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J. M., Vineis, P., Forastiere, F., Saldiva, P., Yorifuji, T., and Loomis, D. 2014. Outdoor particulate matter exposure and lung cancer: A systematic review and meta-analysis. *Environmental Health Perspectives* 122:906-911.

86. Hamra, G. B., Laden, F., Cohen, A., Raaschou-Nielsen, O., Brauer, M., and Loomis. 2015. Lung cancer and exposure to nitrogen dioxide and traffic: A systematic review and meta-analysis. *Environmental Health Perspectives* 123:1107-1112.

87. Harrison, R. M., and Yin, J. X. 2000. Particulate matter in the atmosphere: which particle properties are important for its effects on health? *Science of the Total Environment* 249:85-101.

88. Hart, J. E., Garshick, E., Dockery, D. W., Smith, T. J., Ryan, L., and Laden, F. 2011. Long-term ambient multipollutant exposures and mortality. *American Journal of Respiratory and Critical Care Medicine* 183:73-78.

89. Healy, M. A., and Gilliland, J. A. 2012. Quantifying the magnitude of environmental exposure misclassification when using imprecise address proxies in public health research. *Spatial and Spatio-temporal Epidemiology* 3:55-67.

90. HEI. 2010. Traffic-Related Air Pollution:A Critical Review of the Literature on Emissions, Exposure, and Health Effects. Special Report 17: Health Effects Institute, Boston, MA.

91. HEI. 2015. Diesel Emissions and Lung Cancer: An Evaluation of Recent Epidemiological Evidence for Qunatitative Risk Assessment. Special Report 19: Helth Effects Institute, Boston, MA.

92. Heinrich, J., Thiering, E., Rzehak, P., Kramer, U., Hochadel, M., Rauchfuss, K. M., Gehring, U., and Wichmann, H. E. 2013. Long-term exposure to NO2 and PM10 and all-cause and cause-specific mortality in a prospective cohort of women. *Occupational and Environmental Medicine* 70:179-186.

93. Heo, J., Schauer, J. J., Yi, O., Paek, D., Kim, H., and Yi, S. M. 2014. Fine particle air pollution and mortality importance of specific sources and chemical species. *Epidemiology* 25:379-388.

94. Hesterberg, T. W., Bunn, W. B., McClellan, R. O., Hart, G. A., and Lapin, C. A. 2005. Carcinogenicity studies of diesel engine exhausts in laboratory animals: A review of past studies and a discussion of future research needs. *Critical Reviews in Toxicology* 35:379-411.

95. Hesterberg, T. W., Long, C. M., Bunn, W. B., Lapin, C. A., McClellan, R. O., and Valberg, P. A. 2012. Health effects research and regulation of diesel exhaust: An historical overview focused on lung cancer risk. *Inhalation Toxicology* 24:1-45.

96. Higgins, J. P. T., Thompson, S. G., Deeks, J. J., and Altman, D. G. 2003. Measuring inconsistency in meta-analyses. *British Medical Journal* 327:557-560.

97. Higgins, P. T., and Green, S. 2011. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 ed. London, UK: The Cochrane Collaboration.

98. Hill, A. B. 1965. Environment and disease - Association or causation. *Proceedings of the Royal Society of Medicine-London* 58:295-300.

99. Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., and Briggs, D. 2008a. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* 42:7561-7578.

100. Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., and Briggs, D. 2008b. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* 42:7561-7578.

101. Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., and Kaufman, J. D. 2013. Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environmental Health* 12.

102. Hornung, R. W. 1991. Statistical evaluation of exposure assessment strategies. *Applied Occupational and Environmental Hygiene* 6:516-520.

103. Hornung, R. W., Herrick, R. F., Stewart, P. A., Utterback, D. F., Feigley, C. E., Wall, D. K., Douthit, D. E., and Hayes, R. B. 1996. An experimental design approach to retrospective exposure assessment. *American Industrial Hygiene Association Journal* 57:251-256.

104. Huang, Y. L., and Batterman, S. 2000. Residence location as a measure of environmental exposure: A review of air pollution epidemiology studies. *Journal of Exposure Analysis and Environmental Epidemiology* 10:66-85.

105. Hughes, A. E., and Pruitt, S. L. 2014.Using address histories in health research: Challenges and recommendations for research. http://healthgis.tamu.edu/Program/2014/healthgis-11.pdf

106. Hung, R. J., Boffetta, P., Brockmoller, J., Butkiewicz, D., Cascorbi, I., Clapper, M. L., Garte, S., Haugen, A., Hirvonen, A., Anttila, S., Kalina, I., Le Marchand, L., London, S. J., Rannug, A., Romkes, M., Salagovic, J., Schoket, B., Gaspari, L., and Taioli, E. 2003. CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. *Carcinogenesis* 24:875-882.

107. Hutchings, S., and Rushton, L. 2011. Toward risk reduction: Predicting the future burden of occupational cancer. *American Journal of Epidemiology* 173:1069-1077.

108. Hystad, P., Demers, P. A., Johnson, K. C., Brook, J., van Donkelaar, A., Lamsal, L., Martin, R., and Brauer, M. 2012. Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. *Environmental Health* 11:DOI: 10.1186/1476-069X-11-22.

109. Hystad, P., Demers, P. A., Johnson, K. C., Carpiano, R. M., and Brauer, M. 2013. Long-term residential exposure to air pollution and lung cancer risk. *Epidemiology* 24:762-772.

110. IARC. 2013. Air Pollution and Cancer. World Health Organization, International Agency for Research on Cancer, Geneva, Switzerland.

111. IARC. 2014. Diesel and Gasoline Engine Exhausts and Some Nitroarenes. World Health Organization, International Agency for Research on Cancer, Geneva, Switzerland.

112. IARC. 2016. Outdoor Air Pollution. World Health Organization, International Agency for Research on Cancer, Geneva, Switzerland.

113. Ioannidis, J. P. A. 2008. Why most discovered true associations are inflated. *Epidemiology* 19:640-648.

114. Jacquez, G. 2012. A research agenda: Does geocoding positional error matter in health GIS studies? *Spatial and Spatio-temporal Epidemiology* advanced online publication:doi: 10.1016/j.sste.2012.02.002.

