Otto Hänninen

Probabilistic Modelling of $PM_{2.5}$ Exposures in the Working Age Population of Helsinki Metropolitan Area

Publications of the National Public Health Institute A 10/2005

Department of Environmental Health
National Public Health Institute
Kuopio, Finland

PROBABILISTIC MODELLING OF $PM_{2.5}$ EXPOSURES IN THE WORKING AGE POPULATION OF HELSINKI METROPOLITAN AREA

Otto Hänninen

ACADEMIC DISSERTATION

Presented with the permission of the Faculty of Natural and Environmental Sciences, University of Kuopio, for public examination in auditorium ML2, Medistudia,Savilahdentie 6, on June 17th, 2005, at 12:00.

National Public Health Institute (KTL)
Department of Environmental Health, Kuopio, Finland

Kuopio 2005

Publications of the National Public Health Institute
KTL A10/2005

Julkaisija-Utgivare-Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

ISBN 951-740-521-0
ISSN 0359-3584
ISBN 951-740-522-7 (pdf)
ISSN 1458-6290 (pdf)
http://www.ktl.fi/attachments/suomi/julkaisut/julkaisusarja_a/2005/2005a10.pdf

Kansikuva: © Helena Hänninen. Koululaisia Hakaniemessä.

Edita Prima Oy
Helsinki 2005

Supervised by

Professor Matti J. Jantunen, Ph.D.
National Public Health Institute (KTL)
Department of Environmental Health
Air Research Laboratory
Kuopio, Finland

Professor Juhani Ruuskanen, Ph.D.
University of Kuopio
Department of Environmental Sciences
Kuopio, Finland

Dr. Erik Lebret, Ph.D.
National Institute for Public Health and the Environment
Centre for Environmental Health Research
Bilthoven, The Netherlands

Reviewed by

Professor Jaakko Kukkonen, Ph.D.
Finnish Meteorological Institute
Helsinki, Finland

Dr. Nicole Janssen, Ph.D.
National Institute for Public Health and the Environment
Centre for Environmental Health Research
Bilthoven, The Netherlands

Opponent

Professor Kaarle Hämeri, Ph.D.
University of Helsinki
Department of Physical Sciences
Helsinki, Finland

Hänninen, Otto: **Pääkaupunkiseudun työikäisen väestön pienhiukkasaltistuksen mallittaminen**.
Publications of the National Public Health Insitute, A10/2005. ISBN 951-740-521-0 (pfd 951-740-522-7),
ISSN 0359-3584 (pdf 1458-6290). http://www.ktl.fi/attachments/suomi/julkaisut/julkaisusarja_a/2005/2005a10.pdf

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Pienhiukkaset ovat vuosittain osasyynä satoihin tuhansiin kuolemantapauksiin Euroopassa. Pyrittäessä vähentämään ilmansaasteiden haittoja ensisijaisena keinona on yleinen ilmanlaadun parantaminen ja päästöjen vähentäminen, mutta vähentämistoimet voidaan kohdentaa monin eri tavoin. On selvää, että terveyden kannalta parhaaseen tulokseen päästään vähentämällä nimen omaan väestön altistusta tehokkaasti.

Ilmanlaadun ajallisen ja paikallisen vaihtelun lisäksi altistukseen vaikuttavat väestön ajankäyttö, erityisesti liikenteessä ja toisaalta sisätiloissa vietetty aika. Liikenteessä päästölähteiden läheisyys nostaa päästöjen vaikutusta altistukseen, sisällä oleskeltaessa puolestaan rakennukset suodattavat melko suuren osan ulkoilman pitoisuuksista. Toisaalta oma merkityksensä sisällä tapahtuvaan altistukseen on sisälähteillä, jotka joissain tapauksissa voivat kohottaa sisäilman pitoisuudet kertaluokkia korkeammaksi kuin pitoisuudet ulkona.

Tässä työssä kehitettiin väestön altistusten arviointiin soveltuva simulointimalli, jonka avulla voidaan vertailla erilaisten ympäristönsuojelutoimenpiteiden vaikutusta väestön altistukseen. Malli kuvaa testilaskentojen mukaan väestön altistuksen vaihtelua hyvin ja mallin virheet jäävät väestötutkimusten otantavirheitä pienemmiksi lukuun ottamatta aivan korkeimpia altistustasoja. Mallin soveltuvuutta erilaisten toimenpiteiden vertailuun testattiin tarkastelemalla uudenaikaisten ilmanvaihtojärjestelmien tarjoamaa mahdollisuutta alentaa altistusta ulkoilman pienhiukkasille. Olettaen, että koko rakennuskannassa pääkaupunki-seudulla käytettäisiin tulevaisuudessa koneellista ilmanvaihtoa suodattimineen tavalla, joka on jo käytössä 1990-luvulla rakennetuissa toimistorakennuksissa, voitaisiin altistusta ulkoilman pienhiukkasille laskea 27 % vuosien 1996-97 tasosta. Suuruusluokaltaan tämä vastaa paikallisen liikenteen pakokaasupäästöjen vaikutusta. Rakennusten ilmanvaihdon kehittäminen vaikuttaa lisäksi kaukokulkeutuneisiin hiukkasiin.

Mallin vastaavuus mittauksiin testatuissa tapauksissa oli siis hyvä ja mallin osoitettiin soveltuvan erilaisten tulevaisuuskuvien vertailuun. Altistuksen arviointia ja mallien käyttöä osana ympäristöpolitiikan kehittämistä tulee lisätä.

Asiasanat: pienhiukkaset, altistuminen, mallintaminen, ilman saastuminen, terveysvaikutukset, kaupunkiväestö, simulointi, sisäilma, ilmanvaihtojärjestelmät, tutkimus

Hänninen, Otto: **Probabilistic Modelling of $PM_{2.5}$ Exposures in the Working Age Population of Helsinki Metropolitan Area**. Publications of the National Public Health Insitute, A10/2005. ISBN 951-740-521-0 (pfd 951-740-522-7), ISSN 0359-3584 (pdf 1458-6290).
http://www.ktl.fi/attachments/suomi/julkaisut/julkaisusarja_a/2005/2005a10.pdf

ABSTRACT

Fine particles are associated with hundreds of thousands annual deaths and significant increase in morbidity in Europe. Improvement of air quality and reduction of air pollution emissions are identified as the primary goals, but environmental policies can be targeted in different ways. It is clear, that optimal protection of public health is achieved by policy options reducing population exposures effectively. Besides air quality and associated temporal and spatial variability, the most important factor affecting exposures is population mobility. In traffic environments the proximity of emissions increases exposures, while in indoor environments concentrations of particles entering from outside are reduced by the building shell. Presence of indoor sources, however, may result in indoor concentrations orders of magnitude higher than outdoors.

In the current work a population exposure model was developed to compare the impact of alternative future policy scenarios on population exposures. Comparison with measurements showed that the model predicts the exposures and their variability well. The model errors were smaller than the statistical errors caused by random population sampling in an exposure study, apart from the highest few percentiles. Model applicability to policy evaluation was demonstrated by modelling the potential of ventilation systems equipped with effective particle filters to reduce exposures. Assuming the whole Helsinki metropolitan area building stock would be equipped with such mechanical ventilation systems that is already used in office buildings built in 1990's, the overall population exposure to ambient particles was reduced by 27 %. This is in the order of the effect of local traffic tailpipe emissions, which would have to be completely removed to achieve a similar net effect. Besides, building ventilation system affects also long-range transported particles.

Model correspondence with measurements was good and the model applicability to practical policy options comparison was demonstrated. The general conclusion of the work is that exposure assessment, using models when necessary, should be incorporated with development of effective environmental policies.

Subject terms: air pollution, air pollution, indoor, air pollutants, environmental, ventilation, evaluation studies, urban population, particle size

ACKNOWLEDGEMENTS

This work has been carried out in the inspiring environment of KTL Department of Environmental Health. For more than twenty years **Professor Jouko Tuomisto, M.D., Ph.D.,** guided the journey taking Finnish environmental health research into the international front row. His straight standing, patient, humble, envisioned contribution in workshops, meetings, and personal discussions together with his efforts in providing colleagues of mine and me the opportunity for this work merit my deepest thanks. **Professor Terttu Vartiainen, Ph.D.**, continues his work**.**

Young scientific thoughts grow from the fertile ground laid down by experience. For this I want to express my highest gratitude to my supervisors. My former university teacher, the head of the Air Research Laboratory, **Professor Matti Jantunen, Ph.D.,** gathered an international network for the current study and made it possible to put the latest achievements in exposure science into unequalled use in Europe. **Decan Juhani Ruuskanen, Professor, Ph.D.,** my teacher already in the late '80s in the University of Kuopio, provided his gentle and peaceful sense of reality and approving attitude in a most constructive way. **Dr. Erik Lebret, Ph.D.,** Head of the Unit of Environmental Epidemiology, RIVM, brought in expertise in exposure modelling, and his professional touch kept my work on track.

The reviewers of the theses, **Professor Jaakko Kukkonen, Ph.D.**, from the Finnish Meteorological Institute, and **Dr. Nicole Janssen, Ph.D.**, from the Dutch Institute for Public Health and the Environment, deserve my sincerest appreciation. They spent countless hours reading the work and provided significant insights that made it possible for me to condense and clarify many sections.

The opportunity to study and work together with **Kimmo Koistinen** for two decades is a corner stone of this work. Together we have faced challenges from university to business and science, and he has been my closest workmate and friend, for which I want to thank him.

The members of the *EXPOLIS*-Helsinki team, **Anu Kousa, Jouni Jurvelin, Tuulia Rotko, Tuija Stambej, Virpi Vuori**, **Tuula Pipinen** and **Tirre Hentinen**, the principal investigators **Klea Katsouyanni**, **Nino Künzli**, **Dennis Zmirou**, **Radim Srám**, and **Marco Maroni,** and our international colleagues **Hanneke Kruize, Oscar Breugelmans, Lucy Oglesby**, **Celine Boudet**, **Maria Caparis**, **Evi Samoli**, **Paolo Carrer**, **Domenico Cavallo**, and **Lambros Georgoulis** made my work not only possible but a truly unforgettable journey. Thank you.

