From THE INSTITUTE OF ENVIRONMENTAL MEDICINE
Karolinska Institutet, Stockholm, Sweden

CARDIOVASCULAR AND RESPIRATORY EFFECTS OF AIR POLLUTION – APPLICATION OF DIFFERENT OBSERVATIONAL STUDY DESIGNS AND ANALYSIS APPROACHES

Niklas Berglind

Karolinska Institutet

Stockholm 2017

CARDIOVASCULAR AND RESPIRATORY EFFECTS OF AIR POLLUTION – APPLICATION OF DIFFERENT OBSERVATIONAL STUDY DESIGNS AND ANALYSIS APPROACHES
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Niklas Berglind

Principal Supervisor:
Prof. Tom Bellander
Karolinska Institutet
Institute of Environmental Medicine
Division of Environmental Epidemiology

Co-supervisor:
Assoc. Prof. Fredrik Nyberg
University of Gothenburg
Sahlgrenska Academy
Institute of Medicine
Division of public health
and community medicine

Opponent:
Dr. Anna Oudin
Umeå University
Department of Public Health and clinical Medicine
Division of Occupational Medicine

Examination Board:
Prof. Paul Dickman
Karolinska Institutet
Department of Medical Epidemiology
and Biostatistics
Division of Biostatistics

Prof. Hans-Christen Hansson
Stockholm University
Department of Environmental Science and
Analytical Chemistry
Division of Atmospheric Science

Prof. Gunnar Engström
Lund University
Department of Clinical Sciences, Malmö
Division of Epidemiology

To Anna

ABSTRACT

The effect of air pollution exposure on human health has been a topic of an increasing volume of research over the past half century. Due to the ubiquitous nature of ambient air pollution exposure and the comparatively weak effects that can be expected, one of the most viable and most utilized avenues of research has been to conduct large scale observational studies. This thesis presents four such studies covering several health outcomes in a variety of populations using a number of different observational study designs and analysis methods. The overarching aim of the thesis was to elucidate challenges that come with observational research on health effects of air pollution and to assess how the studies and their designs address those challenges. In addition, the individual works were described, interpreted, and assessed on the basis of how they met their respective objectives.

BAMSE is an ongoing birth cohort in Stockholm of over 4,000 children who were recruited between 1994 and 1996 and are followed with respect to their respiratory health and development of allergies, as well as their environmental exposures. An association was seen between exposure to traffic generated air pollution during their first year of life and early onset persistent wheezing during their first four years of life, as well as lung function (measured by peak expiratory flow) and allergic sensitization at age 4.

HEAPSS was a European five-center register-based study of more than 25,000 first-time acute myocardial infarction (MI) survivors who were followed with respect to re-infarction and mortality, linking the health events to the daily air pollution levels. Daily all-cause mortality among the MI survivors was linked to increased levels of both ultrafine and larger particles during the days preceding the deaths.

ALVA was a study of 211 patients in Stockholm and Gothenburg with implantable cardioverter defibrillators, devices implanted in the chest of the patients and designed to recognize and treat ventricular arrhythmias. Since the devices record the exact time when they have treated an arrhythmia it was possible to very precisely determine the time of an event and link it to the air pollution levels immediately before that time. An association was indicated between particulate air pollution levels in the two hours immediately preceding the event and the onset of ventricular arrhythmia.

The Swedish Onset study comprised 660 first-time MI survivors in Stockholm and was designed to identify triggers for the onset of MI through very detailed interviews conducted shortly after the event about the time leading up to the MI. This allowed for very accurate determination of the exact time of the events and that information could be linked to the air pollution levels preceding that time. We did not observe any association between air pollution exposure and the onset of first-time non-fatal MI.

Major challenges included study design, exposure and outcome assessment, potential confounding and selection bias, as well as study power and analytical strategies. All four studies used state-of-the-art designs and were conducted with great care for data quality and

effort was made to rule out that the observed results were due to misrepresentation of exposures or outcomes, uncontrolled confounding, selection bias, or other unintended effects. All four studies had sample sizes that were comparatively small, which yielded rather imprecise estimates of the weak air pollution effects, and each study have some results that cannot easily be explained. In summary, all four studies met their respective objectives to assess how air pollution exposure affect the respective health outcomes and were designed, conducted, analyzed, presented, and interpreted in a manner that allow them to effectively contribute to the collective scientific evidence of health effects of air pollution.

PREFACE

Dear reader,

What you have in front of you is my doctoral thesis in epidemiology. That's right – epidemiology! Those who knew me when I first started working in the public health field might find it ironic that I have pursued a doctorate in that particular discipline since I used to be somewhat critical of it. As an angry young statistician I was heard accusing epidemiology of being a sorry excuse of a science where overpaid physicians with little or no understanding of statistics systematically abused that fine science to justify horrible over-interpretations and baseless speculation to forward an alarmist agenda. This was a long time ago and I am now *much* older, hopefully a little wiser, slightly less angry, and arguably not even a statistician anymore, and I have completely revised my stance and is now ready to proudly put "Doctor of Epidemiology" on my business card. Here is the story of how I got there:

The job title of my first employment straight out of college was "Assistant to the Epidemiologist" (loosely interpreted from Swedish "Epidemiologassistent"). At the time, I did not know what an epidemiologist was, but I soon came to the conclusion that it was a medical doctor who mostly did statistics (with statisticians as their assistants). Being the angry young statistician that I was, I started to preach to anyone that would listen (and there weren't many who would), that epidemiology should be considered a branch of biostatistics, and I was very upset that epidemiologists were almost exclusively non-statisticians.

Despite this apparent flaw, I worked full time in the field of environmental epidemiology for almost 10 years and enjoyed it tremendously. During my time at the Department of Environmental Medicine I got the chance to work with many great scientists and I made a lot of friends. But when I left the field of epidemiology and embarked on a career as a clinical trials statistician in the pharmaceutical industry in the US, the contrast was quite striking and I was intimidated at first. I came from being the person who knew the most about statistics (for long periods the only statistician in the department) to being the person who knew the least (one of only a handful of statistician without a PhD in a department of more than 50). I was self-conscious of my own short-comings, not only did I lack a PhD but my decade of experience was from a field that most statisticians were quite skeptical about, even though they generally did not know much about it.

Soon, however, I found that my lack of formal training and experience in the mathematical realm was outweighed by my experiences and acquired skills in collaborative research. I had gained proficiency in the arts of compromise and negotiation, I had acquired a skill in understanding other researcher's viewpoints and developed strategies for making others understand mine, and I had a lot of practical experience of walking the thin line of trying to make researchers from other backgrounds understand the sometimes intimidating and complex concepts in statistics without simplifying it to the point where it is corrupted. All these things have turned out to be invaluable skills during my years in the pharmaceutical industry.

I lived and worked in America for six years before I returned to my homeland in 2011 trying to wrap up this thesis (it only took 6 more years…). Returning to Sweden from the U.S. and to epidemiology from clinical trials statistics was an interesting experience. First, the weather in Sweden is even worse than I remembered it. Second, with my new career have come new experiences and new perspectives on certain things and I was worried that I would no longer be able to stand behind these publications and defend what I had contributed to. Fortunately, I had no reason to worry. I am proud of each of these papers and I believe they do justice to the effort put in by those involved in the execution of the studies and the study subjects who gave their consent to participate. This brings me to another thing that I used to preach, and still do at times: There are no good or bad results, only good or bad studies. The results are what they are and as long as they come from a good study they contribute to the collective knowledge and one can and should be happy with them. These papers are all based on good studies. They were all meticulously planned and well-executed and are great contributions to the scientific knowledge.

I no longer believe that epidemiology should be considered a branch of biostatistics. In fact, it is my firm conviction that epidemiology, like most applied sciences, works best when it is a collaborative effort where scientists of different backgrounds, philosophies, and specialties come together and contribute to the end result. However, I do maintain that the statisticians should be equal scientific partners to the physicians rather than their assistants.

Niklas Berglind, Västra Frölunda, April 2017

LIST OF SCIENTIFIC PAPERS

1. Nordling E, **Berglind N,** Melén E, Emenius G, Hallberg J, Nyberg F, Pershagen G, Svartengren M, Wickman M, Bellander T.
   Traffic-related air pollution and childhood respiratory symptoms, function and allergies. *Epidemiology*. 2008;19(3):401-8.

2. **Berglind N**, Bellander T, Forastiere F, von Klot S, Aalto P, Elosua R, Kulmala M, Lanki T, Löwel H, Peters A, Picciotto S, Salomaa V, Stafoggia M, Sunyer J, Nyberg F; HEAPSS Study Group.
   Ambient air pollution and daily mortality among survivors of myocardial infarction. *Epidemiology*. 2009;20(1):110-8.

3. Ljungman PL, **Berglind N**, Holmgren C, Gadler F, Edvardsson N, Pershagen G, Rosenqvist M, Sjögren B, Bellander T.
   Rapid effects of air pollution on ventricular arrhythmias. *Eur Heart J*. 2008;29(23):2894-901.

4. **Berglind N**, Ljungman P, Möller J, Hallqvist J, Nyberg F, Rosenqvist M, Pershagen G, Bellander T.
   Air Pollution Exposure – A Trigger for Myocardial Infarction? *Int. J. Environ. Res. Public Health*. 2010, 7(4), 1486-1499.

CONTENTS

LIST OF ABBREVIATIONS

APHEA	Air Pollution and Health: A European Approach
APHENA	Air pollution and health: A European and North American approach
ALVA	Air pollution and Life threatening Ventricular Arrhythmias
BAMSE	Barn, Allergi, Miljö, Stockholm, Epidemiologi (Children, Allergy, Environment, Stockholm, Epidemiology)
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
HEAPSS	Health Effects of Air Pollution on Susceptible Subpopulations
ICD (code)	International classification of diseases
ICD (device)	Implanted cardioverter defibrillator
IgE	Immunoglobulin E
IQR	Interquartile range
MI	Myocardial infarction
MONICA	Multinational MONItoring of trends and determinants in CArdiovascular disease
NMMAPS	The National Morbidity Mortality Air Pollution Stud
NO_2	Nitrogen dioxide
OR	Odds ratio
O_3	Ozone
PEF	Peak expiratory flow
PM_{10}	Particulate matter with an aerodynamic diameter of up to 10 µm
$PM_{2.5}$	Particulate matter with an aerodynamic diameter of up to 2.5 µm
PNC	Particulate number concentration
SHEEP	Stockholm Heart Epidemiology Program
SO_2	Sulphur dioxide
WHO	World Health Organization

INTRODUCTION

HEALTH EFFECTS OF AIR POLLUTION

In London, England in early December 1952, a combination of massive emissions of air pollution, mainly from coal burning, and unique weather conditions led to what is known as the *London Killer Fog,* which over the following several months was responsible for up to 12,000 excess deaths according to some estimates (Bell and Davis 2001). Even though there had been earlier reports in the scientific literature of episodes of increased air pollution levels leading to suspected increases in mortality and morbidity, the sheer scale of the London disaster made it the landmark event that acted as a catalyst for the scientific study of health effects of air pollution. Before the 1980s most of the research was focused on various health effects of extreme air pollution episodes and in the 1980s there was some research done on average air pollution exposure using area-level data (Nyberg and Pershagen 2000). At this time there was a general consensus that very high concentrations of particulate air pollution could be associated with cardiopulmonary disease, but there was disagreement in the scientific community whether or not low to moderate levels had any effect on human health. In the 1990s this all changed after a number of mainly US studies were published that linked surprisingly low air pollution levels to health effects of all levels of severity, ranging from asthma exacerbations and school absences to mortality (Pope and Dockery 2006). The studies reported in the early 1990s were mainly single-city time series studies of short-term effects on mortality but also two major prospective cohort studies: the Harvard Six Cities Study (Dockery, Pope et al. 1993) and the American Cancer Society Study (Pope, Thun et al. 1995) which both reported long-term effects. These studies were followed by a few very large multi-city studies of short-term effects: the European APHEA (Katsouyanni, Zmirou et al. 1997) and APHEA2 (Katsouyanni, Touloumi et al. 2001; Analitis, Katsouyanni et al. 2006) and the American NMMAPS (Samet, Zeger et al. 2000; Dominici, McDermott et al. 2005) all of which showed modest but consistent mortality and morbidity effects at low and moderate air pollution levels.