115. Jacquez, G. M., Slotnick, M. J., Meliker, J. R., AvRuskin, G., Copeland, G., and Nriagu, J. 2011. Accuracy of commercially available residential histories for epidemiologic studies. *American Journal of Epidemiology* 173:236-243.

116. Jedrychowski, W., Becher, H., Wahrendorf, J., and Basacierpialek, Z. 1990. A case-control study of lung-cancer with special reference to the effect of air pollution in Poland. *Journal of Epidemiology and Community Health* 44:114-120.

117. Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., and Giovis, C. 2005. A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology* 15:185-204.

118. Jerrett, M., Burnett, R. T., Beckerman, B. S., Turner, M. C., Krewski, D., Thurston, G., Martin, R. V., van Donkelaar, A., Hughes, E., Shi, Y. L., Gapstur, S. M., Thun, M. J., and Pope, C. A. 2013. Spatial analysis of air pollution and mortality in California. *American Journal of Respiratory and Critical Care Medicine* 188:593-599.

119. Jerrett, M., and Finkelstein, M. 2005. Geographies of risk in studies linking chronic air pollution exposure to health outcomes. *Journal of Toxicology and Environmental Health-Part A-Current Issues* 68:1207-1242.

120. Johnson, M., Isakov, V., Touma, J. S., Mukerjee, S., and Ozkaynak, H. 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmospheric Environment* 44:3660-3668.

121. Jurek, A. M., Greenland, S., and Maldonado, G. 2008. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? *International Journal of Epidemiology* 37:382-385.

122. Katanoda, K., Sobue, T., Satoh, H., Tajima, K., Suzuki, T., Nakatsuka, H., Takezaki, T., Nakayama, T., Nitta, H., Tanabe, K., and Tominaga, S. 2011. An association between long-term exposure to ambient air pollution and mortality from lung cancer and respiratory diseases in Japan. *Journal of Epidemiology* 21:132-143.

123. Khuder, S. A. 2001. Effect of cigarette smoking on major histological types of lung cancer: A meta-analysis. *Lung Cancer* 31:139-148.

124. Kim, D., Kumfer, B. M., Anastasio, C., Kennedy, I. M., and Young, T. M. 2009. Environmental aging of polycyclic aromatic hydrocarbons on soot and its effect on source identification. *Chemosphere* 76:1075-1081.

125. Kim, J. Y., Burnett, R. T., Neas, L., Thurston, G. D., Schwartz, J., Tolbert, P. E., Brunekreef, B., Goldberg, M. S., and Romieu, I. 2007. Panel discussion review: session two - interpretation of observed associations between multiple ambient air pollutants and health effects in epidemiologic analyses. *Journal of Exposure Science and Environmental Epidemiology* 17:S83-S89.

126. Krall, J. R., Chang, H. H., Sarnat, S. E., Peng, R. D., and Waller, L. A. 2015. Current methods and challenges for epidemiological studies of the associations between chemical constituents of particulate matter and health. *Current Environmental Health Reports* 2:388-398.

127. Krewski, D., Burnett, R. T., Goldberg, M. S., Hoover, K., Siemiatycki, J., Jerrett, M., Abrahamowicz, M., White, W. H., Bartlett, G., Brodsky, L., Calisti, L., Chen, Y., De Luca, P., du Berger, R., Finkelstein, N., Foliart, D. E., Fung, K., Hughes, E., Kader, S., Lakhani, R., Luo, X., Nadon, L., Ma, R., Mallick, R., Mo, F., Richardson, L., Schichtel, B., Schopflocher, T., Shi, Y., Villeneuve, P., and Willis, A. 2000. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Health Effects Institute, Cambridge, MA.

128. Krewski, D., Jerrett, M., Burnett, R. T., Ma, R., Hughes, E., Shi, Y., Turner, M. C., Pope, C. A., Thurston, G., Calle, E. E., and Thun, M. J. 2009. Extended Follow-up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality. Research Report 140: Health Effects Institute, Boston, MA.

129. Kuehni, C. E., Strippoli, M. P. F., Zwahlen, M., and Silverman, M. 2006. Association between reported exposure to road traffic and respiratory symptoms in children: Evidence of bias. *International Journal of Epidemiology* 35:779-786.

130. Kuempel, E. D., Smith, R. J., Dankovic, D. A., and Stayner, L. T. 2009. Rat- and human-based risk estimates of lung cancer from occupational exposure to poorly-soluble particles: A quantitative evaluation. *Journal of Physics: Conference Series* 161:1-12.

131. Laden, F., Hart, J. E., Eschenroeder, A., Smith, T. J., and Garshick, E. 2006a. Historical estimation of diesel exhaust exposure in a cohort study of US railroad workers and lung cancer. *Cancer Causes & Control* 17:911-919.

132. Laden, F., Schwartz, J., Speizer, F. E., and Dockery, D. W. 2006b. Reduction in fine particulate air pollution and mortality - Extended follow-up of the Harvard six cities study. *American Journal of Respiratory and Critical Care Medicine* 173:667-672.

133. Lee, I. M., Sesso, H. D., and Paffenbarger, R. S. 1999. Physical activity and risk of lung cancer. *International Journal of Epidemiology* 28:620-625.

134. Leech, J. A., Nelson, W. C., Burnett, R. T., Aaron, S., and Raizenne, M. E. 2002. It's about time: A comparison of Canadian and American time-activity patterns. *Journal of Exposure Analysis and Environmental Epidemiology* 12:427-432.

135. Lepeule, J., Laden, F., Dockery, D., and Schwartz, J. 2012. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard six cities study from 1974 to 2009. *Environmental Health Perspectives* 120:965-970.

136. Lipfert, F. W., and Wyzga, R. E. 2008. On exposure and response relationships for health effects associated with exposure to vehicular traffic. *Journal of Exposure Science and Environmental Epidemiology* 18:588-599.

137. Lipfert, F. W., Wyzga, R. E., Baty, J. D., and Miller, J. P. 2006. Traffic density as a surrogate measure of environmental exposures in studies of air pollution health effects: Long-term mortality in a cohort of US veterans. *Atmospheric Environment* 40:154-169.