In this project it is difficult to imagine success without the support from **Sari Alm** and **Anna-Maija Piippo**. Sari directed the Laboratory of Air Hygiene while Matti was away, and her experience from the first Finnish exposure studies was invaluable. Annukka practically run our unit; no internal financial report or a single trip abroad could have been completed without her.

Innumerable persons from KTL have supported my work, not possible to list all deserving, but amongst them I want to acknowledge the encouragement from **prof. Juha Pekkanen**, **Raimo O. Salonen**, **Päivi Aarnio**, **Jouni Tuomisto**, **Rufus Edwards**, **Raili Venäläinen**, **Matti Viluksela**, **Matti Vartiainen**, **Yuri Bruinen de Bruin**, **Vito Ilacqua** and **Tarja Yli-Tuomi**.

I want to thank my Mother **Auli Hänninen, M.D.**, and my Father **Professor Osmo Hänninen, M.D., Dr.Med.Sci., Ph.D.,** emeritus head of the Department of Physiology and former chancellor of the University of Kuopio. They set me high standards for pushing forward in life & science.

Last and most important thanks belong to **Maire** and our children, **Henri** and **Minttu**.

Kuopio, June 2005

Otto Hänninen

This work is based on the European Union 4th Framework Research and Technology Development Programme funded multi-centre study "Air Pollution Exposure Distributions of Adult Urban Populations in Europe" (*EXPOLIS*), and has been supported by EU contracts ENV4-CT96-0202 (*EXPOLIS*) and EVK4-CT-2002-00097 (FUMAPEX), Academy of Finland contract № 36586, and intramural funding from KTL and other participating institutes.

ABBREVIATIONS AND DEFINITIONS

These non-comprehensive definitions describe of the use of the terms in the current context.

AirPEX	Air Pollution Exposure model developed in RIVM (Freijer *et al*., 1998).
BS	Black Smoke. An optical measure of the blackness of a filter sample. Associated typically with diesel exhausts.
CA	California. A western state in the U.S.
CD-ROM	Compact Disk Read Only Memory. A CD-disk, typical capacity 650 MB.
CHAD	Consolidated Human Activity Database, a population time-activity database combined from several U.S. studies (McCurdy *et al*., 2000).
CIDB	Combined International Database; the main results from all centres. Available in MS-Access versions 95, 97, and 2000.
CO	Colorado. A state in the U.S.
CO	Carbon monoxide. Toxic gas emitted from incomplete combustion processes.
DOS	Disk Operating System by Microsoft, Inc. A personal computer operating system popular in the 1980's.
Direct mode	Exposure modelling in the current work using directly microenvironment concentration distributions (as opposed to *nested* mode).
EADB	*EXPOLIS* Access Database. The local database used for local data entry and management in each *EXPOLIS* centre. MS-Access version 95.
EC	European Community.
ED-XRF	Energy dispersive X-ray fluorescence (see also XRF).
EPA	U.S. Environmental Protection Agency.
ETS	Environmental Tobacco Smoke. Air pollution (PM, nicotine, CO, etc.) originating from different forms of burning tobacco products to which smoking and non-smoking subjects are exposed in the environment. The total tobacco smoke exposure of active smokers is significantly higher than their ETS exposure, created by themselves and fellow smokers.
EU	European Union.
EXPOLIS	Air Pollution Exposure Distributions within Adult Urban Populations in Europe –study. A multi-centre study conducted in seven cities in 1996-2000 (Jantunen *et al*., 1998).
GerES	German Exposure Survey. A German exposure research program (Seifert *et al*., 2000).
GIS	Geographical Information System. A computer software environment for handling spatially oriented data. E.g. MapInfo.
GPS	Global Positioning System, a satellite network and atomic clock based system for accurate real-time measurement of geographical locations.
GSM	Global System for Mobile Communications (originally Groupé System Mobile), a cellular telephone system.
H^+	Hydrogen ion. Cause of acidity.
HAPEM	Hazardous Air Pollutant Exposure Model by U.S. EPA.
HEDS	Human Exposure Database System, developed by U.S. EPA NERL.
Helsinki	Unless otherwise specifically indicated, the current work refers with this to the Helsinki metropolitan area, consisting of cities Helsinki, Espoo, Kauniainen, and Vantaa. Total population approximately 1 million.
IN	Indiana. A state in the U.S.

KTL	Finnish Public Health Institute (*Kansanterveyslaitos*; www.ktl.fi).
MB	Megabyte. A measure of computer memory device storage capacity. Defined alternatively as 1.000.000 bytes or 2^{20} (1.048.576) bytes depending on the source.
ME	Multilinear Engine. A type of principal component analysis (Paatero and Hopke, 2003).
MEM	Microenvironment monitor. A sampling device that is positioned in a specific micro-environment, typically a (room in the) residence, school, or workplace of the subject.
NC	North Carolina. An eastern state in the U.S.
NERL	National Exposure Research Laboratory of U.S. EPA.
Nested mode	Exposure modelling in the current work using ambient levels to model microenvironment concentrations (as opposed to *direct* mode).
NHEXAS	An exposure research program in 1990's in the U.S. (Clayton *et al*., 2002).
NJ	New Jersey. An eastern state in the U.S.
NO_2	Nitrogen dioxide. An air pollutant.
NV	Nevada. A state in the U.S.
NY	New York. An eastern state in the U.S.
O_3	Ozone. An air pollutant produced by photochemistry in the atmosphere.
ON	Ontario. An east-central province in Canada.
PAH	Polycyclic aromatic hydrocarbons.
PC	Personal Computer. A microprocessor-based computer dedicated to a single user. Originally developed by IBM, Inc. in 1982.
PCA	Principal Component Analysis. A statistical modelling technique.
PCP	Pentachlorophenol.
PEM	Personal exposure monitor. A sampling device that is carried by the subject.
PM, PM_{10}, $PM_{2.5}$	Particulate matter (with aerodynamic cut size diameter smaller than 10, 2.5 µm). Particles consisting of solid and liquid materials, suspended in the air.
PMF	Positive Matrix Factorization. A type of principal component analysis (Hopke *et al*, 2003)
pNEM	Probabilistic version of U.S. EPA National Exposure Model (NEM, Law *et al*. 1997)
PTEAM	Particle-TEAM study, Riverside, CA, U.S. (Özkaynak *et al*., 1996)
p-value	A statistical measure for the probability of an outcome being caused by mere chance.
r^2	Coefficient of determination. A statistical estimate for the fraction of variance being attributable to the independent variable(s) in a regression model.
RIVM	The Dutch Institute for Public Health and the Environment (*Rijksinstituut voor Volksgezondheid en Milieu*; www.rivm.nl)
RSP	Respirable suspended particles. Particulate matter suspended in the air capable of penetrating the respiratory system. Particle size defined differently in different sources, upper limit varying typically from 3.5 to 10 µm.
SD	Standard deviation. A statistical measure of variability of values in a data set.
SHAPE	Simulation of Human Activity and Pollutant Exposure, a probabilistic exposure model developed by Ott *et al*. (1988).
SHEDS	Stochastic Human Exposure and Dose Simulation model by U.S. EPA NERL (Burke *et al*., 2001).
SOP	Standard operating procedure. A quality assurance procedure and document.
TAD, TMAD	Time-(microenvironment-)activity diary. A diary filled by study subjects to record their locations and activities.

TEAM	Total Exposure Assessment Methodology –research program in U.S., started in 1980's.
THEES	Total Human Environmental Exposure Study conducted in Phillisburg, NJ in 1980's (Lioy *et al*. 1990).
THERdbASE	Total Human Exposure Database and Simulation Environment by U.S. EPA NERL (Pandian *et al*., 1990).
TN	Tennessee. A state in the U.S.
TSP	Total Suspended Particles. Particulate matter suspended in air, regardless of the particle size (i.e. including coarse particles up to tens of micrometers).
TX	Texas. A southern state in the U.S.
UK	United Kingdom, consisting of Great Britain and Northern Ireland.
U.S.	United States of America.
VA	Virginia. An eastern state in the U.S.
VOC	Volatile Organic Compounds. A heterogeneous group of innumerable volatile organic compounds, boiling points varying from 50-100°C to 240-260°C (WHO, 1989).
VT	Vermont. An eastern state in the U.S.
WA	Washington. A western state in the U.S.
WHO	World Health Organization of the United Nations.
XRF	X-ray fluorescence spectrometry. An analysis technique for determination of the elemental composition of samples of airborne PM.

MATHEMATICAL SYMBOLS

E	Time-weighted average exposure level [$\mu g\ m^{-3}$]
f	Fraction of time (spent in an microenvironment) [unitless]
C	Concentration [$\mu g\ m^{-3}$]; using subscripts: *a* ambient (outdoors) *ai* ambient originating particles in indoors *ig* indoor generated particles in indoors *i* indoor concentration (sum of ambient originating and indoor generated levels)
F_{inf}	Infiltration factor [unitless]; ratio of C_{ai} and C_a; using superscripts *S* sulphur-containing particles $PM_{2.5}$. fine particles
P	Penetration factor [unitless]
k	Decay rate (indoors) [h^{-1}]
a	Air exchange rate [h^{-1}]
V	Volume (of an indoor space, e.g. apartment) [m^3]
Q	Emission rate (source strength) [$\mu g\ h^{-1}$]
t	time [h]
β_0	Regression constant
β_1	Regression slope; using superscripts *S* sulphur-containing particles $PM_{2.5}$. fine particles

ORIGINAL PUBLICATIONS

This thesis is based on the following seven original articles, published in four peer reviewed scientific journals. The articles are referred in the text by Roman numerals I-VII.

I Jantunen, M.J., Hänninen, O.O., Katsouyanni, K., Knöppel, H., Künzli, N., Lebret, E., Maroni, M., Saarela, K., Srám, R., Zmirou, D., 1998. **Air pollution exposure in European cities: The *EXPOLIS*-study**. *Journal of Exposure Analysis and Environmental Epidemiology* 8 (4): 495-518.