By this time there was a lot of epidemiological evidence for the existence of health effects of ambient air pollution but there was a sense that these results probably didn't reflect a common risk for everyone in the general population and the next big efforts was towards trying to identify the susceptible subpopulations that would be the most vulnerable. A number of studies have suggested that children, elderly, people of lower socioeconomic status, as well as patients with underlying disease of diabetes, obesity, asthma, COPD, congestive heart disease, and previous myocardial infarction (MI) may be at higher risk of experiencing various adverse effects of air pollution (Ruckerl, Schneider et al. 2011). In parallel to the epidemiologic studies there has also been an increasing interest in elucidating the mechanisms and physiological pathways that lie behind the effects seen in the observational studies. This has been done through human, animal, and in-vitro experimental studies (Anderson, Thundiyil et al. 2011) and gene-environment interaction studies (Zanobetti,

Baccarelli et al. 2011). All these different types of studies are necessary in order to tease out the whole picture of health effects of air pollution and one type of study does not replace the other.

CHALLENGES

The World Health Organization has stated that "Action to improve air quality, and in this way to improve people's health, starts by recognizing the pollutants and pollution sources that contribute most to population exposure. The foundation for improving public health here is laid by risk assessment of air pollution and by legislation based on relevant scientific information." (World Health Organization. Regional Office for Europe. 1999). It is indeed quite a long journey from that first recognition of a health effect to actually improved public health. There is a first hypothesis-generating study that identifies an exposure as being associated with a health outcome. The results of that study prompts other studies to confirm and quantify the extent of the association which then provide the basis for legislators and policymakers to make laws, policies, and recommendations that subsequently reduce the publics' exposure, ultimately leading to improvement in the publics' health. This thesis is about one part of that chain from hypothesis to improved public health: the conduct and interpretation of a single observational study. It is not about the question of causality since that is a much broader question that is best answered once there is accumulated data from multiple observational and experimental studies and a systematic and independent approach is taken to synthesize the available information. The rationale for conducting an observational study, as well as most other studies, is based on a belief that there is a "true association" out there, be it positive or negative or null, and the study is conducted in order to achieve the best estimate possible of that true association. In my opinion, the primary responsibility of the author of an individual study is to present that estimate as clearly and as transparently as possible and to deliver it along with a judgment about how well the estimate can be trusted to represent the true association.

There are plenty of challenges in trying to come up with the best estimate possible of an association in an observational study. Some of these challenges stem from the very nature of trying to observe and analyze real life without the tools of randomization, double-blindness, and placebo-control to comfort us, and some stem from trying to simplify reality enough to make sense of the data without simplifying it so much that it no longer represents reality. These challenges include the handling of biases from misrepresentation of exposure and outcome, confounding, as well as the important, and often overlooked considerations about interpretation and presentation of the results.

Misrepresentation of exposure and outcome

The World Health Organization defines exposure as "a contact over space and time between a person and one or more biological, chemical or physical agents. […] Exposure is described by the concentration of an environmental agent in the carrier medium at the point of contact with the body together with the duration of the contact." They further explain that "Exposure

2

assessment identifies and defines the exposures that occur […] in human populations."
(World Health Organization. Regional Office for Europe. 2000). Reading this definition
(crystal clear as it is…) and thinking about it in the context of air pollution exposure, it is
evident that as exposures goes, ambient air pollution is a difficult one to assess. As the word
ambient suggests it can be found all around us, in the very air we all breathe. The level of
exposure is affected of course by its sources but also by the weather, geographical elements,
and it penetrates buildings to various extents. Exposure occurs just about everywhere which
means that the contrast in exposure between individuals is often low and accurate
measurement and estimation of exposure is important. To estimate individual exposure,
ideally one would place a measurement device immediately in front of each subject's
breathing organs to get the most accurate measurements possible of what the subject is
exposed to. This is often impractical, especially in large studies with long duration. Instead
one has to rely on exposure assessment methods or models that utilize approximations and
extrapolations in different ways (Zou, Wilson et al. 2009). These will invariably lead to the
subjects' exposure not being accurately represented. This is a well-known bias that is
extensively discussed in the epidemiology literature and is often referred to as
misclassification (or misrepresentation) *of exposure,* one component of the broader category
of *information bias* (Rothman and Greenland 1998). Misrepresentation of exposure will
affect the observed associations, and they can be affected in a number of different ways. If the
misrepresentation is somehow connected to the health outcomes or how we measure these,
known as differential misrepresentation, there is a risk of observing an association between
exposure and health that is really due to biased measurement, or of masking a real effect. If
the misrepresentation is non-differential, i.e. independent of true as well as measured disease
status, it will create noise that may obscure the real association and will, on average, result in
weakened estimated measures of association. The discussion section of many
epidemiological articles contains a claim of non-differential misrepresentation of exposure
leading to dilution of the effect, especially in cases where the researchers have failed to find
an expected effect or found one weaker than expected. Due to the complexity of the reality
we are trying to observe and the interconnectedness of the human behavior and its
environment, it is often hard to convincingly support a claim that any representation of
exposure is completely independent of disease status and such claims may sometimes be
expressions of wishful thinking rather than plausible assumptions about reality. That doesn't
mean that there aren't biases that dilute the effect; just that one has to be vigilant about
exploring them and not summarily dismiss all biases that stem from misrepresentation of
exposure as being non-differential and thus always biased towards the null. It has also been
shown that non-differential misrepresentation of exposure may indeed bias *away* from the
null in certain circumstances (Dosemeci, Wacholder et al. 1990).

Misrepresentations can also occur when we measure health outcomes, this misrepresentation
has similar effects on the estimates of association as misrepresentation of exposure. Non-
differential misrepresentation tends in most cases to bias towards the null as for that of
exposure. However, an important difference is that in many cases non-differential

misrepresentation of outcome has little effect on relative measures of associations and results in a bias towards the null for absolute measures (Rothman and Greenland 1998).

Misrepresentation of both exposure and outcome are prominently featured in air pollution epidemiology because of the ambient nature of the exposure, the difficulties of measuring personal exposure, and the need for large studies which often comes with less detailed information about outcomes. As mentioned before, misrepresentation of exposure and outcome are well-known biases in epidemiology and the discussion section of most articles that present the results of observational studies have at least some discussion related to it.

Confounding

Confounding is when a factor C (confounder) is correlated to factor A (exposure) and is causally related to B (outcome) so that when the relationship between factors A and B is studied without considering C, the effect of A is estimated incorrectly. In the words of Drs. Rothman and Greenland: "the apparent effect of the exposure of interest is distorted because the effect of an extraneous variable is mistaken for or mixed with the actual exposure effect" (Rothman and Greenland 1998).

In the setting of the observational study, our best way to handle the challenge of confounding is usually through the design of the study. Any potential confounding that can be controlled for by design is one less problem to worry about. However, rare is the study design that can do away with all potential confounding so in one form or another it is usually necessary to handle potential confounding in the analysis, with either a matching procedure like propensity score matching, or directly in the statistical model (Sturmer, Joshi et al. 2006). A regression model can provide estimates for an association between the dependent variable and an independent that is adjusted for a number of other variables. But there is a limit to how many other variables can be included in the model. The limit is partly dependent on the number of observations but also on how the different variables are interconnected. There is also a small penalty to pay for each extra variable in terms of loss of degrees of freedom and the inclusion of variables that do not contribute can introduce unwanted noise (Hosmer and Lemeshow 2000).

A statistical model can be thought of as a way to simplify reality to make it easier to understand. A model that uses all available information will usually give a better picture of reality (have a better fit) than one that only uses some of the available information. However, the more information that is put in the model the more complex it becomes and it no longer makes reality easier to understand. Statistical model selection involves a tradeoff between fit and simplicity and most researchers in the field realize that there is no one best way to make this tradeoff. For every approach A there is an example where it can be shown to be better than approach B, and another example to show that B is better than A (Forster and Sober 1994). This unfortunate state of affairs leaves every attempt at model selection or model building open to criticisms and it is important to be transparent about the process and be able to justify all choices.

4

Model building, or confounder selection, as it is usually applied in observational studies is a difficult and tedious process of judging for each candidate variable if it should be included in the model or not. It can be done in many different ways, but usually there are a number of criteria for a variable to be included in a model. The main criterion would be whether the variable in the current situation is a confounder or not, i.e. if the estimated effect of the exposure of interest changes based on whether the variable is included or not. Another criterion could be if the variable is a risk factor, i.e. is itself associated with the dependent variable. Variables could also be forced into the model regardless if they meet any of the criteria for different reasons. Design variables that are somehow part of the design of the study should usually be included; the most obvious example being stratification variables. Variables that have been shown to be confounders or risk factors in previous studies, regardless if they qualify in the current study, could also be considered, as could what is sometimes referred to as "political" variables, i.e. variables that reviewers expect to see in any study, like smoking status or gender (most of which are common confounders in many settings). Automatic methods for model selection exist (stepwise regression, backwards elimination, forward inclusion etc. (Rothman and Greenland 1998)), but for most purposes these methods are too blunt. Inclusion is usually based only on p-values for the coefficient, and they will not handle the important confounder criterion. The automatic methods have one great feature, which is that they are insensitive to user input (except for the choice of the covariate set to begin with). This is important since model building generally happens after data is available and if the model building process is completely manual there is a risk that the researcher that does the model building is influenced by the results, consciously or unconsciously, or even more importantly, there may be a *perception* that the researcher is influenced by the results which could cast doubt on the process and in turn the validity of the results.

After going through the labor-intensive process of selecting our statistical models for cofounding control to arrive at the best model possible to estimate the associations at hand, there is still one problem: The statistical model can only adjust for potential confounders that were available in the database. There is always a chance that there are confounders out there that we didn't adjust for since we didn't measure them, they may even be unmeasurable. This is the problem of uncontrolled confounding. Indeed, whenever a study is conducted to observe phenomena as they occur in real life, the unpleasant truth is that reality is tremendously complex and the different aspects of the environment, the society, the people's health, the people's behavior and so on, are all interconnected in ways that are very hard, if not impossible, to fully understand. Trying to disentangle how one particular aspect of the environment (in this case air pollution) affects one particular aspect of people's health (be it respiratory heath, cardiovascular events, or mortality) is truly a formidable task. When conducting an observational study to investigate a particular health effect we observe and measure as many of these factors as we can and then we try to tease out to what extent changes in our exposure of interest co-vary with changes in the health outcome of interest, ruling out the role of all the other measured factors. Even then, due to the aforementioned

complexity of reality, and the fact that it is impossible to measure everything, it is hard for the researcher to know if it is the exposure or something else poorly measured, unmeasured or un-measureable that lies behind the observed association. This is something that researchers who deal with observational studies have to live with and in practice it is likely that the actual problem is quite rare, but since it by definition cannot be ruled out, the threat of it is always present and has to be considered in the interpretation and weighing of the evidence of most observational studies. (Francis, Shea et al. 2006)

Interpretation of results

Arguably the most important task in any study is interpreting the results. It is a great responsibility, since the researcher has the accountability to give the study full justice - for all the people who have worked hard in the execution of the study, and for the studies that were never done or had to be modified because this study received the funding, and for the study subjects who have agreed to subject themselves to all kinds more or less invasive or inconvenient sacrifices in order to contribute to the science. That is a responsibility that should not be taken lightly.

Interpreting the results doesn't just mean explaining them but to fairly assess to what extent they can be trusted. To walk the path that led up to the results and try to decipher how we ended up where we did. Which biases are at play? Are the observed results artifacts of the study design? How could the results be affected if the assumptions of the statistical models do not hold? What is the role of missing data? To what extent can the can the observed results be due to uncontrolled confounding? How does chance play in? It is not enough to *assume* that a bias is negligible, or that biases that are not clearly confounding will always be toward the null, or that all missing data are unrelated to the phenomenon under study, or that all confounding is controlled for. One has to back these assumptions up. Supportive and sensitivity analyses, review of the literature, and logical reasoning will help determine to what extent these assumptions may hold. More often than not, they won't (for example, uncontrolled confounding and chance can never be completely ruled out and the reason for missing data is often related to something that is part of the association under investigation). The important thing is not to avoid problems, that cannot be done, but to figure out if the problems are large or small and in which direction a bias would likely change the point estimate and what the magnitude of change could be.