138. Lipsett, M. J., Ostro, B. D., Reynolds, P., Goldberg, D., Hertz, A., Jerrett, M., Smith, D. F., Garcia, C., Chang, E. T., and Bernstein, L. 2011. Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort. *American Journal of Respiratory and Critical Care Medicine* 184:828-835.

139. Lokke, A., Lange, P., Scharling, H., Fabricius, P., and Vestbo, J. 2006. Developing COPD: a 25 year follow up study of the general population. *Thorax* 61:935-939.

140. Loomis, D., Grosse, Y., Ghissassi, F. E., Bouvard, V., Benbrahim-Talla, L., Guha, N., Baan, R., Mattock, H., and Straif, K. 2014. IARC evaluation of the carcinogenicity of outdoor air pollution. *Environment Risque & Sante* 13:347-352.

141. Loomis, D., Grosse, Y., Lauby-Secretan, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Baan, R., Mattock, H., and Straif, K. 2013. The carcinogenicity of outdoor air pollution. *Lancet Oncology* 14:1262-1263.

142. Loomis, D., and Kromhout, H. 2004. Exposure variability: Concepts and applications in occupational epidemiology. *American Journal of Industrial Medicine* 45:113-122.

143. Madsen, C., Gehring, U., Haberg, S. E., Nafstad, P., Meliefste, K., Nystad, W., Carlsen, K. C. L., and Brunekreef, B. 2011. Comparison of land-use regression models for predicting spatial NOx contrasts over a three year period in Oslo, Norway. *Atmospheric Environment* 45:3576-3583.

144. Mao, Y., Pan, S., Wen, S. W., and Johnson, K. C. 2003. Physical activity and the risk of lung cancer in Canada. *American Journal of Epidemiology* 158:564-575.

145. Marshall, J. D., Nethery, E., and Brauer, M. 2008. Within-urban variability in ambient air pollution: Comparison of estimation methods. *Atmospheric Environment* 42:1359-1369.

146. Mayoralas-Alises, S., and Diaz-Lobato, S. 2012. Air pollution and lung cancer. *Current Respiratory Medicine Reviews* 8:418-429.

147. McClellan, R. O. 2012. Re: The Diesel Exhaust in Miners Study: A nested case-control study of lung cancer and diesel exhaust and a cohort mortality study with emphasis on lung cancer. *Journal of the National Cancer Institute* 104:1843-1845.

148. McDonnell, W. F., Nishino-Ishikawa, N., Petersen, F. F., Chen, L. H., and Abbey, D. E. 2000. Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. *Journal of Exposure Analysis and Environmental Epidemiology* 10:427-436.

149. Miller, K. A., Siscovick, D. S., Sheppard, L., Shepherd, K., Sullivan, J. H., Anderson, G. L., and Kaufman, J. D. 2007. Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine* 356:447-458.

150. Mohner, M., Kersten, N., and Gellissen, J. 2013. Diesel motor exhaust and lung cancer mortality: Reanalysis of a cohort study in potash miners. *European Journal of Epidemiology* 28:159-168.

151. Mohner, M., and Wendt, A. 2017. A critical review of the relationship between occupational exposure to diesel emissions and lung cancer risk. *Critical Reviews in Toxicology* 47:185-224.

152. Montagne, D., Hoek, G., Nieuwenhuijsen, M., Lanki, T., Pennanen, A., Portella, M., Meliefste, K., Eeftens, M., Yli-Tuomi, T., Cirach, M., and Brunekreef, B. 2013. Agreement of land use regression models with personal exposure measurements of particulate matter and nitrogen oxides air pollution. *Environmental Science & Technology* 47:8523-8531.

153. Moolgavkar, S. H., Chang, E. T., Luebeck, G., Lau, E. C., Watson, H. N., Crump, K. S., Boffetta, P., and McClellan, R. 2015. Diesel engine exhaust and lung cancer mortality: Time-related factors in exposure and risk. *Risk Analysis* 35:663-675.

154. Morfeld, P., Keil, U., and Spallek, M. 2013. The European "Year of the Air": fact, fake or vision? *Archives of Toxicology* 87:2051-2055.

155. Naess, O., Nafstad, P., Aamodt, G., Claussen, B., and Rosland, P. 2007. Relation between concentration of air pollution and cause-specific mortality: Four-year exposures to nitrogen dioxide and particulate matter pollutants in 470 neighborhoods in Oslo, Norway. *American Journal of Epidemiology* 165:435-443.

156. Nafstad, P., Haheim, L. L., Oftedal, B., Gram, F., Holme, I., Hjermann, I., and Leren, P. 2003. Lung cancer and air pollution: A 27 year follow up of 16 209 Norwegian men. *Thorax* 58:1071-1076.

157. Nuckols, J. R., Ward, M. H., and Jarup, L. 2004. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environmental Health Perspectives* 112:1007-1015.

158. Nyberg, F., Gustavsson, P., Jarup, L., Bellander, T., Berglind, N., Jakobsson, R., and Pershagen, G. 2000. Urban air pollution and lung cancer in Stockholm. *Epidemiology* 11:487-495.

159. O'Neill, M. S., Jerrett, M., Kawachi, L., Levy, J. L., Cohen, A. J., Gouveia, N., Wilkinson, P., Fletcher, T., Cifuentes, L., and Schwartz, J. 2003. Health, wealth, and air pollution: Advancing theory and methods. *Environmental Health Perspectives* 111:1861-1870.

160. Oberdorster, G. 1995. Lung particle overload - Implications for occupational exposures to particles. *Regulatory Toxicology and Pharmacology* 21:123-135.

161. Offermans, N. S. M., Vermeulen, R., Burdorf, A., Peters, S., Goldbohm, R. A., Koeman, T., Van Tongeren, M., Kauppinen, T., Kant, I., Kromhout, H., and van den Brandt, P. A. 2012. Comparison of expert and job-exposure matrix-based retrospective exposure assessment of occupational carcinogens in the Netherlands Cohort Study. *Occupational and Environmental Medicine* 69:745-751.

162. Ostro, B., Lipsett, M., Reynolds, P., Goldberg, D., Hertz, A., Garcia, C., Henderson, K. D., and Bernstein, L. 2010. Long-term exposure to constituents of fine particulate air pollution and mortality: Results from the California Teachers study. *Environmental Health Perspectives* 118:363-369.