II Kruize, H., Hänninen, O.O., Breugelmans, O., Lebret, E., Jantunen, M., 2003. **Description and demonstration of the *EXPOLIS* simulation model: Two examples of modeling population exposure to particulate matter**. *Journal of Exposure Analysis and Environmental Epidemiology* 13 (2): 87-99.

III Hänninen, O.O., Kruize, H., Lebret, E., Jantunen, M., 2003. ***EXPOLIS* Simulation Model: $PM_{2.5}$ Application and Comparison with Measurements in Helsinki**. *Journal of Exposure Analysis and Environmental Epidemiology* 13 (1): 74-85.

IV Hänninen, O.O., Lebret, E., Ilacqua, V., Katsouyanni, K., Künzli, N., Srám, R., Jantunen, M.J., 2004. **Infiltration of ambient $PM_{2.5}$ and levels of indoor generated non-ETS $PM_{2.5}$ in residences of four European cities**. *Atmospheric Environment*, 38 (37): 6411-6423.

V Hänninen, O.O., Lebret, E., Tuomisto, J.T., and Jantunen, M.J., 2005. **Characterization of Model Error in the Simulation of $PM_{2.5}$ Exposure Distributions of the Working Age Population in Helsinki, Finland**. *JAWMA*. 55: 446-457.

VI Hänninen, O.O., Palonen, J., Tuomisto, J., Yli-Tuomi, T., Seppänen, O., Jantunen, M.J., 2005. **Reduction potential of urban $PM_{2.5}$ mortality risk using modern ventilation systems in buildings**. *Indoor Air*. In press (published as *OnlineEarly*).

VII Hänninen, O.O., Alm, S., Katsouyanni, K., Künzli, N., Maroni, M., Nieuwenhuijsen, M.J., Saarela, K., Srám, R., Zmirou, D., Jantunen, M.J., 2004. **The *EXPOLIS* Study: Implications for exposure research and environmental policy in Europe**. *Journal of Exposure Analysis and Environmental Epidemiology*, 14: 440-456.

CONTENTS

1. INTRODUCTION

A glimpse for perspective. Since prehistoric times it's been known to man that the smoke from flames is irritating – anyone who ever sat in front of an open fire outdoors knows that it makes your eyes bleed and throat sore; it has never been news that air pollution is bad for health. The three major factors that have increased exposures to air pollution during the last millenniums are urbanization, industrialization, and the drastic increase of traffic. Urbanization started well in the first millennium before Christ. Growth of the cities during the following two millennia gradually increased the problems of pollution. Industrialization boomed towards the end of the second millennium, starting in the 18th and 19th centuries, but still in those days, merely domestic heating was a significant problem for air quality; a fireplace existed in almost every room of every inhabited building. Photographs from late 19th and early 20th century taken over towns during days when heating was needed, demonstrate the poor state of air quality of that time. The third major step in worsening the air pollution was taken so late as early in the 20th century by the wide acceptance of the use of combustion engine.

The air pollution problem peaked in unfavourable meteorological conditions in places like Meuse Valley, Belgium (Dec. 1-5, 1930, 60 deaths), Donora, Pennsylvania, U.S., (Oct. 27-30, 1948, 20 deaths), and finally in London, UK, (Dec. 5-9, 1952, 3000 excess deaths, added to the one thousand of normally expected ones for such a period) (Bell and Davis, 2001). Severe wide-spread public health effects during these extreme air pollution episodes, including death of thousands, demonstrated beyond any doubt the acute harmfulness of modern air pollution to human health.

Fighting air pollution. In the next decades successful programs were launched to control air pollution, first in the developed world, and then towards the end of the century also globally. Political groups were founded targeting environmental protection in contrast to the struggle between the social classes in the beginning of the century. International collaboration started to fight global pollution and agreements were made to implement new low emission technologies.

Sulphur dioxide was one of the main pollutants that the emission abatement programs focused on in the 1970's. Emissions in many countries were dropped by tens of percents by the end of the century despite of increasing production and energy consumption, but globally the sulphur emissions continued to grow (Lefohn et al., 1999). Since the 1970's strict emission reduction requirements have been set for the auto industry, turning the tailpipe emissions into a slowly lowering trail in spite of the continuously increasing number of vehicles and kilometres. So by the end of the century the developed world had conquered the problem of air pollution – or had it?

The problem persists. After the London episode air quality monitoring has become standard practice in all cities and towns with more than hundred thousand inhabitants in the developed world. Together with the ever-increasing number of details of data collected by health authorities from populations of hundreds of millions, the accumulating data from these air quality monitoring networks has made it possible to study the effects of air pollution on human health with unforeseen sensitivity. During the last decade of the 20th century it became evident that even the prevailed relatively low levels or air pollution were still significantly associated with mortality and other health consequences in urban populations of the developed world. The number of premature deaths associated with air pollution was estimated to be tens of thousands annually in North America (Pope et al., 2002;Pope et al., 1995;Dockery et al., 1993) and in Europe (Samoli et al., 2005;Katsouyanni et al., 2001;Katsouyanni et al., 1997). The most significant association has been repeatedly found for particulate matter (PM), especially fine particles ($PM_{2.5}$) (WHO, 2002;Ezzati et al., 2002).

At the same time that the developed world realized that air pollution is an additional risk factor that increases the statistical probability of death and other adverse health effects caused primarily by cardio-vascular and respiratory diseases, the role of exposure as the actual causal link in the chain from emissions to the health effects became more clearly acknowledged (Ott, 1995). Health effects really having causal connections with the air pollution must be caused by the actual exposures of the affected individuals. Therefore reductions in the health risks must occur via reductions in the exposures – and sometimes emission-based policies have shown to have only negligible effects on exposures (Jantunen, 1998).

Particles originate from a number of different sources, including energy production, industry, vehicles, resuspension of dust, natural sources, and many sources indoors. In terms of emission tons the indoor sources are typically negligible, but their effect on indoor concentrations may be remarkable. Together with the fact that urban populations spend a majority of their time indoors makes the indoor exposures significant, and in some cases totally dominating. In the beginning of the current decade it became obvious that the health effects of ambient and indoor generated pollution should be considered separately (Wilson et al., 2000). The concentrations caused by these do not correlate with each other; the particles have different chemical and physical compositions, presumably different toxicities, and definitely very different controlling options. Consequently, the questions that have risen to a central role in the public health protection concerning particulate matter pollution are:

- Are all particles (equally) harmful?
- What kinds of particles are (more) harmful?
- To whom are they (most) harmful?
- How to reduce the harmful exposures of sensitive population groups efficiently?

Effective public health protection policies must be based on a clear understanding of population exposures and the underlying factors, including microenvironment concentrations and population time-activity (Lioy, 1990). Optimal reduction of exposures can then be achieved by comparing alternative control strategies in terms of costs and exposures. Comparison of hypothetical policy options is really possible only by using models (Ott, 1995;Seifert, 1995;Lioy, 1991;Ryan, 1991;Ott, 1985). Requirements for the reliability of such models, when used in selecting expensive and potentially invasive and limiting policies, are high. Such models must be carefully evaluated against experimental data in existing setups, including a thorough peer review before the models are applied. This is exactly what the current work is about.

2. Aims of the Dissertation

The overall objective of the current doctoral dissertation work was to develop and evaluate a modelling methodology for the estimation of urban population exposures to fine particulate matter in current and future scenarios, including hypothetical scenarios supporting policy options evaluation. The work uses $PM_{2.5}$ data from Helsinki for these purposes.

The specific steps required meeting this overall objective include the following tasks. The original articles that tackle each task in detail are listed in parentheses.

1. Design and carry out a population-based exposure study to collect data on urban population exposure levels, microenvironment concentrations, and population time-activity for development and validation of a probabilistic exposure simulation model (**I**),

2. Develop a conceptual model and supporting software framework for implementing probabilistic exposure models (**II**),

3. Create data analysis methods to estimate model inputs from measured variables, including partitioning of microenvironment concentrations into ambient and indoor generated fractions and analysis of infiltration factors, and selection of appropriate population groups for time-activity modelling (**III**, **IV**, **V**),

4. Study the accuracy of the simulation model by comparing model results with the measured personal exposure distributions in a random population sample (**II, III**, **V**),

5. Clarify the concepts of model evaluation by differentiating between the concepts of model error and assessment of uncertainty (**V**) and discuss the use of independent data,

6. Demonstrate the use of a simulation model in a policy relevant setup by applying it for a selected exposure reduction scenario (**VI**), and

7. Discuss development of effective environmental policies by using exposure analysis and models (**VII**).

3. BACKGROUND

Focus shift from emissions to exposures. Environmental policies are facing new integration and optimization challenges in the 21st century. Health effects which have a causal relationship with air pollution must be caused by the actual personal exposures of the affected individuals (Spengler and Soczek, 1984;Duan, 1982;e.g. Ott, 1982). During the past decade it became clear that straightforward emission reductions are not always cost-effective means to reduce public health risks – in fact they can be costly and yet very ineffective. Perhaps the best-known example of this is the benzene exposure case in Northern California (Jantunen, 1998;Ott, 1995). In the early 1990's the San Francisco Bay Area Air Quality Management District considered that of all ambient air pollutants benzene was contributing the largest risk to the Bay area residents. The Board called for a 50 % reduction in benzene emissions from the largest industrial point sources. However, a source apportionment of the benzene exposures revealed that only 25 % of the exposures were of ambient origin, and only 3 % originated from the point sources. Majority of the exposures came from traffic, tobacco smoke, and various indoor sources and the 50 % reduction in point source emissions yielded only an indistinguishable 1.5 % reduction in the population's exposure and corresponding cancer risk.

The Exposure Paradox. The association between ambient PM pollution and health was observed in epidemiological studies using air quality monitoring data from fixed outdoor sites to describe population exposures. Personal exposures are, however, modified by individual behaviour, time spent in traffic, and especially the indoor environments visited. Many studies have confirmed that personal exposures correlate poorly with ambient levels measured at fixed monitoring sites (Alm et al., 2001;Koistinen et al., 2001;Oglesby et al., 2000;Pellizzari et al., 1999;Wallace, 1996;Morandi et al., 1988;Spengler et al., 1985;Sexton et al., 1984). At first, this was seen as a major objection to the epidemiological finding itself, before it was realized that the health effects associated with fixed station levels are those caused by the particles of ambient origin. Fixed urban background monitoring stations represent well the average population exposures to these particles (Wilson et al., 2000). Other particles, not correlating with the ambient levels, may then have health effects of their own (Mage, 2001;Wilson et al., 2000), but due to the methodological difficulties in assessing these, the toxicities of indoor generated particles – except for ETS (e.g. Zhang et al., 2005) – are still largely unknown.