Results from well-designed and well-executed studies are important to the scientific community, no matter what the results show. It is the responsibility of the researcher to present the results in a clear and transparent fashion so that future researchers can use the results and put them in the context of the totality of evidence. Some results are expected and warmly welcomed by the researcher as they fit neatly into the current ideas of how the world works. Other results are unexpected and are met with disappointment by the researcher since they contradict what was previously "known" or otherwise fail to live up to their expectations. The unexpected results, especially null results, are usually put under rigorous scrutiny to try to find what is wrong with them while the expected results are treated with

fanfare and press conferences and are sometimes even placed in the titles of articles. As understandable as these reactions are, it is important to scrutinize the welcome results just as hard and to not dismiss the unwelcome results as flukes or errors just because they don't deliver what was expected of them. In fact, the greatest leaps of science often have their origins in unexpected results (Dunbar 2012).

Presentation of results

It can be argued that if the reader can't understand the results, the presentation is a failure and in turn the whole study is a failure regardless of the results. It is important to present the results in an appropriate and logical format. Sometimes a graphical presentation is the best choice and sometimes it will only confuse. There is value in using conventions even if the researcher thinks that a novel innovative approach is more clever. E.g., time is perceived by most people to go from left to right, so time should be put on the horizontal axis, or the graph will not be easily understood by most readers (DeSanctis 1984).

Transparency is important in order for the reader to be able to put the results in context. It is important for the reader to know whether an analysis was pre-planned ahead of the researchers getting access to the data or if the analysis was something they came up with while analyzing the data, known as a post hoc analysis. Analyses that were not pre-specified are, on average, more likely to show a strong positive result since the hypothesis it was supposed to address could have been generated by the data. Post hoc analyses can certainly be very valuable and appropriate to include when presenting the results, but it is important that they are clearly identified as such so the reader can judge the likelihood of a bias being present. Supportive and sensitivity analyses mentioned previously can also be important to include. A lot of the work of untangling the biases and trying to figure out to what extent the results can be trusted is mainly to reassure the researchers themselves that it is OK to go ahead and present the results with a clean conscience, but some of the results from supportive or sensitivity analyses can be very helpful for a reader of an article as well. This is especially important in studies that explore new associations or uses novel approaches and there is no established "best way" to analyze and interpret the results. Due to space restrictions, supporting results are usually limited to mentions in the text without showing the supporting data, but as supplemental online appendices are often available in journals today such data could probably be included more often than they are.

Perception is also important and sometimes overlooked. By showing the reader that every effort has been made to try to understand the results and being transparent and allowing the readers to judge for themselves if it was, the research will be perceived as more credible than if the reader is asked to take the researchers word for it, that the results can be trusted.

Caveats are sometimes used when presenting results that are questionable for some reason. This is of course appropriate, there is an absolute obligation to let the reader know if there is something problematic about the results. But, in many cases, even more appropriate may be to *not* present results that are questionable. Caveated results have a tendency to lose their

caveats over time. Something that starts out as "bear in mind that the result of stronger effect in group X is based on very small subgroup analysis in a selected population from an analysis that was not pre-planned and should only be considered hypothesis generating" is likely to end up down the road as "it has been shown that there is a stronger effect in group X". This serves to illustrate that the reader of scientific presentations also has a considerable role and responsibility in the interpretation of data, which should not be forgotten.

PAPERS 1-4

The articles presented in this thesis are four rather disparate research papers and they cover a range of topics in air pollution epidemiology, looking at acute and chronic exposure, pediatric and adult study populations, and a virtual smorgasbord of health endpoints from allergic sensitization and respiratory effects to different cardiac outcomes and death. Published between 2008 and 2010 they all fall under the broad heading of Health Effects of Air Pollution and they are all observational studies. To a varying extent they are all faced with many of the challenges presented above. In the following sections the objectives, methods, and results are briefly described followed by a discussion of each study and study design and how they meet these challenges.

Paper 1 reported the results of analyses of the association between air pollution exposure during the first year of life and health effects at age 4 in the BAMSE cohort. These long awaited results increased the understanding around the role of air pollution in the development of allergic disease and lung function in children.

Paper 2 reported results from the European HEAPSS collaboration, which was initiated to investigate the role of air pollution exposure in the sensitive sub-population of MI survivors. This paper added to the knowledge of mortality effects from ambient air pollution in this frail patient population.

Papers 3 and 4 reported results on acute effects of air pollution and were both unique in the level of accuracy in ascertainment of the time point of the studied events, ventricular arrhythmias and MI respectively.

OBJECTIVES

The overarching aim was to elucidate challenges in using observational studies to assess health effects of air pollution and to assess how the different studies and their designs meet these challenges.

The specific objectives of the work performed were to assess several health effects of air pollution exposure. Specifically to:

1. Assess how exposure to source-specific air pollutants during the first year of life affects wheezing, lung function, and allergic sensitization in children at age 4.
2. Assess how air pollution levels affect daily mortality in myocardial infarction survivors.
3. Assess how very short-term exposure to air pollution affects the risk of ventricular arrhythmia in patients with implanted cardioverter defibrillators.
4. Assess how very short-term exposure to air pollution affects the risk of first-time myocardial infarction.

DESCRIPTION OF THE STUDIES

PAPER 1: BAMSE

The acronym BAMSE stands for Barn Allergi Miljö Stockholm Epidemiologi (Children Allergy Environment Stockholm Epidemiology). It is a prospective birth cohort of children born between 1994 and 1996 in Stockholm, Sweden. The project has multiple aims related to learning more about the development of allergic disease, obstructive lung symptoms and asthmatic disease in children and the role of environmental exposures like pet allergens, indoor and outdoor air quality, as well as genetic and socio-economic factors. The aim for the part of the project presented in Paper 1 was to assess how exposure to source-specific air pollutants during the first year of life affects wheezing, lung function, and sensitization in children at age 4.

More than 30 journal articles using data from the BAMSE cohort have been published since 2002. The articles cover a variety of topics, including: allergy prevention and diagnostics (Wickman, Ahlstedt et al. 2003; Wickman, Melen et al. 2003; Wickman 2004); the development of asthma, allergy, and eczema in children (Almqvist, Egmar et al. 2003; Asarnoj, Ostblom et al. 2008; Saarne, Gronlund et al. 2010; Ballardini, Kull et al. 2012); the genetics of asthma and allergy (Melen, Bruce et al. 2005; Melen, Umerkajeff et al. 2006; Melen, Nyberg et al. 2008); the effects of indoor air quality (Emenius, Svartengren et al. 2004; Emenius, Svartengren et al. 2004; Lannero, Wickman et al. 2006; Lannero, Wickman et al. 2008); and the effects of outdoor air quality (Nordling, Berglind et al. 2008; Gruzieva, Bellander et al. 2012; Schultz, Gruzieva et al. 2012; Gruzieva, Bergstrom et al. 2013).

Materials and methods

A total of 4089 infants were recruited from Child Health Centers, comprising 75% of all eligible children born in predefined areas of four municipalities in Stockholm in 1994-1996. The four municipalities were chosen to represent both urban and different types of suburban environments. Data on parental allergic diseases, pet contact, detailed residential characteristics and socio-economic factors were collected with a postal questionnaire to the parents at recruitment. The median age of the children at recruitment was 2 months. At approximately 1, 2, and 4 years of age the parents received additional questionnaires concerning symptoms related to wheezing and allergic diseases and detailed information on exposure factors. The response rates based on the originally recruited cohort were 96%, 94%, and 91% at ages 1, 2, and 4 respectively. At approximately 4 years of age, 2965 children (73% of the full cohort) attended a clinical investigation at the Department of Environmental Medicine at the Karolinska Hospital, including measurements of lung function and blood sampling.

Exposure to locally emitted air pollution during the children's first year of life was assessed using a methodology developed to estimate long-term source-specific exposure to air pollution, developed originally for a study of long-term exposure to air pollution and lung cancer (Bellander, Berglind et al. 2001). It entails the assignment of geographic coordinates

to subjects' address information (geocoding), and the use of an emission inventory together with dispersion models to map outdoor levels of selected pollutants from selected emission sources over time at the relevant geographical locations. Usually the home address is used, and in this study residential addresses were collected in the questionnaires at recruitment and 1 year, and were geocoded using the geographical information system MapInfo (MapInfo Corporation, Troy, New York).

Emission databases describing traffic-generated nitrogen oxides (traffic-NO_x) and particulate matter less than 10 μm (traffic-PM_{10}), as well as sulfur dioxide from house heating (heating-SO_2), within the county were available for the years 1990 (traffic-NO_x and heating-SO_2 only) and 2000 (traffic-NO_x, heating-SO_2, and traffic-PM_{10}). The emission database for road traffic contains 4600 road segments (sources of road traffic emissions) and for each source there is detailed information on traffic flow, posted speed limits, heavy traffic share, road type, and number of stops per kilometer. From this information and data about emissions from different vehicle classes, emission values were calculated for each source. Sources of domestic heating emissions were modeled as area sources based on estimated heating oil consumption and sulfur content of the heating oil. The spread of the pollutants from the sources were then estimated with a Gaussian dispersion model based on the average distribution of combinations of wind speed, direction and precipitation in synoptic classes for the whole surface of Stockholm County. In addition, because the air pollution levels in the city also depend on very local traffic conditions, for narrow high-traffic streets with buildings on both sides, a street canyon contribution was added for addresses in the most polluted street segments in the city center (Bellander, Berglind et al. 2001).

The geographical coordinates of the residential addresses were matched to the corresponding air pollution levels from the dispersion models at each geographic location. Residential outdoor levels of traffic-NO_x and heating-SO_2 for the children's first year of life (1994-1997) were calculated by interpolation between 1990 and 2000, assuming a linear change in air pollution levels between these years at each geographic location. For traffic-PM_{10}, the levels from the year 2000 were used for the whole study period.

The health outcomes of interest were wheezing, allergic rhinitis, lung function and allergic sensitization. The definition of wheezing was based on questionnaire information and was subdivided in "transient," "persistent," and "late onset," according to reported episodes of wheezing during the early (3 months to 2 years) and recent (last 12 months at 4 years) age periods (See Table 1). Cumulative wheezing up to age four was defined as any of transient, persistent, and late onset wheezing.

Table 1. Definitions of Wheezing. Cells show number of reported episodes of wheezing.

Period	Transient	Persistent	Late onset
Early (3 months to 2 years)	≥3	≥1	none
Recent (last 12 months at 4 years)	none	≥1	≥1

Allergic rhinitis at 4 years was defined as reported sneezing, runny or blocked nose, or red itchy eyes after exposure to pollen or pets; a physician's diagnosis of allergic rhinitis during the last 24 months; or both.

Lung function was assessed through peak expiratory flow (PEF) measurements, performed using normal range Ferraris Peak Flow Meters (Ferraris Medical Limited, Hertford, UK). The highest of several readings was used for analysis. A PEF measurement was accepted if the operator had judged the child's performance as adequate and if the best and second-best peak expiratory flow values were within 15%. A total of 2599 out of the 2965 children who participated in the clinical testing had PEF readings that were acceptable to use in the analysis (88%).

Blood samples were drawn from 2614 children (88% of those who attended the clinical investigation). Serum IgE-antibodies were analyzed in 2543 (97%) of the blood samples. Phadiatop containing inhalant allergens of cat, dog, horse, birch, timothy, mugwort, *Dermatophagoides pteronyssinus* and *Cladosporium* and fx5 containing food allergens of cow's milk, egg white, soy bean, peanut, fish, and wheat were used. Samples that exceeded 0.35 kU$_A$/L were considered positive in Phadiatop or fx5 and were analyzed for specific IgE antibodies to the airborne and food allergens mentioned above.

The associations between air pollution exposure and the categorical endpoints (wheezing, allergic rhinitis, and sensitization) were analyzed using logistic and multinomial logistic regression, presenting the results as odds ratios (ORs) and 95% confidence intervals (CIs). The association between air pollution and peak expiratory flow was analyzed with linear regression. Air pollution exposure was entered as a continuous variable and the results are presented for an increment representing the 5th to 95th percentile difference of exposure within the cohort.