163. Oudin, A., Forsberg, B., Stromgren, M., Beelen, R., and Modig, L. 2012. Impact of residential mobility on exposure assessment in longitudinal air pollution studies: A sensitivity analysis within the ESCAPE project. *Scientific World Journal* .

164. Ozkaynak, H., Baxter, L. K., Dionisio, K. L., and Burke, J. 2013. Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of Exposure Science and Environmental Epidemiology* 23:566-572.

165. Paddle, G. M., and Harrington, J. M. 2000. Environmental epidemiology - Strengths and weaknesses. *International Archives of Occupational and Environmental Health* 73:7-14.

166. Pearce, N., Checkoway, H., and Kriebel, D. 2007. Bias in occupational epidemiology studies. *Occupational and Environmental Medicine* 64:562-568.

167. Pedeli, X., Hoek, G., and Katsouyanni, K. 2011. Risk assessment of diesel exhaust and lung cancer: combining human and animal studies after adjustment for biases in epidemiological studies. *Environmental Health* 10.

168. Peng, R. D., Dominici, F., and Zeger, S. L. 2006. Reproducible epidemiologic research. *American Journal of Epidemiology* 163:783-789.

169. Pereira, F. A. C., Lemos, M., Mauad, T., de Assuncao, J. V., and Saldiva, P. H. N. 2011. Urban, traffic-related particles and lung tumors in urethane treated mice. *Clinics* 66:1051-1054.

170. Pesch, B., Kendzia, B., Gustavsson, P., Jockel, K. H., Johnen, G., Pohlabeln, H., Olsson, A., Ahrens, W., Gross, I. M., Brunske, I., Wichmann, H. E., Merletti, F., Richiardi, L., Simonato, L., Fortes, C., Siemiatycki, J., Parent, M. E., Consonni, D., Landi, M. T., Caporaso, N., Zaridze, D., Cassidy, A., Szeszenia-Dabrowska, N., Rudnai, P., Lissowska, J., Stucker, I., Fabianova, E., Dumitru, R. S., Bencko, V., Foretova, L., Janout, V., Rudin, C. M., Brennan, P., Boffetta, P., Straif, K., and Bruning, T. 2012. Cigarette smoking and lung cancerurelative risk estimates for the major histological types from a pooled analysis of case-control studies. *International Journal of Cancer* 131:1210-1219.

171. Phillips, C. V., and Goodman, K. J. 2006. Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *Emerging Themes in Epidemiology* 3:1-5.

172. Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thurston, G. D. 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama-Journal of the American Medical Association* 287:1132-1141.

173. Pope, C. A., Burnett, R. T., Turner, M. C., Cohen, A., Krewski, D., Jerrett, M., Gapstur, S. M., and Thun, M. J. 2011. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: Shape of the exposure-response relationships. *Environmental Health Perspectives* 119:1616-1621.

174. Pope, C. A., Ezzati, M., and Dockery, D. W. 2009. Fine particulate air pollution and life expectancy in the United States. *New England Journal of Medicine* 360:376-386.

175. Pope, C. A., Thun, M. J., Namboodiri, M. M., Dockery, D. W., Evans, J. S., Speizer, F. E., and Heath, C. W. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. *American Journal of Respiratory and Critical Care Medicine* 151:669-674.

176. Prud'homme, G., Dobbin, N. A., Sun, L., Burnett, R. T., Martin, R. V., Davidson, A., Cakmak, S., Villeneuve, P. J., Lamsal, L. N., van Donkelaar, A., Peters, P. A., and Johnson, M. 2013. Comparison of remote sensing and fixed-site monitoring approaches for examining air pollution and health in a national study population. *Atmospheric Environment* 80:161-171.

177. Puett, R. C., Hart, J. E., Yanosky, J. D., Spiegelman, D., Wang, M. L., Fisher, J. A., Hong, B. L., and Laden, F. 2014. Particulate matter air pollution exposure, distance to road, and incident lung cancer in the nurses' health study cohort. *Environmental Health Perspectives* 122:926-932.

178. Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M. J., Brunekreef, B., Xun, W. W., Katsouyanni, K., Dimakopoulou, K., Sommar, J., Forsberg, B., Modig, L., Oudin, A., Oftedal, B., Schwarze, P. E., Nafstad, P., De Faire, U., Pedersen, N. L., Ostenson, C. G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K. T., Sorensen, M., Tjonneland, A., Ellermann, T., Eeftens, M., Peeters, P. H., Meliefste, K., Wang, M., Bueno-de-Mesquita, B., Key, T. J., de Hoogh, K., Concin, H., Nagel, G., Vilier, A., Grioni, S., Krogh, V., Tsai, M. Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Trichopoulou, A., Bamia, C., Vineis, P., and Hoek, G. 2013. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncology* 14:813-822.

179. Raaschou-Nielsen, O., Andersen, Z. J., Hvidberg, M., Jensen, S. S., Ketzel, M., Sorensen, M., Hansen, J., Loft, S., Overvad, K., and Tjonneland, A. 2011a. Air pollution from traffic and cancer incidence: A Danish cohort study. *Environmental Health* 10:1-67.

180. Raaschou-Nielsen, O., Andersen, Z. J., Hvidberg, M., Jensen, S. S., Ketzel, M., Sorensen, M., Loft, S., Overvad, K., and Tjonneland, A. 2011b. Lung cancer incidence and long-term exposure to air pollution from traffic. *Environmental Health Perspectives* 119:860-865.

181. Raaschou-Nielsen, O., Bak, H., Sorensen, M., Jensen, S. S., Ketzel, M., Hvidberg, M., Schnohr, P., Tjonneland, A., Overvad, K., and Loft, S. 2010. Air pollution from traffic and risk for lung cancer in three Danish cohorts. *Cancer Epidemiology Biomarkers & Prevention* 19:1284-1291.