The main conclusion from these findings is the fact that urban populations are exposed to a large variety of different kinds of particles from different sources; the particles may have different toxicities, and different sources certainly have different control mechanisms. Therefore it is important to assess these exposures separately (Ott, 1995;Sexton et al., 1995a;Wallace, 1993;Girman et al., 1989).

Understanding the underlying source and exposure factors associated with the health effects is crucial for the success in both exposure modelling and in public health risk management. On the population level there are dozens of time-activity factors, and factors that affect local microenvironment concentrations, that together create the individual exposure levels. Some major milestones in the particulate matter exposure analysis studying these factors are reviewed in the following section.

3.1. Population-Based Exposure Research

During 1980-2000 a number field studies were conducted first in the U.S. and later in Europe to collect population-based data for exposure analysis. The following reviews some of the studies that either had a profound contribution to exposure analysis for particulate matter, the design of the current work, or that have been progressing parallel to our study. Some of these studies, which have either preceded the current study and influenced its design, or have been conducted parallel or later to it, are summarized in Table 1 in chronological order and compared with *EXPOLIS*. The studies are identified primarily by the project acronym (if available; otherwise by location or primary researcher).

The reviewed studies can be classified into two categories: (i) those focusing on total exposures of pollutants having multiple routes of entry into the human body, including besides inhalation also dietary and skin exposures. From the point of view of the current work, some of these studies (e.g. TEAM, NHEXAS, GerES, see definitions and details below) have been significant in terms of developing concepts and methods for population exposure assessment. The second category (ii) includes studies of inhalation exposures focusing more or less on particulate matter.

Important exposure concepts developed along the two active decades of population exposure research include exposure distributions, intra- and inter-personal variation, source apportionment, ambient and indoor sources, microenvironment assessment and modelling, indoor-outdoor relationships, and infiltration of particles. Many of these concepts are directly utilized in the modelling in the current work.

Northern America. Early milestones in PM exposure research were set in late 1970's and early 1980's. One of these was the Harvard Six Cities study, a successful long-term research project that produced one of the most significant epidemiological findings on the association between ambient PM and health (Dockery et al., 1993). As a small part of this project, also the indoor-outdoor relationships of respirable particles (RSP) were studied using data from 68 residences over one-year period (Dockery and Spengler, 1981). Somewhat later a similar study was conducted in Suffolk and Onondaga counties in the New York State ERDA –study (Koutrakis et al., 1992), where $PM_{2.5}$ measurements, now including 16 elemental constituents, were conducted in 178 residences. Both of these studies were used to develop models for the indoor-outdoor relationship of particles (see modelling details in **IV**).

One of the important aspects studied in the 1980's was the relationship of short-term and long-term exposures. When short-term exposure measurements are conducted on a population sample, the observed variance of personal exposures includes two components: inter-personal variance (i.e. variance in exposures of different subjects during the same day) and intra-personal variance (variance of exposures of the same persons over different days). This issue was tackled in the Waterbury, Kingston-Harriman, and Phillisburg studies (Table 1). Exposures to respirable suspended particles (RSP) were measured in Waterbury (VT) using 48 subjects (Sexton et al., 1984). Each subject was sampled every other day for two weeks, giving information on the intra-personal day-to-day variation. In Kingston and Harriman (TN) the size of the population sample was 97 (Spengler et al., 1985). In this study RSP personal exposures were monitored for three non-consecutive days together with simultaneous residential indoor concentrations. The longitudinal variation of personal exposures to PM_{10} was studied also in the THEES study in Phillisburg (NJ) (Lioy et al., 1990). The population sample was rather small (14) and not randomly selected, but residential indoor and outdoor concentrations and personal exposures were followed from day to day for a two-week period. Thus the results formed a 14x14 matrix of person days, allowing for analysis of the inter- and intra-day variances of the personal exposures and their relationships to ambient PM_{10} levels.

Table 1. Summary of design features of selected exposure studies focusing on particulate matter (in chronological order from left to right).

	Kingston-Harriman	**Waterbury**	**THEES**	**PTEAM**	**Phillips *et al.* ETS studies**	**Janssen *et al.***	**ULTRA**	**Toronto, Indianapolis manganese**	***EXPOLIS***	**RIOPA**
Timeframe in relation to *EXPOLIS*	Earlier	Earlier	Earlier	Earlier	Earlier	Earlier	Parallel	Parallel	-	Later
Cities/areas	Kingston and Harriman (TN)	Waterbury (VT)	Phillisburg (NJ)	Riverside (CA)	8 European cities	Amsterdam, Wageningen	Amsterdam, Helsinki	Toronto (ON) Indianapolis (IN)	7 European cities	Houston (TX) Los Angeles (CA) Elizabeth (NJ)
Survey year(s)	1981	1982	1988	1990	1992-95	1994-95	1996-1999	1995-96	1996-2000	1999-2000
Compound(s) [1]	RSP	RSP	PM_{10} bentso(a)-pyrene	PM_{10}, $PM_{2.5}$ (RI+RO)	ETS, RSP	PM_{10}	ultrafines (<0.1µm), $PM_{2.5}$	PM_{10} $PM_{2.5}$ manganese	$PM_{2.5}$ + elements + BS 30 VOCs NO_2, CO	$PM_{2.5}$ VOC carbonyls
Population, age range	random, non-smoking adults	voluntary, nonsmoking	voluntary, 28-, nonsmoking	random	random, non-smoking adults	children, elderly volunteers	elderly cardiac patients	random, 16-	random, 25-55	adults & children
Nr of subjects	97	48	14	178	188-255 per city	37 adults, 45 children	82	732 Toronto 240 Indianapolis	501	212 homes
Seasonal time frame	spring	winter-spring	winter	fall	various seasons	various seasons	various seasons	one year (ON) summer (IN)	one year	one year
Air sampling time	24 hours	24 hours	24 hours	2x12 hours	24 hours	24 hours	24 hours	3 days	48 hours	48 hours
Longitudinal sampling	3 non-consequtive days	every other day for two weeks	14 consequtive days	consequtive day+night	none	4-8 measure-ments	upto 13 measurements	repetition with random lag for a subsample of 190 in Toronto	2 consequtive days for CO	repetition after 3 month lag for a subsample
Air sampling micro-environments [2]	RI, P	RI, RO, P	RI, RO, P	RI, RO, P	P	RI, P, A class rooms	RI, P, A	RI, RO, A, P	RI, RO, W, P	RI, RO, P
Reference(s)	Spengler et al. 1985	Sexton et al. 1984	Lioy et al. 1990	Clayton et al. 1993	Phillips et al. 1994-1999	Janssen et al. 1997-1999	Pekkanen et al. 2002	Pellizzari et al. 1999	I	Weisel et al. 2005

[1] See Abbreviations for symbol definitions

[2] RI = Residential indoor, RO = Residential outdoor, P = Personal, A = Ambient, W = Workplace indoor

Perhaps the best-known exposure research program in the 1980's was the Total Exposure Assessment Methodology (TEAM) focusing on multi-route exposures. Inhalation exposure compounds like carbon monoxide (CO), nitrogen dioxide (NO_2), total suspended particles (TSP), respirable (PM_{10}) and fine particles ($PM_{2.5}$), acid aerosols, environmental tobacco smoke (ETS), and ozone were included, but in a minor role in these studies and benefited mainly from the methodological developments in population exposure assessment. The other exposure routes, dietary and skin exposures, however, have a profound role for many other substances including VOC's (e.g. benzene, toluene, limonene, styrene, chlorinated hydrocarbons, different forms of xylene), pentachlorophenol (PCP), lead, cadmium, polycyclic aromatic hydrocarbons (PAH), and pesticides. Population samples in the TEAM studies varied from small and non-representative to quite large random or stratified random samples. Inhalation exposures were measured typically for one day, but some designs allowed also for longitudinal exposure analyses (Hartwell et al., 1987;Spengler et al., 1985;Sexton et al., 1984).

Concerning PM exposures, the most important study before *EXPOLIS* was initiated by the series of earlier TEAM studies and was called Particle TEAM (PTEAM, Table 1). This study was conducted in 1990 in Riverside (CA) using a random population sample of 178 subjects. Residential indoor and outdoor PM_{10} levels were monitored for two consecutive 12-hour periods (day and night) together with corresponding personal exposures. Residential indoor and outdoor $PM_{2.5}$ concentrations were also measured, allowing for modelling of $PM_{2.5}$ exposures and assessment of the ratio of PM_{10} and $PM_{2.5}$ exposures. Elemental compositions were also determined and used for infiltration modelling and analysis of the decay and penetration terms required by the mass-balance model (Özkaynak et al., 1996;Clayton et al., 1993;Thomas et al., 1993;Clayton et al., 1991). Similar analysis was developed further using the *EXPOLIS* data in **IV**.

Parallel to the current work was conducted the Ethyl Corporation funded study by Research Triangle Institute (NC) for $PM_{2.5}$ and manganese exposures in Toronto (Ontario, Canada; Table 1). This is the largest population based PM study so far with it's 732 measured subjects. Manganese used as a gasoline additive in Canada was suspected to have public health effects. A sub sample of 190 subjects was measured again within the one-year study period with a random lag. Besides personal levels also residential concentrations were measured indoors and outdoors. Each person was monitored for 3-day period. Supplementary data on traffic, meteorology, occupation, and time activity of subjects were also collected. Databases were developed to store the data and to support the data analysis. (Pellizzari et al., 1999;Clayton et al., 1999a)

A parallel manganese study was conducted in Indianapolis (IN; Table 1) to get comparable exposure levels from a city where the same gasoline additive was not used (Pellizzari et al., 2001a). In general the Indianapolis PM levels were somewhat higher than the corresponding levels in Toronto. The Mn levels, as expected, were lower in Indianapolis, especially when excluding occupational exposures. All PM_{10} levels in Toronto and microenvironment PM_{10} levels in Indianapolis were clearly lower than the PM_{10} levels in PTEAM study, Riverside (Pellizzari et al., 2001a).