A model selection strategy with strict criteria was chosen to select the covariates to include in the respective statistical models. The set of candidate covariates to be included in the model was chosen before any results were seen and a base model was created using the air pollution variable and the design variable geographic area and sex, then the covariates were added one by one. Impact on the air pollution effect was assessed by percent change in the air pollution coefficient (in any direction), and impact on the dependent variable was assessed by the p-value for the covariate effect. If any of the criteria was met, the variable was added, if none were met, the variable was not added at this stage. However, as soon as a new variable was added to the model, all the discarded variables were again added one by one and assessed

based on the criteria, and so on until none of the variables left had any impact as either confounders or risk factors. This procedure was carried out for each outcome variable. Complete confounder sets were available for 3515 children (86% of the enrolled cohort) for the analysis of wheezing, for 2565 (63%) for the analysis of peak expiratory flow, and for 2543 (62%) for analysis of sensitization.

Results

Wheezing up to the age of 4 was reported in 22% of the children. Persistent wheezing was associated with residential exposure to traffic-PM_{10} and traffic-NO_x but not to heating-SO_2 (Table 2). Transient and late onset wheezing did not seem to be associated with exposure to any of the air pollutants and neither did all wheezing, reported doctor's diagnosis of asthma in combination with current symptoms, or allergic rhinoconjunctivitis at the age of 4 years.

Table 2. Association of residential levels of locally emitted ambient air pollution during the first year of life with various wheezing outcomes up to age 4 years (n=3515).

Wheezing outcomes	No.	Traffic-PM_{10} OR (95% CI)[a]	Traffic-NO_x OR (95% CI)[a]	Heating-SO_2 OR (95% CI)[a]
No wheezing[b]	2725	1.00	1.00	1.00
Any wheezing up to age 4	790	1.17 (0.76 to 1.80)	1.13 (0.84 to 1.53)	0.84 (0.56 to 1.27)
Transient	269	0.90 (0.45 to 1.81)	0.82 (0.48 to 1.40)	1.06 (0.57 to 1.97)
Late onset	194	0.94 (0.42 to 2.11)	0.87 (0.47 to 1.60)	0.82 (0.39 to 1.73)
Persistent	327	1.64 (0.90 to 3.00)	1.60 (1.09 to 2.36)	0.69 (0.37 to 1.29)
Girls	194	2.32 (0.93 to 5.79)	1.94 (1.07 to 3.50)	0.69 (0.27 to 1.78)
Boys	133	1.52 (0.68 to 3.40)	1.55 (0.92 to 2.63)	0.72 (0.31 to 1.66)

[a]Odds ratios are calculated for a difference in the source-specific air pollution level corresponding to the difference between the 5th and 95th percentile range in the cohort (6 µg/m^3 for PM_{10}, 44 µg/m^3 for NO_x, and 3 µg/m^3 for SO_2). Adjusted for municipality, socioeconomic status, heredity, mother's smoking during pregnancy and infancy, year that house was built, damp or mold in the home at birth, and sex of the child.
[b]Reference category.

Exposure to traffic-PM_{10} during the first year of life was associated with lower peak expiratory flow at age 4 and a similar tendency was seen for traffic-NO_x and to a lesser degree for heating-SO_2 (Table 3).

Table 3. Associations of residential levels of air pollution during the first year of life with peak expiratory flow (L/min) at age 4 years (n=2565).

	No.	Traffic-PM_{10}[a] Point Estimate (95% CI)	Traffic-NO_x[a] Point Estimate (95% CI)	Heating-SO_2[a] Point Estimate (95% CI)
All subjects[b]	2565	- 5.36 (- 10.67 to – 0.053)	- 3.08 (- 6.84 to 0.68)	- 2.07 (-7.28 to 3.14)
Girls[c]	1251	- 5.73 (- 11.73 to 0.29)	- 3.33 (- 7.80 to 1.14)	- 4.51 (-10.70 to 1.68)
Boys[c]	1314	- 5.00 (- 11.03 to 1.04)	- 2.81 (- 7.36 to 1.73)	0.47 (-5.79 to 6.74)

[a]Effects (L/min) are calculated for a difference in the source-specific air pollution level corresponding to the difference between the 5th and 95th percentile range in the cohort (6 µg/m^3 for PM_{10}, 44 µg/m^3 for NO_x, and 3 µg/m^3 for SO_2).
[b]Adjusted for sex, age, height, and municipality.
[c]Adjusted for age, height, and municipality.

Exposure to air pollution from traffic or house heating during the first year of life showed a tendency to association with sensitization overall (inhalant or food allergens). For air pollution from traffic, but not from home heating, this was driven by a strong effect for sensitization to pollen (Table 4).

Table 4. Association of residential levels of locally emitted ambient air pollution during the first year of life with allergic sensitization (IgE antibodies ≥0.35 kU$_A$/L) at age 4 years (n=2543).

Wheezing outcomes	No.	Traffic-PM_{10} OR (95% CI)[a]	Traffic-NO_x OR (95% CI)[a]	Heating-SO_2 OR (95% CI)[a]
None[b]	1929	1.00	1.00	1.00
Any food or inhalant[c]	614	1.42 (0.88 to 2.29)	1.30 (0.94 to 1.80)	1.37 (0.87 to 2.16)
Any food	406	1.06 (0.60 to 1.88)	1.16 (0.79 to 1.72)	1.41 (0.84 to 2.39)
Any inhalant	389	1.73 (0.98 to 3.04)	1.34 (0.91 to 1.98)	1.22 (0.71 to 2.11)
Pollen[d]	281	2.30 (1.23 to 4.29)	1.67 (1.10 to 2.53)	1.03 (0.55 to 1.93)
Furred pets[e]	189	0.93 (0.40 to 2.16)	0.77 (0.40 to 1.51)	0.94 (0.44 to 2.02)

[a]Odds ratios are calculated for a difference in the source-specific air pollution level corresponding to the difference between the 5th and 95th percentile range in the cohort (6 µg/m^3 for PM_{10}, 44 µg/m^3 for NO_x, and 3 µg/m^3 for SO_2). Adjusted for municipality, socioeconomic status, heredity, mother's smoking during pregnancy and infancy, year that house was built, damp or mold in the home at birth, and sex of the child.
[b]Reference category.
[c]Food allergens according to fx5, inhalant allergens according to Phadiatop.
[d]Birch, timothy, or mugworth.
[e]Cat, horse, or dog.

PAPER 2: HEAPSS

The acronym HEAPSS stands for Health Effects of Air Pollution on Susceptible Subpopulations and is a European multi-center study of first-time Myocardial Infarction survivors which had several different aims. The first aim was to study the relation between air pollution exposure and the risk of a first MI (Lanki, Pekkanen et al. 2006). A second aim was to study the risk for a readmission of MI in the patients that survived their first MI (von Klot, Peters et al. 2005). The third aim is the one presented here and concerns the association between air pollution exposure and mortality among survivors of MI and aimed to assess how air pollution levels affect daily mortality in myocardial infarction survivors.

Materials and methods

The 5 centers included in the HEAPSS study were Augsburg (Germany), Barcelona (Spain), Helsinki (Finland), Rome (Italy), and Stockholm (Sweden). For recruitment of the cohorts of first-time MI cases age 35 or older, dedicated MI registers were used in Augsburg and Barcelona, and administrative hospital discharge registers in the other cities. To reduce the possibility that the death was simply a continuation of the initial MI event, the follow-up time with respect to mortality started on the 29th day after the date of the qualifying event, thus including only 28-day survivors in the cohorts, in alignment with the definition of fatal and non-fatal MI used in the WHO MONICA Project (Tunstall-Pedoe, Kuulasmaa et al. 1994)

Subjects included in the cohorts were identified through a qualifying non-fatal MI event. In the two cities with dedicated MI registers cases were identified based on the definitions used by the registers and were confirmed to be first-time MI cases. Cases in the centers using hospital discharge registers were defined by the first occurrence of an MI diagnosis during the recruitment period as primary diagnosis for a patient admitted to an acute care hospital. First-time MI was defined as no registered MI within 3 years before the index event. Any concurrent ICD code indicating a previous MI and discharge within 3 days of the index MI diagnosis (indicating likely miscoding) led to exclusion.

Mortality data were collected from national mortality registers in Helsinki and Stockholm and from regional mortality registers in Barcelona. Municipal registers and death certificates were used to check vital status in Augsburg and Rome. The investigated outcomes were all-cause non-traumatic mortality and cardiovascular mortality. Deaths with trauma as an underlying cause were considered censored on the death date. A person was considered to be at risk from the 29th day after the index MI event until the first of the following events: death, migration out of study area, censored follow-up, or end of follow-up. If the location of non-traumatic death was known to be outside the study area, the person was censored on that day and the death not included for analysis.

In each city, air pollution data were collected from a number of fixed monitors, and a daily city average was calculated for each pollutant for each city. Data for the whole study period were available in all cities for CO, NO_2, and Ozone (O_3). PM_{10} data were unavailable for parts of the study periods in Augsburg and Barcelona and were estimated using available data for

total suspended particles and black smoke. Daily averages were calculated for CO, NO_2, sulfur dioxide (SO_2), and PM_{10} concentrations, and maximum 8-hour averages were used for ozone. As a primary hypothesis, analyses of ozone were restricted to the warm season (April–September) because levels and correlations with other pollutants depend on season, and health effects of ozone have generally been associated with summer-time levels.

Particle number concentrations were assessed retrospectively. In early 2001, condensation particle counters were set up in each location to measure the total particle number concentration of ambient particles. The purpose of these measurements was to retrospectively estimate a measure of ultrafine particles (diameter ≤100 nm). Since ultrafine particles constitute most of all particles, particle number concentration (PNC) is a good indicator of the number concentration of ultrafine particles. City-specific statistical models were developed using available data on other air pollutants and meteorological variables during a period when particle number concentration was measured in parallel with other measurements, to retrospectively estimate particle number concentration for the study follow-up periods (Paatero, Aalto et al. 2005). For Augsburg, Helsinki, Rome, and Stockholm, the follow-up period also contained some periods when particle number concentrations were measured. For these days, measured, rather than estimated, particle number concentration was used. Meteorological variables collected included temperature, dew point temperature, relative humidity, solar radiation, barometric pressure, and wind speed and direction. Many of the meteorological variables were mainly used for modeling of PNC.

A separate part of the HEAPSS project aimed to compare different analytical strategies to study short-term effects in a cohort setting like this. The comparison was carried out for the re-hospitalization part of the project and has been published separately (Peters, von Klot et al. 2006). Briefly, the main conclusion from that comparison as it applied to this part of the project was that all methods compared yielded similar point estimates for the air pollution effects and that from a practical standpoint Poisson regression analysis is the best choice for quantifying the relationship between air pollution exposure and daily mortality in the cohort of MI survivors.

A protocol for the model building strategy was finalized prior to access to the data in order to ensure that the model building were not influenced by the results. Once data was available, Poisson model specification was done separately for each city. We used generalized additive models to allow the inclusion of smooth functions for covariates and used an offset-term in all Poisson models to allow for the variable number of persons at risk. Penalized regression splines were used to model the continuous confounder variables as smooth functions. In a hierarchical approach, potential confounders were evaluated and then a core model was selected before adding air pollution as an independent variable. Of the evaluated confounders, long-term trend was forced into all models, as was at least one temperature variable (same day temperature or the difference between same day and the average of the preceding 3 days). Same-day relative humidity, same-day barometric pressure, day of week, holidays, and days of population decrease (added to the model in that order) were included only if they improved

the model. Decisions for keeping a covariate in a model were based on judgment using a number of different model parameters to guide the judgment (p-value, the generalized cross-validation score, the autocorrelation function, visual inspection of the shape of the smooth function, and partial auto correlation function). Trend was included in the model as a penalized spline with 1 to 6 knots per year (the choice based on review of the model parameters), to control for long-term trends, seasonality, and changes in the baseline risk. Once the core models were final, the air pollution variables were included as 2-day and 5-day moving averages, as pre-specified in the model building strategy.

The city-specific effect estimates were pooled using a method for meta-analysis. A test of heterogeneity was performed to decide which method to use. If heterogeneity was indicated, the combined effect estimate was calculated using a random effects model and if heterogeneity was not indicated an inverse variance-weighted average of city-specific regression coefficients was calculated.