182. Raaschou-Nielsen, O., Beelen, R., Wang, M., Hoek, G., Andersen, Z. J., Hoffmann, B., Stafoggia, M., Samoli, E., Weinmayr, G., Dimakopoulou, K., Nieuwenhuijsen, M., Xun, W. W., Fischer, P., Eriksen, K. T., Sorensen, M., Tjonneland, A., Ricceri, F., de Hoogh, K., Key, T., Eeftens, M., Peeters, P. H., Bueno-de-Mesquita, H. B., Meliefste, K., Oftedal, B., Schwarze, P. E., Nafstad, P., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Penell, J., De Faire, U., Korek, M., Pedersen, N., Ostenson, C. G., Pershagen, G., Fratiglioni, L., Concin, H., Nagel, G., Jaensch, A., Ineichen, A., Naccarati, A., Katsoulis, M., Trichpoulou, A., Keuken, M., Jedynska, A., Kooter, I. M., Kukkonen, J., Brunekreef, B., Sokhi, R. S., Katsouyanni, K., and Vineis, P. 2016. Particulate matter air pollution components and risk for lung cancer. *Environment International* 87:66-73.

183. Raimondi, S., Paracchini, V., Autrup, H., Barros-Dios, J. M., Benhamou, S., Boffetta, P., Cote, M. L., Dialyna, I. A., Dolzan, V., Filiberti, R., Garte, S., Hirvonen, A., Husgafvel-Pursiainen, K., Imyanitov, E. N., Kalina, I., Kang, D., Kiyohara, C., Kohno, T., Kremers, P., Lan, Q., London, S., Povey, A. C., Rannug, A., Reszka, E., Risch, A., Romkes, M., Schneider, J., Seow, A., Shields, P. G., Sobti, R. C., Sorensen, M., Spinola, M., Spitz, M. R., Strange, R. C., Stucker, I., Sugimura, H., To-Figueras, J., Tokudome, S., Yang, P., Yuan, J. M., Warholm, M., and Taioli, E. 2006. Meta- and pooled analysis of GSTT1 and lung cancer: A HuGE-GSEC review. *American Journal of Epidemiology* 164:1027-1042.

184. Reymao, M. S. F., Cury, P. M., Lichtenfels, A. J. F. C., Lemos, M., Battlehner, C. N., Conceicao, G. M. S., Capelozzi, V. L., Montes, G. S., Junior, M. F., Martins, M. A., Bohm, G. M., and Saldiva, P. H. N. 1997. Urban air pollution enhances the formation of urethane-induced lung tumors in mice. *Environmental Research* 74:150-158.

185. Rhomberg, L. R., Goodman, J. E., Bailey, L. A., Prueitt, R. L., Beck, N. B., Bevan, C., Honeycutt, M., Kaminski, N. E., Paoli, G., Pottenger, L. H., Scherer, R. W., Wise, K. C., and Becker, R. A. 2013. A survey of frameworks for best practices in weight-of-evidence analyses. *Critical Reviews in Toxicology* 43:753-784.

186. Ris, C. 2007. US EPA health assessment for diesel engine exhaust: A review. *Inhalation Toxicology* 19:229-239.

187. Robinson, A. L., Donahue, N. M., and Rogge, W. F. 2006. Photochemical oxidation and changes in molecular composition of organic aerosol in the regional context. *Journal of Geophysical Research-Atmospheres* 111.

188. Rooney, A. A., Boyles, A. L., Wolfe, M. S., Bucher, J. R., and Thayer, K. A. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. *Environmental Health Perspectives* 122:711-718.

189. Rose, N., Cowie, C., Gillett, R., and Marks, G. B. 2009. Weighted road density: A simple way of assigning traffic-related air pollution exposure. *Atmospheric Environment* 43:5009-5014.

190. Rothman, K. J., Greenland, S., and Lash, T. L. 2008. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott, Williams & Wilkins.

191. Roux, A. V. D., Merkin, S. S., Hannan, P., Jacobs, D. R., and Kiefe, C. I. 2003. Area characteristics, individual-level socioeconomic indicators, and smoking in young adults - The Coronary Artery Disease Risk Development in Young Adults Study. *American Journal of Epidemiology* 157:315-326.

192. Rucker, G., Schwarzer, G., Carpenter, J. R., and Schumacher, M. 2008. Undue reliance on I2 in assessing heterogeneity may mislead. *Bmc Medical Research Methodology* 8.

193. Rushton, L. 2009. Retrospective exposure assessment in environmental epidemiology. *Occupational and Environmental Medicine* 66:572-573.

194. Rushton, L., Hutchings, S. J., Fortunato, L., Young, C., Evans, G. S., Brown, T., Bevan, R., Slack, R., Holmes, P., Bagga, S., Cherrie, J. W., and Van Tongeren, M. 2012. Occupational cancer burden in Great Britain. *British Journal of Cancer* 107:S3-S7.

195. Ryan, P. H., and LeMasters, G. K. 2007. A review of land-use regression for characterizing intraurban air models pollution exposure. *Inhalation Toxicology* 19:127-133.

196. Sahmel, J., Devlin, K., Paustenbach, D., Hollins, D., and Gaffney, S. 2010. The role of exposure reconstruction in occupational human health risk assessment: Current methods and a recommended framework. *Critical Reviews in Toxicology* 40:799-843.

197. Samet, J. M., Avila-Tang, E., Boffetta, P., Hannan, L. M., Olivo-Marston, S., Thun, M. J., and Rudin, C. M. 2009. Lung cancer in never smokers: Clinical epidemiology and environmental risk factors. *Clinical Cancer Research* 15:5626-5645.

198. Sanmartin, C., Fines, P., Khan, S., Peters, P., Tjepkema, M., Bernier, J., and Burnett, R. 2013. Modelling risk factor information for linked census data: The case of smoking. *Health Reports* 24:9-15.

199. Savitz, D. A. 2012. A niche for ecologic studies in environmental epidemiology. *Epidemiology* 23:53-54.

200. Schikowski, T., Adam, M., Marcon, A., Cai, Y. T., Vierkotter, A., Carsin, A. E., Jacquemin, B., Al Kanani, Z., Beelen, R., Birk, M., Bridevaux, P. O., Brunekeef, B., Burney, P., Cirach, M., Cyrys, J., de Hoogh, K., de Marco, R., de Nazelle, A., Declercq, C., Forsberg, B., Hardy, R., Heinrich, J., Hoek, G., Jarvis, D., Keidel, D., Kuh, D., Kuhlbusch, T., Migliore, E., Mosler, G., Nieuwenhuijsen, M. J., Phuleria, H., Rochat, T., Schindler, C., Villani, S., Tsai, M. Y., Zemp, E., Hansell, A., Kauffmann, F., Sunyer, J., Probst-Hensch, N., Kramer, U., and Kunzli, N. 2014a. Association of ambient air pollution with the prevalence and incidence of COPD. *European Respiratory Journal* 44:614-626.