Another significant U.S. program in population based exposure research in general, but having only a minor contribution to PM research, is the National Human Exposure Assessment Survey (NHEXAS) that followed the TEAM studies in assessing multi-route multi-media exposures. NHEXAS targeted the whole population of the U.S. and to this end developed geographical, urban-rural and sociodemographic stratification levels for population sampling. In respect to pollutants studied, NHEXAS was more focused than the TEAM-studies; there was a clear view that the compounds selected for such a large study should be documented or suspected human health hazards and there should be a need for exposure information for them. Pollutants of especial interest according to these criteria included benzene, pentachlorophenol, formaldehyde, mercury, and lead (Lioy and Pellizzari, 1995). Besides these, dozens of heavy metals, VOCs and pesticides were considered (Callahan et al., 1995;Sexton et al., 1995b). NHEXAS acknowledged the need to characterize population distributions of exposures, including information on both the base line exposures as well as the high percentiles and estimates on the highest exposed individual levels for both the general population as well as for population sub groups. The program was divided into three phases. Phase I targeted planning, designing and testing, phase II implemented the national survey and in depth special studies were allocated to phase III. After that, NHEXAS was envisioned to be a continuous research activity, to be repeated every three to six years. (Sexton et al., 1995b)

NHEXAS phase I studies were conducted in three different areas; (i) Arizona, (ii) EPA region 5, consisting of six states in the Great Lakes area, and (iii) Maryland. NHEXAS Arizona measured residential indoor, outdoor and personal concentrations of 25 metals, 4 pestisides and 25 VOCs for 175 subjects (study phase 3). The measurements were conducted during all seasons. (Gordon et al., 1999;Robertson et al., 1999;O'Rourke et al., 1999a;O'Rourke et al., 1999b). The NHEXAS EPA region 5 study panned six states, where selected metals and 4 VOCs were measured for a random sample of 250 subjects during an 18-month period in 1995-97. Six-day samples of residential indoor, outdoor and personal VOC levels were collected besides extensive set of other samples. (Clayton et al., 2002;Pellizzari et al., 2001b;Clayton et al., 1999b;Pellizzari et al., 1995). In Maryland the NHEXAS studies were more focused on selected specific issues. Buck et al. (1995) studied statistical aspects of estimating long-term exposures from short-term measurements. MacIntosh et al. (2001) and Pang et al. (2002) studied population exposures to pesticides, especially chlorpyrifos. Inhalation exposure related 24-hour measurements were conducted only in residential indoors of 80 subjects during a one-year study period. Longitudinal aspects were studied by repeating measurements on population sub samples up to six times.

The most recent PM study is the Relationships of Indoor, Outdoor, and Personal Air (RIOPA, Table 1) study in U.S. The concentrations of 18 volatile organic compounds (VOCs), 17 carbonyl compounds, and fine particulate matter mass ($PM_{2.5}$) were measured using 48-h outdoor, indoor and personal air samples collected simultaneously. PM2.5 mass, as well as several component species (elemental carbon, organic carbon, polyaromatic hydrocarbons, and elemental analysis) were also measured in 1999-2000 in Houston (TX), Los Angeles (CA) and Elizabeth (NJ) in 212 non-randomly sampled homes. Personal samples were collected from non-smoking adults and a portion of children living in the target homes. The population sample was stratified according to the residence location in relationship to major freeways, industry and other recognised emission sources. (Meng et al., 2005;Weisel et al., 2005)

Analysis results of the RIOPA data have just started to appear in the published literature. The first results include similar analysis of indoor-outdoor relationships of $PM_{2.5}$ levels that was earlier presented by Dockery and Spengler (1981) and Koutrakis et al (1992), and that was conducted also in the *EXPOLIS* study (**IV**).

Europe. One of the most significant early exposure studies in Europe were the German Environmental Surveys (GerES) that was first conducted in the former West Germany 1985-86 and then repeated in 1990-92, now including the whole united Germany. GerES studied representative population samples for exposures to dozens of metals and other toxicants.

Inhalation exposures to VOCs were measured only on a sub sample of 113 adult subjects, PM exposures not at all. (Hoffmann et al., 2000a;Seifert et al., 2000a;Hoffmann et al., 2000b;Seifert et al., 2000b)

In Finland the first exposure studies were conducted by Alm *et al*. (2001;2000;1998;1994) and Mukala *et al*. (2000;1996). They measured personal carbon monoxide and nitrogen dioxide exposures of pre school children panels in Helsinki in 1990-91. Personal NO_2 levels were found to be lower than levels at the day care centres and the fixed station levels. Personal CO levels were higher than fixed station levels, and they were affected by the presence of gas stove at home. Respiratory symptoms were also connected to NO_2 exposures. Both NO_2 and CO exposures were affected by tobacco smoking in the home. These studies had a significant contribution for the practical implementation of the *EXPOLIS* studies.

A significant number of PM exposure studies in Europe were conducted by Phillips et al. in more than a half dozen European cities in collaboration with local institutes in each city (Table 1). These studies, however, were solely focused on ETS and nicotine exposures. The population samples were fairly large and representative in all cities (188-255 subjects per study), including only non-smoking subjects. Particle concentrations were measured mostly with cyclone pre-separator with 50% removal efficiency at 3.5 µm (the earliest study used no pre-separator and very low flow rate). Besides gravimetric RSP particle measurement various analytical methods were used to measure tobacco smoke originating particle concentrations (ultraviolet, fluorescence and solanesol measurements). (Phillips et al., 1999;1998a;1998b;1997a;1997b;1996;1994)

Important early European PM exposure studies were conducted by Janssen et al in the Netherlands (Table 1). They measured the $PM_{2.5}$ and PM_{10} exposures of school children and elderly people in Wageningen and Amsterdam in 1994-95. Panels of 45 children and 37 adults were sampled during 4-8 periods for 24 hours. Besides personal and residential levels, also concentrations in the school classrooms were measured (1999a;1999b;1998a;1998b;1997a;1997b). From the point of view of the *EXPOLIS* study some experience in the development of silent microenvironment and personal monitors were acquired from the Dutch experiences. Data analysis benefited, too, from the publications that appeared in the literature during the active period of *EXPOLIS* data analysis.

The Dutch studies were followed by the Exposure and risk assessment for fine and ultrafine particles in ambient air (ULTRA, Table 1). Cohorts of elderly cardiovascular patients were followed for six months in Amsterdam and Helsinki, including biweekly health inspection and ultrafine PM and $PM_{2.5}$ exposure measurements (Vallius et al., 2003;Pekkanen et al., 2002;Ruuskanen et al., 2001;Janssen et al., 2000).

3.2. Databases Supporting Exposure Modelling

The enormous amounts of valuable data produced in the population based exposure studies could potentially be utilised very effectively in exposure analysis outside the original study scope, if only the data was properly documented and made available (Burke et al., 1992). The value of databases designed for this purpose has been recognized since early 1990's (Sexton et al., 1994;Burke et al., 1992;Graham et al., 1992;Sexton et al., 1992), when the revolution brought by the Internet-based networking really started to make a difference in the ways that exposure related data is collected and stored. Due to the technical nature of such databases, however, little has been written about them in the scientific literature.

A lot of effort was put in the current work in developing a researcher-friendly, efficient, and reliable database system for collecting, storing, and distributing the various subsets of data from the *EXPOLIS* centres. The databases described in the Material and Methods –section have been used in data analysis for dozens of scientific papers, and in preparation a dozen doctoral dissertations. Therefore a short review of the thin literature concerning such databases is appropriate here to foster the use and publication of exposure databases to maximise the usability of data collected on public funding.

The role of exposure databases in exposure analysis and exposure model development – the context for the current work – is depicted in Figure 1. The database provides data needed both for the process of constructing the model as well as data for the model runs.

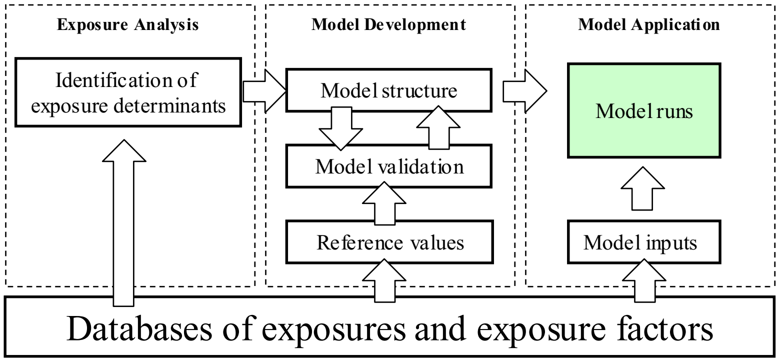

Figure 1. The triple role of exposure database s in exposure model development.

This topic was so urgent in the early 1990's that a workshop designed specifically to examine exposure-related databases was conducted in January 21-23, 1992, in Virginia Beach (VA). Participants, including scientists from federal and state agencies, the private sector and academic community, examined the utility of existing databases from different perspectives. Sexton *et al.* (1992) concluded that the existing databases of that time contained a substantial amount of relevant information, but that it was clear that the quality of the data was inconsistent and it was difficult to access the data. These statements are still valid. The studied systems demonstrated a striking absence of data on actual human exposures – a factor that has improved since. *EXPOLIS* database is one of the European milestones in this area.

Graham *et al.* (1992) recommended in the Virginia Beach workshop risk management workgroup that more human exposure measurement studies should be conducted and that new databases should be developed to meet critical data needs. The databases should emphasize quality assurance and control and they should be accessible to exposure and risk assessors. These are exactly the driving motivation for the current study: combination of conduction a population based European exposure study and development of an extensive exposure database for exposure analysis, modelling, and model validation purposes.