As a sensitivity analysis we performed an extended Cox proportional hazards model that included the same covariates for each city as the Poisson regression but with quadratic terms instead of penalized splines since the Cox proportional hazards model is a fully parametric model which cannot accommodate the semi-parametric method of penalized splines. Subgroup analyses were performed to assess potential effect modification by time since enrollment and age at baseline. As a post-hoc analysis, 15-day moving average of the air pollution variables was analyzed.

Results

Table 5 shows the percent change in daily mortality per a specified unit increase in each pollutant averaged over 2, 5, and 15 days, combined for all 5 cities. For the 2-day average, the strongest effect estimates were found for the particle measurements—both around 5%. For the 5-day average, NO2 and CO were also associated with risk. The results for SO2 and O3 do not indicate an association with daily deaths except for the 15-day average of SO2; a sensitivity analysis using full-year O3 data similarly did not show any association. For all pollutants except NO2 and O3, the strongest association was observed for the 15-day average. The sensitivity analysis using the Cox model generally yielded similar results as the Poisson regression. Subgroup analyses revealed no difference between cause-specific cardiovascular mortality and all-cause mortality, no clear trend in age, and a tendency toward stronger effects for deaths occurring within 1 year after the first MI compared with deaths occurring after 1 year of follow-up.

Table 5. Percent change in daily non-trauma deaths per unit change in air pollutants. Overall pooled results of Poisson regressions over all 5 centers for the common age range 35-74 years.

Pollutant[a]	Unit Change	Mean of Lag 0 and 1 % (95% CI)	Mean of Lag 0–4 % (95% CI)	Mean of Lag 0–14 % (95% CI)
PNC	$10,000/cm^3$	5.62[b] (2.83 to 8.47)	6.01 (3.41 to 8.68)	8.68[b] (5.35 to 12.1)
PM_{10}	$10\ \mu g/m^3$	5.08[b] (1.06 to 9.27)	3.92[b] (1.19 to 6.72)	4.91[b] (1.30 to 8.65)
CO	$0.2\ mg/m^3$	2.61[b] (–0.26 to 5.56)	3.82[b] (1.00 to 6.72)	4.92[b] (2.11 to 7.81)
NO_2	$8\ \mu g/m^3$	2.31[b] (–1.26 to 6.01)	3.25[b] (0.19 to 6.39)	2.54 (–1.49 to 6.74)
SO_2[c]	$2\ \mu g/m^3$	0.09 (–2.23 to 2.46)	1.60[b] (–1.28 to 4.57)	8.06[b] (4.38 to 11.9)
O_3	$15\ \mu g/m^3$	1.04[b] (–6.32 to 8.96)	0.51[b] (–8.97 to 11.0)	–0.52[b] (–10.2 to 10.2)

[a]Daily 24-h mean, except ozone: maximum 8-h average (April-September).
[b]City-specific estimates were heterogeneous and a random effects pooling technique was used.
[c]SO_2 not available for Barcelona.

PAPER 3: ALVA

ALVA stands for Air pollution and Life threatening Ventricular Arrhythmias and is a study which was aimed at assessing how very short-term exposure to air pollution affects the risk for ventricular arrhythmia in patients with implanted cardioverter defibrillators (ICDs).

Materials and methods

All patients in Stockholm and Gothenburg who previously had received ICDs or were implanted with ICDs during the course of the study period were invited to participate in the study, provided that they were deemed mentally fit to cooperate. All included patients except one had implants due to an earlier index episode of ventricular tachyarrhythmia. Data were extracted from the records of each patient concerning pre-implantation diagnosis, type of index-arrhythmia, ejection fraction, type of ICD, medication, and co-morbidities. The patients were asked to contact the clinic within 3 days after sensing arrhythmia being treated by their ICD. Information from the ICD concerning time and date of arrhythmia, type of arrhythmia, therapy administered, and any change in programming was downloaded and documented. All intracardiac electrograms from the devices were reviewed by an electrophysiologist and only ventricular tachyarrhythmias leading to shock therapy or symptomatic ATP therapy were included. Events within 7 days after a qualifying ventricular event were excluded to try to achieve some degree of independence of events. Patients were interviewed concerning symptoms and possible triggering activities for both the 2 and 24 h periods preceding the ICD discharge as well as the location indoors or outdoors and their geographic location at the time of event in order to calculate the distance from the air pollution monitor.

Air pollution and meteorological data were obtained from one fixed centrally located roof-top monitor in each city reflecting urban background levels. In Stockholm, NO_2 values were available from two monitoring sites, and the mean was used. In the construction of moving averages, complete hourly air pollution data were required for the 2 h exposure analysis, and a maximum of 25% of hourly air pollution data was deemed an acceptable level of missing data for 24 h analysis. During the entire study period, hourly means were provided for PM_{10} and NO_2 for both Gothenburg and Stockholm and PM2.5 only for Stockholm. In addition, meteorological data including temperature, relative humidity, and barometric pressure were obtained.

The association between ventricular arrhythmias and air pollution exposure was analyzed using a case-crossover design where the end of the case period was defined as the starting time of a confirmed ventricular arrhythmia rounded back to the nearest preceding hour. Control periods were matched to the case period on time of day, day of week, calendar month, and year within each subject. Air pollution levels and meteorological data were averaged in 2 and 24 h windows preceding the time of arrhythmia for the case period and the corresponding time for the control periods. A conditional logistic regression model that included a linear term for air pollution, and penalized splines for temperature, relative humidity, and barometric pressure for the same averaging times as the air pollution

parameters, was used to calculate the effect estimates. The effects are expressed as odds ratios with 95 percent confidence intervals for an interquartile range (IQR) increase in mean concentration for each pollutant and averaging time.

In addition to the main analyses, we used the event- and subject specific data for a series of interaction analyses, exploring possible effect modification. Covariates that were assessed for effect modification were the following predictors of cardiovascular events: ischemic heart disease, ejection fraction, diabetes, use of beta-blockers, age, and body mass index. We also included the number of arrhythmias as a covariate, which has been previously reported to show associations, and factors potentially affecting exposure or misrepresentation such as indoor or outdoor location at the time of arrhythmia and distance from event to air pollution monitor. For time-variant factors such as location indoors/outdoors and distance from monitor, we had access to observations only for the case period. Interaction analyses were performed by including an interaction term between air pollution and the potential effect modifier in the conditional logistic regression model.

Results

Increased levels of PM_{10} in the 2 and 24 hours immediately preceding the events recorded by the ICD devices, were associated with an increased risk for ventricular arrhythmias. NO_2 showed a weaker association and $PM_{2.5}$ (only available in Stockholm) showed associations similar in magnitude to PM_{10} (Table 6). Interaction analyses showed stronger associations for PM_{10} in Gothenburg compared to Stockholm, in events occurring closer to the air pollution monitor compared to events occurring further away, in subjects experiencing three or more events compared to those experiencing 1 or 2, and a tendency for events occurring outdoors compared to events occurring indoors. NO_2 generally showed similar but weaker patterns.

Table 6. Odds Ratios of ventricular tachyarrhythmia for an interquartile range increase in air pollutants in patients with implantable cardioverter defibrilators.

Pollutant ($\mu g/m^3$)	Moving average (h)	IQR[a]	Subjects	Events	OR (95% CI)
PM_{10}	2	13.2	65	101	1.31 (1.00 to 1.72)
	24	10.3	67	106	1.24 (0.87 to 1.76)
$PM_{2.5}$[b]	2	7.5	33	49	1.23 (0.84 to 1.80)
	24	5.2	35	53	1.28 (0.90 to 1.84)
NO_2	2	15.3	69	109	1.09 (0.84 to 1.42)
	24	11.0	70	110	1.07 (0.81 to 1.42)

[a]Interquartile range for joint distribution of pollutants in Stockholm and Gothenburg used in case-crossover analyses, except for $PM_{2.5}$ where only data from Stockholm are presented.
[b]Stockholm only.

PAPER 4: ONSET

The Swedish ONSET study is a study which was designed to investigate triggers of first-time myocardial infarction. Published results include effects of heavy lifting (Hallqvist, Moller et al. 2000), sexual activity (Moller, Ahlbom et al. 2001), and stressful life events (Moller, Theorell et al. 2005). The analyses presented in Paper 4 aimed to assess how very short-term exposure to air pollution affects the risk of first-time myocardial infarction.

Materials and methods

The ONSET study is similar in design to the ALVA study described in Paper 3. It is a case-crossover study which stems from a large population based case-control study of causes of myocardial infarction—the Stockholm Heart Epidemiology Program (SHEEP) (Reuterwall, Hallqvist et al. 1999). First-time myocardial infarction cases were recruited between April 1993 and December 1994 from all 10 emergency hospitals in Stockholm and a total of 699 patients were interviewed for the case-crossover study. An important consideration was to determine the exact timing of onset; therefore this information was collected both from the medical records and from the interview. After exclusion of patients with unreliable information on time of onset or with a high percentage of missing or clearly inaccurate answers, 660 cases remained for analysis. The cases were not specifically chosen based on survivor status, but since inclusion was conditional on the ability to be interviewed, the likelihood of including fatal myocardial infarctions (death within 28 days) was small. As it turned out, all included cases were alive 28 days after onset of the myocardial infarction; hence we refer to them as non-fatal myocardial infarctions.

The cases were interviewed by specially trained nurses, during their hospital stay or shortly afterwards. The timing of the interview ranged from the day of the myocardial infarction to more than four months after, with a median interval of 15 days between myocardial infarction onset and the interview. At the interview, detailed information was first obtained on all episodes of pain (clock time, type, duration, *etc.*), other symptoms, and circumstances during the four days before the myocardial infarction to determine the precise time of disease onset.

Comprehensive information on common cardiovascular disease risk factors was available from the questionnaire of the main SHEEP Study that was administered after the Onset interview. The response rate of the questionnaire was 91% among cases already interviewed in the Onset study. Disease history, the presence of other risk factors for cardiovascular disease, and regular medication use were determined from the questionnaire information and from measurements made at the SHEEP health examination (generally performed three months after diagnosis when a metabolically stable status was assumed).

Air pollutants available for the study period were PM_{10}, NO_2, and CO from fixed centrally located roof-top monitors reflecting urban background levels, and O_3 from rurally located monitors representing regional background levels. Two hour and 24 hour means were calculated using the same methods as described for Paper 3.

The statistical analysis was the same as described for Paper 3.

Results

The estimated risk of onset of myocardial infarction for an interquartile increase in air pollution levels were close to unity for all pollutants and both averaging periods, see Table 7.

Table 7. Association of air pollutants with onset of myocardial infarction. Estimated odds ratios (OR) are adjusted for temperature and relative humidity.

Pollutant	Unit	Moving average (h)	IQR	Events	OR (95% CI)
PM_{10}[a]	µg/m^3	2	12.2	342	0.93 (0.81 to 1.08)
		24	9.1	341	0.99 (0.85 to 1.15)
NO_2	µg/m^3	2	18.1	657	0.97 (0.84 to 1.11)
		24	12.3	654	0.97 (0.85 to 1.11)
CO	mg/m^3	2	0.32	649	0.94 (0.82 to 1.07)
		24	0.23	651	0.97 (0.85 to 1.11)
O_3	µg/m^3	2	29.6	629	1.02 (0.85 to 1.23)
		24	24.7	631	0.92 (0.75 to 1.13)

[a]Monitoring for PM_{10} began in March 1994.

Potential effect modifiers investigated were: age; sex; smoking status; BMI; physical inactivity; a history of hypertension, diabetes, and angina; the presence or absence of premonition; and the location of the admitting hospital (Central Stockholm or elsewhere). Interaction analyses provided no clear evidence of effect modification with any of the investigated variables for any of the pollutants or averaging times.

DISCUSSION

Air pollution has been linked to a large number of health effects and has been studied in a multitude of settings with a multitude of different types of studies using a multitude of different methods. This collection of papers covers some of the different areas of research and represents different study designs and analysis strategies. The individual papers discuss the results of the respective studies in detail and put the results in the context of the current literature. Here, the studies are discussed from the viewpoint of the challenges of conducting, analyzing, interpreting, and presenting observational studies of health effects of air pollution.