201. Schikowski, T., Mills, I. C., Anderson, H. R., Cohen, A., Hansell, A., Kauffmann, F., Kramer, U., Marcon, A., Perez, L., Sunyer, J., Probst-Hensch, N., and Kunzli, N. 2014b. Ambient air pollution: a cause of COPD? *European Respiratory Journal* 43:250-263.

202. Schootman, M., Sterling, D. A., Struthers, J., Yan, Y., Laboube, T., Emo, B., and Higgs, G. 2007. Positional accuracy and geographic bias of four methods of geocoding in epidemiologic research. *Annals of Epidemiology* 17:464-470.

203. Schweizer, C., Edwards, R. D., Bayer-Oglesby, L., Gauderman, W. J., Ilacqua, V., Jantunen, M. J., Lai, H. K., Nieuwenhuijsen, M., and Kunzli, N. 2007. Indoor time-microenvironment-activity patterns in seven regions of Europe. *Journal of Exposure Science and Environmental Epidemiology* 17:170-181.

204. Sheehan, M. C., Lam, J., Navas-Acien, A., and Chang, H. H. 2016. Ambient air pollution epidemiology systematic review and meta-analysis: A review of reporting and methods practice. *Environment International* 92-93:647-656.

205. Sheppard, L., Burnett, R. T., Szpiro, A. A., Kim, S. Y., Jerrett, M., Pope, C. A., and Brunekreef, B. 2012. Confounding and exposure measurement error in air pollution epidemiology. *Air Quality Atmosphere and Health* 5:203-216.

206. Shi, X. Q., Zhou, S. H., Wang, Z. X., Zhou, Z. C., and Wang, Z. Z. 2008. CYP1A1 and GSTM1 polymorphisms and lung cancer risk in Chinese populations: A meta-analysis. *Lung Cancer* 59:155-163.

207. Siemiatycki, J., Krewski, D., Shi, Y. L., Goldberg, M. S., Nadon, L., and Lakhani, R. 2003. Controlling for potential confounding by occupational exposures. *Journal of Toxicology and Environmental Health-Part A* 66:1591-1603.

208. Silverman, D. T., Samanic, C. M., Lubin, J. H., Blair, A. E., Stewart, P. A., Vermeulen, R., Coble, J. B., Rothman, N., Schleiff, P. L., Travis, W. D., Ziegler, R. G., Wacholder, S., and Attfield, M. D. 2012. The Diesel Exhaust in Miners Study: A nested case-control study of lung cancer and diesel exhaust. *Journal of the National Cancer Institute* 104:855-868.

209. Smith, J. D., Mitsakou, C., Kitwiroon, N., Barratt, B. M., Walton, H. A., Taylor, J. G., Anderson, H. R., Kelly, F. J., and Beevers, S. D. 2016. London hybrid exposure model: Improving human exposure estimates to NO2 and PM2.5 in an urban setting. *Environmental Science & Technology* 50:11760-11768.

210. Soerjomataram, I., Oomen, D., Lemmens, V., Oenema, A., Benetou, V., Trichopoulou, A., Coebergh, J. W., Barendregt, J., and de Vries, E. 2010. Increased consumption of fruit and vegetables and future cancer incidence in selected European countries. *European Journal of Cancer* 46:2563-2580.

211. Son, J. Y., Bell, M. L., and Lee, J. T. 2010. Individual exposure to air pollution and lung function in Korea Spatial analysis using multiple exposure approaches. *Environmental Research* 110:739-749.

212. Soulakova, J. N., Hartman, A. M., Liu, B. M., Willis, G. B., and Augustine, S. 2012. Reliability of adult self-reported smoking history: Data from the tobacco use supplement to the current population survey 2002-2003 cohort. *Nicotine & Tobacco Research* 14:952-960.

213. Steenland, K., Loomis, D., Shy, C., and Simonsen, N. 1996. Review of occupational lung carcinogens. *American Journal of Industrial Medicine* 29:474-490.

214. Stewart, P. A., and Herrick, R. F. 1991. Issues in performing retrospective exposure assessment. *Applied Occupational and Environmental Hygiene* 6:421-427.

215. Stinn, W., Teredesai, A., Anskeit, E., Rustemeier, K., Schepers, G., Schnell, P., Haussmann, H. J., Carchman, R. A., Coggins, C. R. E., and Reininghaus, W. 2005. Chronic nose-only inhalation study in rats, comparing room-aged sidestream cigarette smoke and diesel engine exhaust. *Inhalation Toxicology* 17:549-576.

216. Sun, Y., Bochmann, F., Nold, A., and Mattenklott, M. 2015. Diesel exhaust exposure and the risk of lung cancer - A review of the epidemiological evidence. *International Journal of Epidemiology* 44:157.

217. Swaen, G., and van Amelsvoort, L. 2009. A weight of evidence approach to causal inference. *Journal of Clinical Epidemiology* 62:270-277.

218. Szpiro, A. A., Paciorek, C. J., and Sheppard, L. 2011a. Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology* 22:680-685.

219. Szpiro, A. A., Sheppard, L., and Lumley, T. 2011b. Efficient measurement error correction with spatially misaligned data. *Biostatistics* 12:610-623.

220. Tardon, A., Lee, W. J., Delgado-Rodriguez, M., Dosemeci, M., Albanes, D., Hoover, R., and Blair, A. 2005. Leisure-time physical activity and lung cancer: a meta-analysis. *Cancer Causes & Control* 16:389-397.

221. Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., and Frame, M. 2011. Data sharing by scientists: Practices and perceptions. *Plos One* 6.