One of the extensive exposure databases developed based on these needs was the Total Human Exposure Database and Simulation Environment (THERdbASE) by U.S. EPA's National Exposure Research Laboratory, Las Vegas (NV) THERdbASE started as a DOS-based database system for information gathered in the TEAM studies to allow for (i) an ordered storage base for exposure-related environmental data and (ii) a convenient base for building total human exposure models (Pandian et al., 1990). In the 1990's the system evolved into a Windows based system capable of handling large databases and complex models in a networked PC-environment. Number of models and a variety of databases, including selected 1990 U.S. Census data were incorporated into one software platform. The database was peer-reviewed by a panel of national experts in December 1997. The database was downloadable from the Internet till 2004 when EPA dropped support for it, and it was adopted as a standard platform for exposure modelling across many offices within the U.S. EPA. (http://www.epa.gov/heasd/edrb/therd/therd-home.htm)

To survey the availability and quality of federally sponsored databases in the U.S. Sexton *et al.* (1994) made an inventory of databases potentially relevant for estimating human exposures to environmental agents. The inventory, reviewing and classifying 67 American databases, was compiled through a joint effort of EPA, the National Center for Health Statistics, and the Agency for Toxic Substances and Disease Registry. The inventory allowed for comparison of databases according to (i) type of exposure estimators, (ii) sample/media types, (iii) compounds, (iv) geographic scope and location coding (e.g. latitude/longitude, zip code, county) and (v) sampling frequency. The inventory showed that a significant number of the data systems contained useful information for exposure analysis, but it also was apparent that the data varied substantially according to the relevance, quality, and availability. Few databases collected representative population samples.

In the area of population time-activity the National Exposure Research Laboratory (NERL), EPA, developed the Consolidated Human Activity Database (CHAD). CHAD combined originally data from 12 U.S. studies related to human activities. CHAD, accessible in the Internet at http://www.epa.gov/chadnet1/, contains data from pre-existing human activity studies that were collected at city, state, and national levels. CHAD is intended to be an input data source for exposure/intake dose modelling and statistical analysis. CHAD is a master database providing access to other human activity databases using a consistent format. This facilitates access and retrieval of activity and questionnaire information from databases that EPA currently uses in its regulatory analyses. (McCurdy et al., 2000)

NERL produced also the Human Exposure Database System (HEDS), which is putting the NHEXAS data on-line. NHEXAS data was originally managed independently in different centres (Lebowitz et al., 1995). HEDS contains chemical composition data for air, soil drinking water, house dust, food, beverage, blood and urine (Robertson et al., 2001). The data includes pesticides, metals, VOCs and polynuclear aromatic hydrocarbons (PAHs). Questionnaire and diary responses are also included, addressing residential, life style demographic, occupational and health characteristics, time activity patterns and food consumption information (Robertson et al., 2001). HEDS is on-line at http://www.epa.gov/heds.

3.3. Theoretical Context for Exposure Modelling

Exposure models are used to estimate the concentrations of chemicals or other substances in an exposure media when in contact with the target subject. The media may be a surface becoming into contact with the skin or it may be e.g. foodstuff entering the digestive system. In the current work, focusing on inhalation exposures to fine particles, exposure is defined as airborne particle concentration in the breathing air of the subjects.

Exposure models may be developed to estimate exposures of individuals, susceptible population groups, or entire populations. They may estimate exposures as continuous variables, or integrate over time from short-term periods like minutes and hours to days, to long-term periods like years to lifetime. Modelled exposure variables may include instant values, mean exposure levels, and distribution parameters like standard deviations, quartiles, and percentiles. Consequently, exposure models range over a wide variety of complexity, approach, inputs, and outputs as discussed shortly below to put the current work into perspective with alternative methods and approaches.

3.3.1. What Are Exposure Models Needed For?

Exposure is the mediating link between man and the environment. Modelling of exposures is of interest to both exposure scientists as well as those in charge of developing environmental and public health policies. Modelling can serve three purposes:

- Understanding a phenomenon
- Estimation in lack of measurements
- Forecasting

The first category is mainly of interest to scientists, for whom sometimes merely a weak but statistically significant correlation between two variables is an interesting finding. The second one may interest both user groups equally and includes exposure modelling for epidemiology and risk assessment. The last one, forecasting, can be interpreted as a special case of the second, where the lack of measurements is due to the fact of looking into future. Forecasting may concern short (hours to days) or long-term (years to lifetime) models. Some approaches integrate modelling and measurements, e.g. data assimilation in meteorological models. Models of the third category belong to the most important tools for formulating science-based and effective public health policies (Ott, 1984).

3.3.2. Causality and Statistics

Models considered here are numerical constructs, quantifying the relationships between independent and dependent phenomena based on a theory. Independent phenomena, or events, are entered into the model as values of input variables to estimate the numerical values describing the corresponding dependent events. The dependent events that are statistically, logically, or causally related to the inputs, are then described using the output variables.

In his classical examination on exposure modelling principles, Ryan (1991) categorized exposure models into three categories: (i) *statistical*; (ii) *physical*; and (iii) *physical-stochastic*. Selection of the best approach for each modelling task is driven by the relationships of the independent and dependent variables – exposures and their determinants - in the target system. Dependency of variables can result from three alternative relationships depicted in Figure 2. The first rows represent causal relationships, where the state of the output variable is directly or indirectly caused (or more often in reality: influenced) by an input variable. In the third row the Effect 1 is irrelevant in causal sense and could be ignored, if only some alternative variable would be available to describe the underlying cause. However, in lack of such a variable the statistical relationship resulting for Effects 1 and 2 can be used for prediction of Effect 2 when observations of Effect 1 are available. A classical example of the last type of relationship is the correlation of ice cream consumption and drowning deaths –both are (causally) influenced by a warm weather, leading to an apparent relationship. In this case the true causal variable is measurable and should be used.

Relationship type	Input	Underlying	Output	Most suitable modelling
1. Direct causal	**Cause** →		**Effect**	Physical
2. Indirect causal	**Cause** →	Effect 1 →	**Effect 2**	Physical & Statistical
3. Common cause	**Effect 1** ←	Cause →	**Effect 2**	Statistical

Figure 2. Different relationships between model variables and model types suitable for them.

In reality, usually there are many independent phenomena affecting the dependent one, and in many cases feedback loops connect the output back to (some of) the inputs. The more direct and simple the causal chain is, the easier it is to model. An increasing number of intermediate variables in the causal chain shifts the relationship towards diminishing causality and weaker statistical correlation. Causal models are generally more reliable than models based on statistical associations exactly due to the increasing complexity of the chain of events in between the input and output variables. Physical modellers often take this as a disadvantage of statistical models, but as Ryan (1991) points in his Venn-diagram depicting the rather small overlapping application area of statistical and physical models, it is more reflecting the nature of the modelled phenomena for which the statistical approach becomes handy.

3.3.3. Researchers Standard Tools: Statistical Models

Statistical models are standard tools of scientists. When the actual causal mechanisms in the system under study are not yet known, they can be revealed by building hypotheses based on current theoretical understanding and testing them using statistical methods. Statistical models are typically used for description of relationships of variables when analysing a collected dataset. Sometimes the causal mechanisms are too complex, or some of the causal variables are not available, making physical modelling impossible. In such cases statistical modelling with existing empirical variables is the only option available for the modeller. More complex models also considered statistical include neural networks.

It is both strength and a weakness of empirical models that they do not require nor imply any causal relationships between the model variables. An empirical model carries on to the result all the interdependencies existing in the data, regardless of whether they are causal or introduced by chance, or considered by the modeller. Empirical models require both input and output variables to be known in the model development system. Because the outputs must be measured anyway, empirical models are not at their best in estimation of unmeasured parameters, excluding perhaps the special case of modelling missing values within a dataset or cases where time-series data is used for statistical forecasting in the same target system. In most cases their data-set dependency restricts their use for making future predictions.

Regression models. By far the most common form of statistical model is the classic regression model. In its simplest form, a regression model solves the constant β_0 and coefficients $\beta_1 \dots \beta_n$, by minimizing the model errors (residuals) for the dependent variable. A standard equation is used to describe the relationships of the independent variables, resulting in the major benefit of statistical modelling techniques that variables having incompatible units of measure can be used together, including continuous and classification variables. Classification variables are typically transformed to binary dummies for studying the effects of a given questionnaire category on the dependent variable. Advantages of regression models include the capability of estimating the coefficient of determination (r^2) as a measure of how large a fraction of the variation of the dependent variable can be explained with the independent variables, and statistical significance (p) as a measure of the statistical probability of the model relationship being caused by mere chance.

Multiple regression exposure models can include concentrations in many microenvironments, and dummy variables for parameters such as smoking, form of commuting, type of work, gas stove, air conditioning, and other appliances. The terms of the resulting model are specific to the data set from which they have been calculated, and there are no grounds other than expert judgement to assess their applicability to some other location, time, or population. Examples of regression models used in exposure analysis include models for carbon monoxide exposures in Athens (Georgoulis et al., 2002) and Milan (Bruinen de Bruin et al., 2004a), and models for $PM_{2.5}$ and NO_2 exposures in Helsinki (Kousa et al., 2002b;Koistinen et al., 2001;Rotko et al., 2001;Kousa et al., 2001b;Rotko et al., 2000a).

Factor analysis. Another commonly applied statistical modelling technique in exposure-related studies is factor analysis. Principal component analysis (PCA), most common type of factor analysis, has been successfully applied to apportion observed air pollutant concentrations to different emission sources, or source categories, in a number of studies (e.g. Koistinen et al., 2004;Vallius et al., 2003;Edwards et al., 2001). Alternative forms of factor analysis have also been applied to environmental concentration data, including positive matrix factorization (PMF) and multilinear engine (ME) (Hopke et al., 2003;Paatero and Hopke, 2003;Basunia et al., 2003;Yli-Tuomi et al., 2003a;Yli-Tuomi et al., 2003b).

Advantages of factor analysis in source apportionment include the fact that the actual emission profiles of the different sources need not to be known. The corresponding disadvantage is that the results are largely data-set specific, and there are difficulties in comparing factors obtained from the same dataset using different methods, or factors from different studies. However, factor analysis is currently the mainstream technique to identify emission sources from concentration and exposure observations.