MISREPRESENTATION OF EXPOSURE

In the BAMSE study reported in paper 1, the conceptual goal of the exposure assessment was to get an estimate of the exposure to the relevant air pollutants for each child in the BAMSE cohort as close as possible to the actual amount of air pollution that the child had been exposed to during the first year of life. In reality, the more important estimate was the relative amount of air pollution each child had been exposed *in comparison* to the other children, since the effect estimates are expressed in terms of a relative difference in outcome probability by a difference in exposure. The chosen approach to exposure assessment was to model the average annual levels of locally emitted pollutants at the geographic location of the child's reported residence, i.e. to estimate geographical differences in exposure. The exposure model was quite complex and was based on a large number of assumptions and estimates of traffic parameters and consequently the accuracy of the estimated exposure is dependent on the accuracy of those assumptions and estimates. Even if there are limitations to the model and the estimated pollutant level at a specific address may be inaccurate, it was shown as part of this study that the contrast between different geographic locations on average was fairly accurate. But that only goes for the estimated outdoor levels across the geographic area. Most children don't spend their first year breathing on the front stoop of their houses. For all we know they might have spent most of the time at grandma's house with completely different air quality, or they might have had a ventilation system in the house that filters the air, or a number of other factors that make the air outside their house different from the air that they actually were exposed to. The amount of time spent outside may also differ across geographic location depending on factors like residential type, socioeconomic factors, and parents' life style choices. There are studies that shed some light on the indoor vs. outdoor aspect of this, concluding that there is on average substantial infiltration of outdoor air pollutants to most indoor environments, and that studies using outdoor levels may suffer from exposure misrepresentation (Wichmann, Lind et al. 2011; Bellander, Wichmann et al. 2012). The rigorous model building strategy included covariates that could be related to differences in how well the estimated exposure represents actual exposure and if there would have been a large amount of differential misrepresentation of exposure that would have been handled by the statistical models. Also see section below about confounding control below.

In contrast to the assessment of the effect of air pollution based on variation across geographic locations in paper 1, papers 2-4 deal with short-term temporal variation in air

pollution exposure. In paper 2, the goal of the exposure assessment was to estimate ambient daily air pollution exposure levels for a large number of subjects and eventually compare the number of deaths on days with high vs. low exposure levels. In this study we knew that the subjects were somewhere in the study area when they became eligible for the study, i.e. when they had their first MI, and for those who died, we know whether they were in the study area or not at the time of death. In-between we knew next to nothing of their whereabouts. This is obviously a limitation.

In papers 3, and 4, the goal was to estimate the air pollution levels that the subjects were exposed to in the even shorter time period immediately preceding a certain event (ventricular arrhythmia in paper 3 and acute MI in paper 4), and at certain other time points when the subjects did not experience the event in question. The comparison would then be made of the exposure levels at the time points where the subjects did experience events and time points when they didn't. In both these studies, we knew little about the subjects' whereabouts at the control time points, but we did have some information about the time of the event. In the study presented in paper 3 we knew the geographic location at the time of the event, and we knew whether they were indoors or outdoors at the time of the event. This information was used in sensitivity analyses and the results of those indicated that in cases where we could suspect that the air pollution exposure estimates were likely to be poor estimates of the true exposure (when subjects were far from the air pollution monitor and when they were indoors) the effect estimates were closer to unity. In the study presented in paper 4 we knew if the subjects were admitted to a hospital in Central Stockholm (close to where the air pollution monitors are) or outside of Central Stockholm. Like most analyses in this study, the sensitivity analyses addressing this aspect were inconclusive.

For the three studies presented in papers 2-4 we chose to estimate the individual subjects' exposure through roof-top measurements in the city center. For many (if not most) subjects, this is a very poor estimate of the actual air pollution levels that they were exposed to at a particular time point. Even though we were aware that the exposure we assigned to each subject was likely to be fairly inaccurate, we considered it likely that the time points for which we estimated a high air pollution levels would *on average* have higher levels than time points for which we estimated low levels. I.e. the relative exposure time point to time point would be reflected by the roof-top measurements. However, there are local sources of air pollution whose temporal variation is not picked up by the roof-top levels and there are patterns to peoples' behavior that may be linked to the weather (going to the beach when the weather in nice etc.), and we know that the air pollution levels are also affected by the weather. Variation in levels measured at the roof-top monitors is partly due to long-range transport of the air pollution, meaning that "air packages" containing air pollution generated elsewhere are carried to the city by the wind. This part of the air pollution mixture will be similar over a large geographic area and the roof top measurements may therefore more accurately reflect temporal variation in individual exposure to long-range transport air pollution (Bellander, Wichmann et al. 2012)

Another challenge related to exposure assessment, which is specific to studies carried out in relatively clean places like Sweden, is the question of whether or not the air pollution levels are high enough to cause any harmful effects at all. Most studies in the literature that have shown effects of air pollution were carried out in places where the air pollution levels were much higher than in Sweden. Even though the statistical models generally estimate linear effects there could be thresholds under which no effects exists. Few studies are large enough and contain a wide enough range of air pollution levels to be able to shed light on this issue and the literature on the topic of threshold effects is limited. The joint European and North American collaboration APHENA attempted to evaluate possible threshold effects and even though they had access to a huge database they concluded that if a threshold exists it would be difficult to detect. The analysis they did perform showed little evidence of any threshold (Katsouyanni, Samet et al. 2009).

MISREPRESENTATION OF OUTCOME

In paper 1, our measures of wheezing were reported in a questionnaire by the parents who were asked to remember the number of episodes of wheezing during the past year. A questionnaire can be an inexact and blunt tool of measurement. Some parents may be more attentive to their children's health than others. Some parents may tend to over-report the number of episodes and others may tend to under-report. The term "wheezing" ("pipande/väsande andning" in Swedish) may also mean different things to different parents. The best we can hope for is that *on average* the children that the questionnaire classified as wheezers, wheezed more than those classified as non-wheezers. And, perhaps most importantly, we hope that the factors that may make parents report differently are *not* directly or indirectly correlated with the air pollution levels their children are exposed to. This may be too much to hope for since it is likely that at least some parents may choose their place of residence based on air quality of the location, and this choice may be correlated to some unmeasurable trait that also affects how they answer the questionnaires. At the time of inception of the BAMSE cohort (1994-96), the public awareness of health effects from air pollution was however much lower than today.

The HEAPSS study reported in paper 2 was a 5-center European study and the MI cases that form the study population were identified in different ways in the different countries. In some centers cases came from dedicated MI registers that employed standardized criteria for identification and diagnosis. These registers were specifically aimed at identifying MI cases and used great detail about the events in order to achieve the correct diagnosis. Other centers relied on administrative registers that have a much broader scope and are a cruder tool, where identification is dependent on the treating physicians' diagnosis. The treating physician would presumably use standardized criteria as well but there is no control over this from the registers' point of view. This difference in the degree of sophistication when identifying MI cases may lead to slightly different composition of the cohorts in the different countries and thus potentially different susceptibility. The main outcome in this study was the seemingly uncomplicated endpoint of death, but even here there could be uncertainty in the accuracy of

the exact time point of death as well as potential differences across centers since some countries had very reliable mortality registers while others didn't and vital status instead had to be collected from municipal registers and causes of death were collected by manually inspecting death certificates.

The ALVA study reported in paper 3 consisted of subjects who had an implanted ICD device. The device records its activity with great precision and much detail making the determination of ventricular arrhythmias a fairly reliable outcome measure, both as to the exact time of the event and the certainty of the type of event since it provide the researchers with intracardiac electrogram print-outs in addition to its own classification of the event. The Swedish ONSET study reported in paper 4 also has very reliable assessments of the outcome myocardial infraction. That is because this study was specifically designed to gain the most precision possible on the time of onset and other aspects of the MI through in-depth interviews with the subjects during the period of hospitalization or soon after discharge. In fact, it is the reliability of the timing of the events in both these studies that made the studies possible in the first place.

CONFOUNDING CONTROL

BAMSE, presented in paper 1, has a classic cohort study design where the children are recruited, their exposure is measured, and they are followed over time with respect to health outcomes. The prospective nature of the study type is often held forth as a good property since it reduces the risk for exposure estimates being influenced by the outcome, but the study design does nothing to control for confounding by design. The dependent variables were both categorical (wheezing and allergic sensitization) and continuous (PEF) so the statistical models were fairly straightforward regression models, logistic or multinomial logistic for the categorical and linear for the continuous. Regression models lend themselves well to adjustment for potential confounding since the models can handle multiple independent variables as covariates, especially when the sample size is large as it was in this study. Even so, there is a desire to keep the model as parsimonious as possible while including the important covariates, and a rigorous model building strategy was implemented. To minimize the risk of the model selection being influenced by user input, a manual method with strict criteria was chosen as described in the methods section. This procedure, which included a manual iterative process of evaluating many combinations of the potential confounders, was rather cumbersome but well worth it, since we could feel confident that we had come up with appropriate models without introducing too much bias.

In the HEAPSS study presented in paper 2, confounding from any subject-related individual factors is effectively controlled for by the study design as a time series analysis. The unit of observation in this study is *day* rather than the more conventional *subject*, and the dependent variable is the number of deaths on each day. The fact that we don't have to worry about confounding from individual factors provides little comfort, since there is plenty of potential confounding left to control for. It is probably safe to assume that many deaths that occur have little to do with air pollution exposure, even in a potentially susceptible subpopulation like MI

survivors. So to try to tease out what influence air pollution exposure has on the number of deaths per day seems like a rather difficult task and we need a very large sample and very good statistical models to be able to tease out this effect among all the other reasons that the number of deaths per day fluctuate - if the effect indeed is there. Daily event rates fluctuate seasonally but there are also year-to-year differences, long-term trends, and day-of-week effects, all of which also are associated with air pollution levels. Air pollution is also strongly interconnected with weather and factors like temperature, humidity, precipitation and barometric pressure can also be linked to some health events making them candidates for confounding. The statistical methodology for analyzing time series studies of air pollution and acute events has been the topic of a lot of discussion in the research community for the past few decades and the different methods have been scrutinized and refined to the point where most researchers now agree on the appropriate methods. (Dominici, McDermott et al. 2002; Peng, Dominici et al. 2006) The current paradigm is the use of Poisson regression models with smooth functions for trend and other continuous variables like temperature and humidity. The choice of the type of smooth function has been under much debate but in recent years the penalized spline function used in this study has been widely used. This study has the added complexity that the time series analysis is performed on a well-defined dynamic cohort rather than in the general population. In studies of the general population it is not too controversial to assume that the underlying population at risk is constant over time and that the event rates are fairly constant over time, both of which are assumptions of the Poisson regression model. In the case of the time series within a cohort of MI survivors, these assumptions no longer hold. The population at risk is clearly not constant since subjects are recruited into the cohort by their qualifying MI events and are followed for a specific amount of time, meaning that the population at risk increases dramatically in the beginning as the cohort is recruiting and decreases towards the end as subjects leave the cohort. This problem was solved by adding an offset term to the model which essentially functions as a denominator for the Poisson process. The constant hazard assumption is also clearly violated since all subjects suffered an MI within the past month as they entered the study and are clearly at high risk, which means that in the beginning of the study all subjects are at high risk while towards the end everyone still in the study has survived the critical time after their MI and their risk is lower. This problem is not solved by the addition of the offset term, but by modeling trend properly the problem should be reduced. To our knowledge no-one had previously tackled the problem of performing a time series study of air pollution and health events within a cohort, and we spent some time comparing and contrasting different models and even published a separate paper on it (Peters 2006). Since this was relatively uncharted territory we felt it was prudent to include the results of a sensitivity analysis where we analyzed the same data using an extended Cox model that avoided these particular problems. The sensitivity analysis showed similar results, which made us inclined to believe that the violated constant hazard assumption did not dramatically influence the results. In this study we also employed a pre-defined model building strategy to minimize bias from user input.