222. Teschke, K., Olshan, A. F., Daniels, J. L., De Roos, A. J., Parks, C. G., Schulz, M., and Vaughan, T. L. 2002. Occupational exposure assessment in case-control studies: opportunities for improvement. *Occupational and Environmental Medicine* 59:575-593.

223. Thomas, D. C. 2009. *Statistical Methods in Environmental Epidemiology*. New York, NY: Oxford University Press.

224. Tielemans, E., Marquart, H., De Cock, J., Groenewold, M., and Van Hemmen, J. 2002. A proposal for evaluation of exposure data. *Annals of Occupational Hygiene* 46:287-297.

225. Tolbert, P. E., Klein, M., Peel, J. L., Sarnat, S. E., and Sarnat, J. A. 2007. Multipollutant modeling issues in a study of ambient air quality and emergency department visits in Atlanta. *Journal of Exposure Science and Environmental Epidemiology* 17:S29-S35.

226. Tomczak, A., Miller, A. B., Weichenthal, S. A., To, T., Wall, C., van Donkelaar, A., Martin, R. V., Crouse, D. L., and Villeneuve, P. J. 2016. Long-term exposure to fine particulate matter air pollution and the risk of lung cancer among participants of the Canadian National Breast Screening Study. *International Journal of Cancer* 139:1958-1966.

227. Tsai, M. Y., Hoek, G., Eeftens, M., de Hoogh, K., Beelen, R., Beregszaszi, T., Cesaroni, G., Cirach, M., Cyrys, J., de Nazelle, A., de Vocht, F., Ducret-Stich, R., Eriksen, K., Galassi, C., Grazuleviciene, R., Grazulevicius, T., Grivas, G., Gryparis, A., Heinrich, J., Hoffmann, B., Iakovides, M., Keuken, M., Kramer, U., Kunzli, N., Lanki, L., Madsen, C., Meliefste, K., Merritt, A. S., Molter, A., Mosler, G., Nieuwenhuijsen, M. J., Pershagen, G., Phuleria, H., Quass, U., Ranzi, A., Schaffner, E., Sokhi, R., Stempfelet, M., Stephanou, E., Sugiri, D., Taimisto, P., Tewis, M., Udvardy, O., Wang, M., and Brunekreef, B. 2015. Spatial variation of PM elemental composition between and within 20 European study areas - Results of the ESCAPE project. *Environment International* 84:181-192.

228. Tsoi, C. T., and Tse, L. A. 2012. Professional drivers and lung cancer: A systematic review and meta-analysis. *Occupational and Environmental Medicine* 69:831-836.

229. Turner, M. C., Cohen, A., Jerrett, M., Gapstur, S. M., Diver, W. R., Pope, C. A., Krewski, D., Beckerman, B. S., and Samet, J. M. 2014. Interactions between cigarette smoking and fine particulate matter in the risk of lung cancer mortality in Cancer Prevention Study II. *American Journal of Epidemiology* 180:1145-1149.

230. Turner, M. C., Krewski, D., Pope, C. A., Chen, Y., Gapstur, S. M., and Thun, M. J. 2011. Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *American Journal of Respiratory and Critical Care Medicine* 184:1374-1381.

231. USEPA. 2002. Health Assessment Document for Diesel Engine Exhaust. EPA/600/8-90/057F: National Center for Environmental Assessment, U.S. Environmental Protection Agency, Washington, DC.

232. Vermeulen, R., Silverman, D. T., Garshick, E., Vlaanderen, J., Portengen, L., and Steenland, K. 2014. Exposure-response estimates for diesel engine exhaust and lung cancer mortality based on data from three occupational cohorts. *Environmental Health Perspectives* 122:172-177.

233. Vienneau, D., de Hoogh, K., Beelen, R., Fischer, P., Hoek, G., and Briggs, D. 2010. Comparison of land-use regression models between Great Britain and the Netherlands. *Atmospheric Environment* 44:688-696.

234. Vineis, P., Hoek, G., Krzyzanowski, M., Vigna-Taglianti, F., Veglia, F., Airoldi, L., Autrup, H., Dunning, A., Garte, S., Hainaut, P., Malaveille, C., Matullo, G., Overvad, K., Raaschou-Nielsen, O., Clavel-Chapelon, F., Linseisen, J., Boeing, H., Trichopoulou, A., Palli, D., Peluso, M., Krogh, V., Tumino, R., Panico, S., Bueno-de-Mesquita, H. B., Peeters, P. H., Lund, E. E., Gonzalez, C. A., Martinez, C., Dorronsoro, M., Barricarte, A., Cirera, L., Quiros, J. R., Berglund, G., Forsberg, B., Day, N. E., Key, T. J., Saracci, R., Kaaks, R., and Riboli, E. 2006.Air pollution and risk of lung cancer in a prospective study in Europe. 169-174.WOS:000237430200024

235. Vlaanderen, J., Portengen, L., Schuz, J., Olsson, A., Pesch, B., Kendzia, B., Stucker, I., Guida, F., Bruske, I., Wichmann, H. E., Consonni, D., Landi, M. T., Caporaso, N., Siemiatycki, J., Merletti, F., Mirabelli, D., Richiardi, L., Gustavsson, P., Plato, N., Jockel, K. H., Ahrens, W., Pohlabeln, H., Tardon, A., Zaridze, D., Field, J. K., Mannetje, A. 'T., Pearce, N., McLaughlin, J., Demers, P., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Dumitru, R. S., Bencko, V., Foretova, L., Janout, V., Boffetta, P., Forastiere, F., Bueno-de-Mesquita, B., Peters, S., Bruning, T., Kromhout, H., Straif, K., and Vermeulen, R. 2014. Effect modification of the association of cumulative exposure and cancer risk by intensity of exposure and time since exposure cessation: A flexible method applied to cigarette smoking and lung cancer in the SYNERGY study. *American Journal of Epidemiology* 179:290-298.

236. von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gotzsche, P. C., and Vandenbroucke, J. P. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement guidelines for reporting observational studies. *Epidemiology* 18:800-804.

237. Wang, J. J., Zheng, Y., Sun, L., Wang, L., Yu, P. B., Li, H. L., Tian, X. P., Dong, J. H., Zhang, L., Xu, J., Shi, W., and Ma, T. Y. 2011. CYP1A1 Ile(462)Val polymorphism and susceptibility to lung cancer: a meta-analysis based on 32 studies. *European Journal of Cancer Prevention* 20:445-452.