3.3.4. Physical Models

Capability to build reliable physical models is the best proof that all aspects of a phenomenon are well understood. Physical models, based on actual quantified physical and causal relationships between variables, are therefore, by definition, better suited for making predictions for alternative future policies than statistical models. In his overview of exposure models, Ryan (1991) divided physical models into deterministic and probabilistic ones. Short comparison of these techniques below introduces the main reasons why probabilistic modelling was selected for the current work. Deterministic techniques have their specific strengths in some exposure domains, as will be discussed in more detail later on when looking at the dimensions along which exposure data are aggregated.

Deterministic models are calculated for selected individuals using input variables describing physical processes, physicochemical characteristics, and mass-balances specific to the target individuals, locations, and points in time. Deterministic models need intensive sets of data when applied to anything more than few individuals and relatively short periods.

Dispersion models are a common exposure-related application area for deterministic techniques. Dispersion models describe emissions and atmospheric boundary layer conditions for estimating outdoor air pollutant concentrations. Such models are used for retrospective analysis of air quality and scenario analyses for policy options evaluation. Compared to air quality monitoring networks, dispersion models have tremendously better spatial resolution, and in addition support detailed analysis of concentrations caused by various emission sources. (Kousa et al., 2001a;Kukkonen et al., 2001a;Kukkonen et al., 2001b;Kukkonen et al., 2001c)

A common technique to overcome some of the limitations set by available data is to use population averages instead specific values for some variables. Typical examples of this include the use of fixed infiltration value; for O_3 in the AirPEX model in the Netherlands (Freijer et al., 1998), for $PM_{2.5}$ and NO_2 in a GIS-based population exposure model EXPAND in Helsinki (Kousa et al., 2002a) and for H^+ and sulphate in the U.S. (Suh et al., 1993). The use of population averages of input parameters instead of actual values does not pose significant problems for estimating mean exposures, but when distributions are estimated, it always reduces the modelled variance and biases individual model outputs towards the corresponding mean. It specifically leads to underestimation of the highest levels. Sometimes in cases when all causal effects cannot be included, physical models may apply physical factors estimated statistically from representative data sets (Karppinen et al., 2004b;Suh et al., 1993). Use of such factors biases results towards the mean, too.

Probabilistic models apply laws of probability to overcome the limitations of unavailable deterministic data for specific individuals, and to still capture the exposure variability in a given population. This is achieved by using the limited available data for estimating probability distributions of the values in the population in question. The population exposures are then simulated using physical equations from input values randomly sampled by the computer from them. In terms of data needs and model complexity, probabilistic modelling is the most efficient technique for estimation of population exposure distributions.

Because the input data in the probabilistic models are drawn randomly from defined statistical distributions, results of individual iterations are essentially random. Combination of a large number of them provides an estimate for population distribution. Originally probabilistic techniques were adapted to exposure analysis to model population variability. However, during the 1990's the methods were taken into use also in analysing uncertainty (Burke et al., 2001;Cox, 1999;Hattis and Burmaster, 1994;Morgan and Henrion, 1990). The uncertainty in a model (or in an analysis) can be described using probability distributions similarly as in Bayesian techniques (Rovers et al., 2005;Wikle and Berliner, 2005;Kashiwagi, 2004;Gangnon and Clayton, 2004), and numerical computer simulation can be used to propagate them through the calculations. The resulting distributions do not represent variability in the values, but uncertainty in them. In this sense simulation of uncertainty is closely related to classic statistical methods for estimation of confidence intervals. Second-order simulations include variability and uncertainty components in the same model (Burke et al., 2001;Cullen and Frey, 1999;Frey and Rhodes, 1996).

During the last few decades several research groups have applied probabilistic modelling for population exposures. The earliest models in the 1980's targeted CO and VOC exposures, but since 1990's also particulate matter exposures have been modelled. Some of the works were mainly targeted on model validation (Law et al., 1997;Ott et al., 1988), others have been focusing on developing tools for policy evaluation (especially models by EPA, e.g. Burke et al., 2001). Yeh and Small (2002) applied probabilistic 1-microenvironment model as a research tool in their analysis of health effects associated with $PM_{2.5}$ exposures. The current work combines the aspects of model validation and development of a tool for policy evaluation. The latest $PM_{2.5}$ models developed in parallel to the current work are summarized shortly below.

U.S. EPA National Exposure Research Laboratory (NERL) developed one of the current models in parallel with the *EXPOLIS* study. The objectives set for this model, the Stochastic Human Exposure and Dose Simulation model (SHEDS) were defined as: (i) prediction of population distributions of daily PM exposures in an urban area; (ii) estimation of contribution of PM of ambient origin to total PM exposure; (iii) determination of factors influencing personal exposures to PM; and (iv) identifying factors contributing to uncertainty in the model predictions (Burke et al., 2001).

SHEDS was applied for daily $PM_{2.5}$ exposures in Philadelphia (PA, USA) by Burke *et al.* (2001). Residential indoor concentrations were modelled based on a single-compartment mass-balance equation. Residential indoor emissions were modelled for cooking, smoking, and "other sources". For the other microenvironments (vehicle, office, school, store, restaurant, bar, other indoor) the distributions of PM concentrations were determined using linear regression equations from concurrent indoor and outdoor measurement data. Target population was divided into twelve groups by age and gender. Simulation results were presented, besides for total $PM_{2.5}$ exposures (mean ± SD: 30 ± 32 μgm^{-3}), separately for partial exposures in different microenvironments, and for exposures of ambient origin. The dominating role of residential indoor environment was obvious due to the large fraction of time spent there. Burke *et al.* compared their model outputs for Philadelphia with measurement results from Toronto, Canada, (Pellizzari et al., 1999) and Basle, Switzerland, (Oglesby et al., 2000); mean population exposures were 30, 28 and 24 μgm^{-3} , respectively. Levels excluding exposures to ETS in Philadelphia and Basle were 20 and 18 μgm^{-3}.

Yeh and Small (Yeh and Small, 2002) simulated population exposures to $PM_{2.5}$ and PM_{10} as part of their work where they compared ambient monitoring epidemiology (AME) approach to individual exposure simulation (IES) model in predicting the number of annual excess deaths caused by PM exposures in Los Angeles county (CA, USA). Same toxicity was assumed for all particles. The probabilistic IES model uses microenvironment approach with two microenvironments combined with mass-balance equation estimation of indoor concentrations caused by mixing of ambient air and emissions from indoor sources (smoking, cooking, other) and additional personal cloud concentration. The mass-balance equation parameters were estimated using data from two household databases (Murray, 1997;Murray and Burmaster, 1995) and the PTEAM study in Riverside (CA, USA) (Özkaynak et al., 1996), but now only residential indoor microenvironments were modelled. Simulated personal exposures were attributed to sources, but not compared to exposure measurements. The estimated number of annual premature deaths was slightly (5 and 10% for $PM_{2.5}$ and PM_{10}, respectively) smaller for the IES model compared to the AME model.

Ott et al. (1988) and Law et al. (1997) used a large population-based CO dataset from Denver, U.S., collected in the early 1980's to simulate population exposures using SHAPE and pNEM models, respectively. These modelling exercises are examined in more detail in **III** and in the section discussing model validation later on in this chapter.

3.3.5. Exposure Dimensions: Individuals in Space and Time

Personal exposures to fine particles vary in time, sometimes even second by second. Each subject is located differently and is in motion in the environment throughout the day, week, and year. As a theoretical mind game, the complete description of population exposure for a given time period, say a year, may be defined as consisting of instantaneous exposures second by second for each individual in the population throughout the year. Such a data structure is impossible to be obtained using current exposure measurement techniques and even modelling of it meets insurmountable problems, if not computationally, then at least in obtaining the necessary data. These difficulties will hardly be overcome.

Therefore to be able to estimate exposures and to draw meaningful conclusions on them, aggregation methods must be used to reduce this imaginary data set into a meaningful one that can be collected and used in exposure analysis. Common aggregation techniques include averaging and description of variability using various kinds of distributions. In the simplest and most common form, variability can be described using mean and standard deviation or other corresponding parameters.

Aggregation of the data occurs along the dimensions of the exposure data – individuals, locations, and time. In the aggregated end of the scale is the long-term mean exposure of the whole population, a significant health measure by its own (e.g. Pope et al., 2002). Each of these dimensions and techniques for handling them in modelling are discussed below.

Individuals and populations. Epidemiological studies have shown that a remarkable number of deaths are associated with fine particle exposures. Therefore estimation of the overall population exposure is one of the main interests. On the other hand, more detailed exposure analysis requires focusing on smaller groups (e.g. exposure studies), or even on few individuals.

Exposures of large populations can be estimated by drawing representative random samples. Standard statistical laws can then be utilized to estimate the uncertainty about the underlying true population values caused by the random sampling process. This method is commonly applied in the population-based exposure studies (see references in the section about population exposure studies earlier in this chapter).

Probabilistic modelling has become a standard technique adapted for modelling of variability of personal exposures in populations (Yeh and Small, 2002;Burke et al., 2001;Lunchick, 2001;Mitchell and Campbell, 2001;Hunter Youngren et al., 2001;Hamey, 2001;Mekel and Fehr, 2001;Price et al., 2001;Cullen and Frey, 1999;Law et al., 1997;Taylor, 1993, I, II, III, V). These population distributions could in principle be estimated using deterministic models for a statistically adequate number of randomly drawn individuals. However, in the population-based exposure studies this has been rarely done (or not reported in the literature).

Depending on the modelling approach, large target populations are usually divided into groups, or cohorts, that are handled separately within the model. Examples of such groups are age cohorts, men and women, and geographic, socioeconomic, and occupational groups. Whenever the exposures of different population groups are expected to be different from each other, their exposures probably need to be modelled separately. Recent studies have reported findings of heterogeneity in the toxicity of particles from different sources and in the sensitivity of different population groups (e.g. Samoli et al., 2005). Especially the elderly, patients with some medical conditions (including respiratory diseases, cardiovascular diseases, and diabetes) and infants have been suspected for higher sensitivity. While toxicologists and epidemiologists are trying to identify the most toxic particles and the most sensitive population groups, modellers are developing methods to estimate specifically the exposures of the susceptible individuals to the most toxic particles.