The studies presented in paper 3 and 4 both utilize the elegant design of the case-crossover study (Maclure 1991). By assigning a time period immediately prior to the event as the case period and other time periods without events as control-periods and comparing the air pollution levels between case and control periods within each subject, any factor that is constant within subject is controlled for by design. By choosing the controls as the same time points, the same day-of-the-week, within the same calendar month, day-of-week effects, seasonal effects and to a certain extent long-term time trends are also controlled for by design. That leaves factors that vary over time, in this case meteorology, as the most obvious potential confounders to control for in the statistical model. Since there weren't that many candidates for confounding control, a model building strategy was not necessary and we added the available weather variables as smooth terms in the conditional logistic regression model. Another elegant feature of this design is that the model lends itself well to appropriate subgroup analyses by simply adding interaction terms with the air pollution variable to the model. This we took full advantage of, and we presented interaction analyses with no less than 12 different factors in paper 3 and 10 in paper 4, for each pollutants and each of the averaging times, yielding a total of 48 interaction analyses in paper 3 and 80 in paper 4. A drawback with the case-crossover design is that it is not very powerful. Compared to for instance the time series analysis approach where we contrast the air pollution exposure of the day of the event with the exposure every other day on the study, which often goes on for years, each case period in the in the case-crossover design is compared with only 3 or 4 control periods. This yields large uncertainties and requires large studies to be able to draw firm conclusions. Unfortunately, neither of studies 3 and 4 were very large in the end and the point estimates are accompanied by a large amount of uncertainty.

INTERPRETATION OF THE RESULTS

Paper 1: BAMSE

Considering the different potential health-related effects, respiratory effects of air pollution are perhaps the most intuitive since the respiratory system is one of the first body systems to come in contact with the air, and its pollution, as we breathe, and many of us have experienced discomfort when we find ourselves breathing heavily polluted air. It is also intuitive that children with their newly developed respiratory systems could be a particularly sensitive subgroup. In paper 1, the BAMSE birth cohort was studied and we found an association between estimated traffic generated outdoor air pollution levels at the children's address during their first year of life and persistent wheezing up to age four as reported by their parents. We also observed an association to lung function as measured by PEF, with lower lung function at four years of age observed in children exposed to higher levels of air pollution. A third finding was that air pollution was associated with allergic sensitization to pollen at age four.

The design and execution of the study along with our statistical analysis strategy (especially the rigorous model building strategy) gave us some confidence in the observed results, but the results were far from clear cut and the effect estimates had wide confidence intervals and

were not seen consistently for both measures of traffic generated air pollution (PM_{10} and NO_x). Associations were not observed for some other investigated measures of respiratory illness, like rhinoconjunctivitis, doctors' diagnosis of asthma, or the other measures of wheezing (transient, early onset, and cumulative). Given the numerous challenges in exposure assessment, outcome assessment, and confounding control discussed earlier we would always expect the results to be rather shaky and even though BAMSE in some sense is a large study, it is on the small side to study air pollution effects where we expect large variability. So the fact that we observed at least some consistency in the effects, despite all the uncertainties, did lead us to conclude that the risk for airway disease before school age may indeed be increased by early exposure to air pollution.

Paper 2: HEAPSS

To our knowledge the HEAPSS study was unique in that the short-term association between air pollution and health events was analyzed in a subgroup of potentially susceptible individuals – MI survivors. In the part of the study that is included in Paper 2, an association was seen between increases in air pollution, especially particles, and mortality in the cohort of MI survivors. Compared to effect estimates from studies conducted in the general population, the effect estimates for the MI survivors were substantially larger. For the gaseous pollutants the relationship was less clear. Longer averaging times seemed to yield somewhat stronger effects, and the longest averaging time (which showed the strongest effects for most pollutants) was the result of a post hoc analysis likely to be affected by some upward bias since the hypothesis was in part generated based on the data. There was a large heterogeneity in effect across centers and subgroup analyses failed to provide any additional insight.

In addition to the challenges that are common to all time series studies of air pollution and health events, the fact that the time series approach was applied to a dynamic cohort presented a number of unique statistical challenges that had to be tackled. The study was also comparatively small which is always problematic in terms of interpretation of the results. The heterogeneity across centers is probably the most troublesome aspect of the study. Despite all the efforts made to conduct the studies in the same way at each center, there were enough differences between the centers that the substantial heterogeneity came as no surprise. Most differences were unavoidable, like the composition of the study populations, the registers used to collect the data, the national health care systems, medical practices, sources of air pollution, measurement of air pollution, climate, meteorology, etc. The small size of some centers, in terms of the cohort of MI survivors followed and the rarity of events also contributes to the heterogeneity just by the large variability that can be expected from the smaller centers. The heterogeneity was taken into consideration when choosing the method for the meta-analysis but that mostly just attaches the proper uncertainty to the effect estimates. It does nothing to reassure the reader that the combined estimate is a good estimate of the true effect. In any meta-analysis, the meta-analyst must consider whether or not it is appropriate or valuable to summarize all the different effect estimates into just the one overall effect, which may not be representative of any of the contributing subpopulations. Are the

contributing centers similar enough that the individual estimates are estimates of the same true effect, or are the contributing subpopulations different enough that the individual estimates are estimates of different true effects? In the latter case the pooled estimate is an estimate of average effect and only interpretable as such. An example of the former could be when a drug company perform multiple clinical trials of the same treatment effect in the same population and then perform a meta-analysis to come up with an overall estimate of the true effect. An example of the latter could be when a large well conducted observational study yields a relative risk of 2 in women and 0.5 in men; in this case the combined relative risk estimate of 1 is not meaningful from the perspective of estimating the effect the exposure has on the outcome since it is representative of neither sex but has some interpretability as an average effect across sexes. In our case we are likely somewhere in-between. The centers are fairly different, but are they so different that we would expect different true effects on mortality from air pollution in MI survivors? We also observed a pattern where the smaller centers, which also were based on a different type of register, have the stronger effects. Just based on the size we would expect larger variability in the effect estimates, but not the fairly consistent strong positive effects. This led us to speculate that the register type could be a reason for this heterogeneity but we did not have any solid support for this. Combining effect estimates from different cities through meta-analysis is a common practice in this area of research and is in fact necessary to obtain the kind of sample size and statistical power needed to arrive at effect estimates with reasonable precision.

Any study which is the first to investigate a phenomenon should be considered exploratory and the results should be considered hypothesis generating. The results should be presented along with supporting analyses and transparency is crucial since the methodology is not well established. We included result of a sensitivity analyses as an online appendix that used an alternative statistical approach with different assumptions and we included a graphical presentation of the city specific results, which added the appropriate transparency to illuminate the substantial geographical heterogeneity. We were also transparent about the 15 day averaging time being a post hoc analysis. Given the substantial challenges in conduct, analysis, and interpretation of the results, our conclusion was the appropriately modest suggestion of an association.

Paper 3: ALVA

Although this was the first such study performed in Europe, the ALVA study was the eighth published study of air pollution as a trigger for ventricular arrhythmias in ICD patients, all other studies being from the U.S. and Canada (Ljungman, Berglind et al. 2008). What set ALVA apart from the previous studies was that we were able to use as short averaging time as 2 hours prior to the event and that we knew more about where the events took place and were able to consider that in an analysis. Unfortunately our study was one of the smaller ones in terms of the number of events which limits the ability to draw conclusions. We observed effects of PM10 that were in line with the strongest effects reported in the literature but the estimates had a large measure of uncertainty. One important finding was that a subgroup

analysis showed stronger effects when the event took place closer to the air pollution monitor making intuitive sense since the exposure assessment likely is better for events that occur closer to the monitors. Our study was conducted in two Swedish cities and the effects were only observed in one.

As previously discussed, this study has a number of features that removes a lot of potential problems by design. The ability to accurately determine the exact time point of the events from the implanted devices reduces the risk for temporal misrepresentation, which is important since the comparison is of the temporal contrast in air pollution levels. The case-crossover design also limits the potential for confounding by time varying factors. This makes the main remaining challenge with this study its small size. It is reasonable to assume that not all arrhythmias are triggered by air pollution exposure, which means that in order to detect a signal, if it is there, a large sample is needed. The results of this study are in some ways rather exciting, with strong effect estimates and subgroup analyses that show factors that may reduce misrepresentation of exposure indeed giving stronger estimates as could be expected (being closer to the monitor and being outdoors). Unfortunately, the wide confidence intervals make the results less impressive and the heterogeneity across the two cities is also of great concern. When starting out with a small sample size, the ability to perform supporting sensitivity analyses by restriction and subgrouping is very limited and in the end, we were unable to give any explanation to the heterogeneity apart from speculation.

We compared our study with the literature and even included an online appendix which contained a comparison table of the 8 studies contrasting features that could be important to explain the differences in results of the studies. Our conclusion from that comparison was that the studies that had characterized temporal exposure better, i.e. made sure that the exposure preceded the event, showed stronger effects. Our overall conclusion was that there appeared to be an association with a caveat about the small sample size.

Paper 4: ONSET

The Swedish ONSET study was not initially designed to investigate air pollution as a trigger for MI, but a multitude of other triggers. The use of this study to investigate air pollution exposure came from opportunity; the material existed so why not use it? As previously mentioned, the quality of the outcome data was very high and the design of the study is very clever, but despite these features the results showed no effect of any of the investigated pollutants for any of the averaging times. A large number of interaction analyses were performed and not only did none of the analyses indicate an interaction but also in none of the subgroup estimates could unity be ruled out.

Even though there are arguably no good or bad results, only good or bad studies, it is difficult to work up much excitement for these results. The study was certainly a good study, albeit a little small to study the anticipated weak associations of air pollution exposure, but the lack of excitement stems from the fact that it is impossible to prove a negative, or put in another way: absence of evidence is not evidence of absence. Our conclusion was appropriately that it

provides no support for an elevated risk, but that doesn't say much. When a study fails to show an association it is extra problematic if it was underpowered to begin with. It may very well be that there is no association to be found. In fact we spent most of the discussion in paper 4, including 2 tables, arguing that our null result was in line with most previous studies, but in the end we didn't really prove anything.

CONCLUSIONS

By use of four observational studies, a number of challenges in assessing health effects of air pollution were identified and discussed. The choices of study designs and the use of methods to address potential biases were assessed for each of the four studies. The conclusion was that design, analysis, presentation, and interpretation were systematic, cautious, and adequate, allowing each of the studies to effectively contribute to the collective scientific knowledge in the field of health effects of air pollution.

The specific objectives of the work performed were successfully addressed, allowing the following conclusions:

1. Exposure to moderate levels of air pollution from traffic during the first year of life may increase the risk of airway disease in preschool children.

2. Exposure to traffic related air pollution was suggested to be associated with all-cause mortality in myocardial infarction survivors, with effect estimates generally substantially higher than those for the general population.

3. Increased levels of PM10 appeared to be associated with an increased risk of ventricular arrhythmias already within 2 hours of exposure.

4. There was no support that 2 hour or 24 hour mean levels of air pollution in Stockholm contributed to triggering of a first-time non-fatal myocardial infarction

ACKNOWLEDGEMENTS

As I touched upon in the preface, my first job out of college was as "Assistant to the Epidemiologist" and over the years I have assisted a number of epidemiologists and epidemiologists in training and I would like to acknowledge some of them as well as some statisticians and others who have been influential to me academically, professionally, and otherwise.

I will start with my advisors Tom Bellander and Fredrik Nyberg. Tom and I started at the department of Environmental Medicine within weeks back in the spring of 1997 and we have worked together in one form or another more or less ever since. Tom is that rare bird of a non-physician who has made a decent career within epidemiology (now Professor!) which is very inspirational. Tom has been a great mentor and friend over the years and has always taken the time to engage in the most esoteric, philosophic, and personal discussions.

Fredrik was a PhD student when I first joined the department and we have also worked together in some respect ever since. Fredrik has a very thorough knowledge of both the theoretical and practical aspect of epidemiology and we have had many stimulating discussions over the years. He was also a role-model of sorts, in that he was the first of my close colleagues to take the leap over from academia to the pharmaceutical industry, and in a round-about fashion I have followed in his footsteps to where we now again work together at AstraZeneca.

Other senior epidemiologists I have worked with and learned a lot from are Lars Järup, Magnus Wickman, Magnus Svartengren, Johan Hallqvist and Gösta Bluhm and of course the Big Boss himself Professor Göran Pershagen.