238. Wang, L. F., Zhang, X. W., Liu, J., Shen, L., and Li, Z. Q. 2014a. Tea consumption and lung cancer risk: A meta-analysis of case-control and cohort studies. *Nutrition* 30:1122-1127.

239. Wang, M., Beelen, R., Basagana, X., Becker, T., Cesaroni, G., de Hoogh, K., Dedele, A., Declercq, C., Dimakopoulou, K., Eeftens, M., Forastiere, F., Galassi, C., Grazuleviciene, R., Hoffmann, B., Heinrich, J., Iakovides, M., Kunzli, N., Korek, M., Lindley, S., Molter, A., Mosler, G., Madsen, C., Nieuwenhuijsen, M., Phuleria, H., Pedeli, X., Raaschou-Nielsen, O., Ranzi, A., Stehanou, E., Sugiri, D., Stempfelet, M., Tsai, M. Y., Lanki, T., Udvardy, O., Varro, M. J., Wolf, K., Weinmayr, G., Yli-Tuomi, T., Hoek, G., and Brunekreef, B. 2013. Evaluation of land use regression models for NO2 and particulate matter in 20 European study areas: The ESCAPE project. *Environmental Science & Technology* 47:4357-4364.

240. Wang, M., Beelen, R., Bellander, T., Birk, M., Cesaroni, G., Cirach, M., Cyrys, J., de Hoogh, K., Declercq, C., Dimakopoulou, K., Eeftens, M., Eriksen, K. T., Forastiere, F., Galassi, C., Grivas, G., Heinrich, J., Hoffmann, B., Ineichen, A., Korek, M., Lanki, T., Lindley, S., Modig, L., Molter, A., Nafstad, P., Nieuwenhuijsen, M. J., Nystad, W., Olsson, D., Raaschou-Nielsen, O., Ragettli, M., Ranzi, A., Stempfelet, M., Sugiri, D., Tsai, M. Y., Udvardy, O., Varro, M. J., Vienneau, D., Weinmayr, G., Wolf, K., Yli-Tuomi, T., Hoek, G., and Brunekreef, B. 2014b. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. *Environmental Health Perspectives* 122:843-849.

241. Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., and Brunekreef, B. 2012. Systematic evaluation of land use regression models for NO2. *Environmental Science & Technology* 46:4481-4489.

242. Wang, Y. P., Li, F., Wang, Z. Z., Qiu, T., Shen, Y., and Wang, M. Z. 2015. Fruit and vegetable consumption and risk of lung cancer: A dose-response meta-analysis of prospective cohort studies. *Lung Cancer* 88:124-130.

243. Warheit, D. B., Kreiling, R., and Levy, L. S. 2016. Relevance of the rat lung tumor response to particle overload for human risk assessment-Update and interpretation of new data since ILSI 2000. *Toxicology* 374:42-59.

244. Weed, D. L., and Gorelic, L. S. 1996. The practice of causal inference in cancer epidemiology. *Cancer Epidemiology Biomarkers & Prevention* 5:303-311.

245. Weichenthal, S., Crouse, D. L., Pinault, L., Godri-Pollitt, K., Lavigne, E., Evans, G., van Donkelaar, A., Martin, R. V., and Burnett, R. T. 2016. Oxidative burden of fine particulate air pollution and risk of cause-specific mortality in the Canadian Census Health and Environment Cohort (CanCHEC). *Environmental Research* 146:92-99.

246. Wells, A. J., English, P. B., Posner, S. F., Wagenknecht, L. E., and Perez-Stable, E. J. 1998. Misclassification rates for current smokers misclassified as nonsmokers. *American Journal of Public Health* 88:1503-1509.

247. Wu, J., Wilhelm, M., Chung, J., and Ritz, B. 2011. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. *Environmental Research* 111:685-692.

248. Xie, Y., Qin, J., Nan, G., Huang, S., Wang, Z., and Su, Y. 2016. Coffee consumption and the risk of lung cancer: an updated meta-analysis of epidemiological studies. *European Journal of Clinical Nutrition* 70:199-206.

249. Yang, P., Cerhan, J. R., Vierkant, R. A., Olson, J. E., Vachon, C. M., Limburg, P. J., Parker, A. S., Anderson, K. E., and Sellers, T. A. 2002. Adenocarcinoma of the lung is strongly associated with cigarette smoking: Further evidence from a prospective study of women. *American Journal of Epidemiology* 156:1114-1122.

250. Yorifuji, T., Kashima, S., Tsuda, T., Ishikawa-Takata, K., Ohta, T., Tsuruta, K., and Doi, H. 2013. Long-term exposure to traffic-related air pollution and the risk of death from hemorrhagic stroke and lung cancer in Shizuoka, Japan. *Science of the Total Environment* 443:397-402.

251. Young, R. P., Hopkins, R. J., Christmas, T., Black, P. N., Metcalf, P., and Gamble, G. D. 2009. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *European Respiratory Journal* 34:380-386.

252. Young, S. S., and Karr, A. 2011. Deming, data and observational studies: A process out of control and needing fixing. *Significance* 8:116-120.

253. Zandbergen, P. A. 2007. Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *Bmc Public Health* 7:doi:10.1186/1471-2458-7-37.

254. Zandbergen, P. A. 2008. A comparison of address point, parcel and street geocoding techniques. *Computers Environment and Urban Systems* 32:214-232.

255. Zeger, S. L., Dominici, F., McDermott.A., and Samet, J. M. 2008. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000-2005). *Environmental Health Perspectives* 116:1614-1619.

256. Zou, B., Wilson, J. G., Zhan, F. B., and Zeng, Y. N. 2009. Air pollution exposure assessment methods utilized in epidemiological studies. *Journal of Environmental Monitoring* 11:475-490.

Concawe
Environmental Science
for European Refining

Concawe
Boulevard du Souverain 165
B-1160 Brussels
Belgium

Tel: +32-2-566 91 60
Fax: +32-2-566 91 81
e-mail: info@concawe.eu
http://www.concawe.eu

ISBN 978-2-87567-096-0