Among the junior epidemiologists I have worked with I would like to begin by acknowledging Mats Rosenlund, with whom I shared my first office at the department back in 97 and later shared many adventures and hotel stays with, as well as a few drinks. We should keep in touch! Gunnel Emenius, Lotta Egmar, and Inger Kull showed me the ropes when I first started and taught me a lot about hands-on environmental health research through the BAMSE project. Petter Ljungman and Erik Melén helped me overcome my early misguided prejudice about physicians and showed me that MDs can not only be excellent researchers but great friends!

I would also like to mention a few other co-workers at the Department of Environmental Medicine over the years: André Lauber, Emma Nordling, Eva Hallner, Stina Gustavsson, and Marie Lewné.

Through HEAPSS I got the opportunity to work with colleagues in other parts of Europe and it was a great experience to work with some of the leading researchers in the field, Francesco Forastiere, Annette Peters, Jouha Pekkanen, and Jordi Sunyer, and I had a lot of fun with my more junior counterparts Stephanie von Klot, Sally Picciotti, and Timo Lanki.

REFERENCES

Almqvist, C., A. C. Egmar, et al. (2003). "Direct and indirect exposure to pets - risk of sensitization and asthma at 4 years in a birth cohort." Clin Exp Allergy **33**(9): 1190-1197.

Analitis, A., K. Katsouyanni, et al. (2006). "Short-term effects of ambient particles on cardiovascular and respiratory mortality." Epidemiology **17**(2): 230-233.

Anderson, J. O., J. G. Thundiyil, et al. (2011). "Clearing the Air: A Review of the Effects of Particulate Matter Air Pollution on Human Health." J Med Toxicol.

Asarnoj, A., E. Ostblom, et al. (2008). "Sensitization to inhalant allergens between 4 and 8 years of age is a dynamic process: results from the BAMSE birth cohort." Clin Exp Allergy **38**(9): 1507-1513.

Ballardini, N., I. Kull, et al. (2012). "Development and comorbidity of eczema, asthma and rhinitis to age 12: data from the BAMSE birth cohort." Allergy **67**(4): 537-544.

Bell, M. L. and D. L. Davis (2001). "Reassessment of the lethal London fog of 1952: novel indicators of acute and chronic consequences of acute exposure to air pollution." Environ Health Perspect **109 Suppl 3**: 389-394.

Bellander, T., N. Berglind, et al. (2001). "Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm." Environ Health Perspect **109**(6): 633-639.

Bellander, T., J. Wichmann, et al. (2012). "Individual exposure to NO2 in relation to spatial and temporal exposure indices in Stockholm, Sweden: the INDEX study." PLoS One **7**(6): e39536.

DeSanctis, G. (1984). "Computer Graphics as Decision Aids: Directions for Research." Decision Sciences **15**(4): 463-487.

Dockery, D. W., C. A. Pope, 3rd, et al. (1993). "An association between air pollution and mortality in six U.S. cities." N Engl J Med **329**(24): 1753-1759.

Dominici, F., A. McDermott, et al. (2005). "Revised analyses of the National Morbidity, Mortality, and Air Pollution Study: mortality among residents of 90 cities." J Toxicol Environ Health A **68**(13-14): 1071-1092.

Dominici, F., A. McDermott, et al. (2002). "On the use of generalized additive models in time-series studies of air pollution and health." Am J Epidemiol **156**(3): 193-203.

Dosemeci, M., S. Wacholder, et al. (1990). "Does nondifferential misclassification of exposure always bias a true effect toward the null value?" Am J Epidemiol **132**(4): 746-748.

Dunbar, K. (2012). Scientific Thinking and Reasoning. The Oxford Handbook of Thinking and Reasoning. R. G. M. Keith J. Holyoak. Oxford, Oxford University Press,**:** 701-718.

Emenius, G., M. Svartengren, et al. (2004). "Building characteristics, indoor air quality and recurrent wheezing in very young children (BAMSE)." Indoor Air **14**(1): 34-42.

Emenius, G., M. Svartengren, et al. (2004). "Indoor exposures and recurrent wheezing in infants: a study in the BAMSE cohort." Acta Paediatr **93**(7): 899-905.

Forster, M. and E. Sober (1994). "How to tell when simpler, more unified, or less ad-hoc theories will provide more accurate predictions." British Journal for the Philosophy of Science **45**(1): 1-35.

Francis, J. A., A. K. Shea, et al. (2006). "Challenging the epidemiologic evidence on passive smoking: tactics of tobacco industry expert witnesses." Tob Control **15 Suppl 4**: iv68-76.

Gruzieva, O., T. Bellander, et al. (2012). "Traffic-related air pollution and development of allergic sensitization in children during the first 8 years of life." J Allergy Clin Immunol **129**(1): 240-246.

Gruzieva, O., A. Bergstrom, et al. (2013). "Exposure to air pollution from traffic and childhood asthma until 12 years of age." Epidemiology **24**(1): 54-61.

Hallqvist, J., J. Moller, et al. (2000). "Does heavy physical exertion trigger myocardial infarction? A case-crossover analysis nested in a population-based case-referent study." Am J Epidemiol **151**(5): 459-467.

Hosmer, D. W. and S. Lemeshow (2000). Applied logistic regression. New York, Wiley.

Katsouyanni, K., J. M. Samet, et al. (2009). "Air pollution and health: a European and North American approach (APHENA)." Res Rep Health Eff Inst(142): 5-90.

Katsouyanni, K., G. Touloumi, et al. (2001). "Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project." Epidemiology **12**(5): 521-531.

Katsouyanni, K., D. Zmirou, et al. (1997). "Short-term effects of air pollution on health: a European approach using epidemiologic time series data. The APHEA Project. Air Pollution Health Effects--A European Approach." Public Health Rev **25**(1): 7-18; discussion 19-28.

Lanki, T., J. Pekkanen, et al. (2006). "Associations of traffic related air pollutants with hospitalisation for first acute myocardial infarction: the HEAPSS study." Occup Environ Med **63**(12): 844-851.

Lannero, E., M. Wickman, et al. (2006). "Maternal smoking during pregnancy increases the risk of recurrent wheezing during the first years of life (BAMSE)." Respir Res **7**: 3.

Lannero, E., M. Wickman, et al. (2008). "Exposure to environmental tobacco smoke and sensitisation in children." Thorax **63**(2): 172-176.

Ljungman, P. L., N. Berglind, et al. (2008). "Rapid effects of air pollution on ventricular arrhythmias." Eur Heart J **29**(23): 2894-2901.

Maclure, M. (1991). "The Case-Crossover Design: A Method for Studying Transient Effects on the Risk of Acute Events." American Journal of Epidemiology **133**(2): 144-153.

Melen, E., S. Bruce, et al. (2005). "Haplotypes of G protein-coupled receptor 154 are associated with childhood allergy and asthma." Am J Respir Crit Care Med **171**(10): 1089-1095.

Melen, E., F. Nyberg, et al. (2008). "Interactions between glutathione S-transferase P1, tumor necrosis factor, and traffic-related air pollution for development of childhood allergic disease." Environ Health Perspect **116**(8): 1077-1084.

Melen, E., S. Umerkajeff, et al. (2006). "Interaction between variants in the interleukin-4 receptor alpha and interleukin-9 receptor genes in childhood wheezing: evidence from a birth cohort study." Clin Exp Allergy **36**(11): 1391-1398.

Moller, J., A. Ahlbom, et al. (2001). "Sexual activity as a trigger of myocardial infarction. A case-crossover analysis in the Stockholm Heart Epidemiology Programme (SHEEP)." Heart **86**(4): 387-390.

Moller, J., T. Theorell, et al. (2005). "Work related stressful life events and the risk of myocardial infarction. Case-control and case-crossover analyses within the Stockholm heart epidemiology programme (SHEEP)." J Epidemiol Community Health **59**(1): 23-30.

Nordling, E., N. Berglind, et al. (2008). "Traffic-related air pollution and childhood respiratory symptoms, function and allergies." Epidemiology **19**(3): 401-408.

Nyberg, F. and G. Pershagen (2000). "Epidemiologic studies on the health effects of ambient particulate air pollution." Scand J Work Environ Health **26 Suppl 1**: 49-89.

Paatero, P., P. Aalto, et al. (2005). "Estimating time series of aerosol particle number concentrations in the five HEAPSS cities on the basis of measured air pollution and meteorological variables." Atmospheric Environment **39**(12): 2261-2273.

Peng, R. D., F. Dominici, et al. (2006). "Model choice in time series studies of air pollution and mortality." Journal of the Royal Statistical Society: Series A (Statistics in Society) **169**(2): 179-203.

Peters, A., S. von Klot, et al. (2006). "Comparison of different methods in analyzing short-term air pollution effects in a cohort study of susceptible individuals." Epidemiol Perspect Innov **3**: 10.

Pope, C. A., 3rd and D. W. Dockery (2006). "Health effects of fine particulate air pollution: lines that connect." J Air Waste Manag Assoc **56**(6): 709-742.

Pope, C. A., 3rd, M. J. Thun, et al. (1995). "Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults." Am J Respir Crit Care Med **151**(3 Pt 1): 669-674.

Reuterwall, C., J. Hallqvist, et al. (1999). "Higher relative, but lower absolute risks of myocardial infarction in women than in men: analysis of some major risk factors in the SHEEP study. The SHEEP Study Group." J Intern Med **246**(2): 161-174.

Rothman, K. J. and S. Greenland (1998). Modern epidemiology. Philadelphia, PA, Lippincott-Raven.

Ruckerl, R., A. Schneider, et al. (2011). "Health effects of particulate air pollution: A review of epidemiological evidence." Inhal Toxicol **23**(10): 555-592.

Saarne, T., H. Gronlund, et al. (2010). "Cat sensitization identified by recombinant Fel d 1 several years before symptoms--results from the BAMSE cohort." Pediatr Allergy Immunol **21**(2 Pt 1): 277-283.

Samet, J. M., S. L. Zeger, et al. (2000). "The National Morbidity, Mortality, and Air Pollution Study. Part II: Morbidity and mortality from air pollution in the United States." Res Rep Health Eff Inst **94**(Pt 2): 5-70; discussion 71-79.

Schultz, E. S., O. Gruzieva, et al. (2012). "Traffic-related air pollution and lung function in children at 8 years of age: a birth cohort study." Am J Respir Crit Care Med **186**(12): 1286-1291.

Sturmer, T., M. Joshi, et al. (2006). "A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods." J Clin Epidemiol **59**(5): 437-447.

Tunstall-Pedoe, H., K. Kuulasmaa, et al. (1994). "Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents." Circulation **90**(1): 583-612.

Wichmann, J., T. Lind, et al. (2011). "PM2.5, soot and NO2 indoor-outdoor relationships at homes, pre-schools and schools in Stockholm, Sweden." Atmospheric Environment **44(35)**: 4536-4545.

Wickman, M. (2004). "Experience with quantitative IgE antibody analysis in relation to allergic disease within the BAMSE birth cohort--towards an improved diagnostic process." Allergy **59 Suppl 78**: 30-31.

Wickman, M., S. Ahlstedt, et al. (2003). "Quantification of IgE antibodies simplifies the classification of allergic diseases in 4-year-old children. A report from the prospective birth cohort study--BAMSE." Pediatr Allergy Immunol **14**(6): 441-447.

Wickman, M., E. Melen, et al. (2003). "Strategies for preventing wheezing and asthma in small children." Allergy **58**(8): 742-747.

von Klot, S., A. Peters, et al. (2005). "Ambient air pollution is associated with increased risk of hospital cardiac readmissions of myocardial infarction survivors in five European cities." Circulation **112**(20): 3073-3079.

World Health Organization. Regional Office for Europe. (1999). Overview of the environment and health in Europe in the 1990s : third Ministerial Conference on Environment and Health, London, 16-18 June 1999. Copenhagen, WHO Regional Office for Europe.

World Health Organization. Regional Office for Europe. (2000). Methods of assessing risk to health from exposure to hazards released from waste landfills. Copenhagen, WHO Regional Office for Europe.

Zanobetti, A., A. Baccarelli, et al. (2011). "Gene-air pollution interaction and cardiovascular disease: a review." Prog Cardiovasc Dis **53**(5): 344-352.

Zou, B., J. G. Wilson, et al. (2009). "Air pollution exposure assessment methods utilized in epidemiological studies." J Environ Monit **11**(3): 475-490.