On the individual side detailed deterministic models have been developed to model personal exposures of small numbers of specified subjects in a limited time frame (Gulliver and Briggs, 2004;Briggs et al., 2003). A historical solution adapted into use in occupational hygiene to account for variability of exposures among the target population included definitions of hypothetical individuals, like the theoretical maximally exposed person. The exposure of this hypothetical individual is calculated (=modelled) by setting all variables to their worst possible values. Exposure estimates calculated this way are higher than the highest exposure of any true person in the target population. Practice has shown that such approach may, indeed, produce exposure estimates that are orders of magnitude higher than any of the actual exposures. The calculation of conservative point estimates provides no information on the actual level of conservatism in the estimate; therefore the development has shifted towards probabilistic assessments in the occupational settings, too. Probabilistic assessment is used to describe the exposure variability, including the prevalence of the highest levels, as accurately as possible, including quantitative estimates for model uncertainty when needed.

Locations. Highly variable environmental pollution fields and mobility of individuals make the spatial dimension utterly important for exposure analysis. The pollutant concentrations can vary rapidly outdoors in space and time due to changes in emission sources and meteorology, but often an even more significant modifier of exposures is the fact that a majority of time in developed urban areas is spent in indoors (Wilson et al., 2000;Wallace, 1996). Outdoor particles penetrate indoors with rather high efficiency along the air intake, but the gradual air exchange makes the concentrations indoors lag behind the outdoor ones smoothing out some of the variability. Indoors the particles are removed from the air by settling on surfaces and other processes, resulting in lower levels of particles. On the other hand, other particles may be generated indoors by resuspension and emissions from especially smoking and cooking, but sometimes also other sources, and by chemical reactions (Wallace, 1996, see also IV). As an outcome the indoor environment is a significant modifier of personal exposures to particles.

Two different approaches have been developed to handle the variation of concentrations in space: spatial techniques and the microenvironment approach. Spatial techniques preserve the actual geographical locations, where the exposures occur. The common computer technique to do this is to use geographical information systems (GIS). Most air pollution dispersion models produce concentration estimates for geographical outdoor locations (Karppinen et al., 2004a;Kousa et al., 2001a;Kukkonen et al., 2001a). Detailed models of indoor air quality have also been developed, but have not been combined with larger scale models of urban air quality mainly due to the difficulties in obtaining the needed detailed data on air exchange systems in individual buildings.

Most detailed spatial modelling follows specific individuals in space and time, modelling the concentrations for the exact locations and times where the individuals are. An example of such approach is the work conducted in the Imperial College, London (Gulliver and Briggs, 2004;Briggs et al., 2003). Moving towards population level makes it impossible to follow all the individuals in space and time. Jensen *et al*. have developed techniques utilising administrative databases to model locations of population members and combine these with air pollutant concentrations from a dispersion model (Hertel et al., 2001;Jensen, 1998).

In Helsinki a statistical approach to population locations has been adapted and combined with dispersion models (Kousa et al., 2002a). Locations of residences and workplaces are retrieved from public databases and an hourly statistical population time-activity model is used to allocate the population members to the residences, workplaces (both as employees and as customers), and to traffic. Results are displayed over the whole metropolitan area using a 100 m x 100 m grid. Infiltration of pollution indoors is modelled using a population average value observed in the *EXPOLIS* study. Population members are not followed across the hours and therefore daily personal exposures cannot be estimated.

The alternative approach, commonly used in probabilistic modelling and selected for the current work, is the microenvironment approach, which classifies different locations visited by the subject into so-called microenvironments (one of the early references Fugas, 1975). The concentration field within the microenvironment is described in this approach using an average value. This is often stated in the literature as assuming the concentration field to be constant within the microenvironment, but this, of course, does not need to be true. Exposure is then calculated as the time-weighted average concentration level across the microenvironments visited (Burke et al., 2001;Freijer et al., 1998;Ryan et al., 1986;Letz et al., 1984;Dockery and Spengler, 1981, II, III, V).

The microenvironment concept has been developed for two different purposes. The first is the fact that the exposure levels of many pollutants are often more similar in e.g. two similar residences or two similar offices across the city than inside and outside of the same building. In other words, the microenvironment category may be equally or more important than the geographical location. The microenvironment concept simplifies exposure modelling dramatically when combined with probabilistic techniques by reducing the millions of actual locations into a limited number of categorised microenvironments.

The probabilistic approach assumes that the concentrations of all outdoor or indoor locations grouped together into a microenvironment can be described by the same probability distribution. The concentrations for simulated microenvironments are then sampled from the defined distributions using computer and random number generator. Exposure contributions of each microenvironment are calculated according to time activity model (e.g. original version of SHAPE, Ott et al., 1988, II, III, V) or using measured time activity patterns (SHEDS using CHAD+HAPEM, Burke et al., 2001;AirPEX, Freijer et al., 1998;pNEM/CO, Law et al., 1997).

The microenvironment approach simplifies spatial modelling substantially. Deterministic time activity model for a large target population would require the geographical locations of each subject to be recorded. Global positioning system (GPS) devices, or the even more up to date GSM based positioning techniques that function also indoors would technically allow for registering such data. The computational requirements, however, are also much reduced when geographical and indoor locations can be combined into a small number of microenvironments.

Time. Temporal scales affect exposure assessments in two ways. Any exposure data are related to some temporal time frame. Emissions, meteorology, populations, activities, and many other environmental factors all change in time, and thus any data on exposures will definitely change too. The relationship of exposure data to the time dimension is often implicit; the limitations are not clearly stated, nor are they always even known. An example of these kind of unknown limitations could be the measured personal and population exposures in the *EXPOLIS* study. The exposure measurements were carried out in 1996-2000, and probably describe the exposures in the seven European cities for some years before and after the measurements. But for how long? Limitations may be specific to a given city, or to a sub population within a specific city, and can only be judged by expert opinion. Depending on the study or model design, exposure data may be representative of a specific time of a year (e.g. summer), days of week (e.g. work days in the *EXPOLIS* study) or time of day.

Another equally important temporal aspect is the averaging time of exposures (Ryan, 1991). Biological doses are functions of uptake and removal processes and therefore the health effects depend on the temporal variation of the exposures. Same integrated personal exposure to CO that as a short-term peak would be lethal is harmless as constant annual level. Similar results have been observed for fine particles; the relative risk for additional mortality associated with daily concentration variations (i.e. short-term exposures) has been estimated to be around 1.5%, while relative risks up to and above 15% have been suggested for long-terms exposures (WHO, 2000). In the case of short-term exposures, epidemiologists find also different lags from the exposure to the health effects (Samoli et al., 2005;Katsouyanni et al., 2001;Penttinen et al., 2001;Roemer et al., 1998;Pekkanen et al., 1997).

Best compilation of current knowledge about health-relevant exposure averaging times are reflected in the definition of air quality guidelines (e.g. WHO, 2000). Several averaging times are needed to protect the public from health effects caused by some pollutants; for others a single time value - with varying averaging times - is considered to provide adequate protection. The relationship of health effects and the temporal exposure profiles is still poorly known. Short-term peak exposure values cannot be assessed from long-term average concentrations, nor can long-term averages be estimated from short samples, if the intrapersonal variation in exposure levels in not known. Therefore the selection of the relevant averaging time must be done properly when designing the model and obtaining the corresponding input data.

3.3.6. Conceptual Model and It's Implementation

The issues of aims of modelling, types of causal relationships and corresponding modelling types, discussed above, affect the development of a conceptual model, which defines the phenomena included in the model, selection of the dependent and independent events, spatial and temporal scales, and equations describing the modelled relationships (Law and Kelton, 1991). Before the model can actually be used, it has to be transformed into definitions of variables, formulas, and a logical flow of computations (Figure 3): the model has to be implemented. The conceptual model looks at the principles, but the implementation has to take care of all the details.

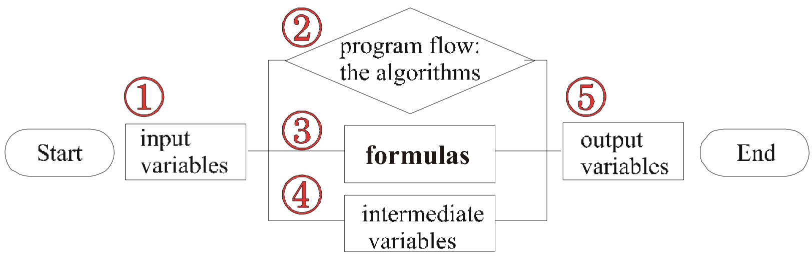

Figure 3. Five components of a computer model: 1) input variables, 2) algorithms, 3) formulas , 4) intermediate variables, and 5) output variables.

A neat conceptual model takes a lot of effort and dirty details before it has been turned into a reliable piece of computer software. The details that must be taken care include behaviour in the case of missing data and other special conditions, often created by such an unexpected source as the laws of mathematics. Besides driving the technical aspects of model reliability, implementation directly creates the user interface, consisting of methods for entering input variables, selecting model options, running the model, and retrieving the results. Model implementation has a significant effect to the required type of model documentation. In the case of a clear implementation, the model documentation needs mainly to concern about introducing the conceptual model. However, often the technical complexities in the model implementation totally drive the type of instructions required to use the model. Good model documentation should always first describe the conceptual model with its underlying assumptions and limitations clearly – good and intuitive implementation should then minimize the need for technical details.

3.3.7. Model Validation

The outset of the current work was the insight that exposure modelling is an important and necessary tool for science-based development of environmental policies. Environmental policies should pose as little limitations and costs to the society as possible while ensuring safe environment for all. But what if a model used in the development of such policies would be unreliable? All conclusions based on such a model would be dubious at best, and total garbage at worst. A model is useful only, when the limits of its applicability and its accuracy are known.

On the other hand, Oreskes *et al*. (1994) shoot calmly down any attempts to 'validate' any model that describes one part of an open system for good. Environmental exposure definitely takes place in such a system. Models work, at best, as long as the rules of the system do not change. As an example we can think of the Newton's law of gravity (Newton, 1687), thought to be the greatest of all laws in the Nature and newer to change, before Einstein was able to see beyond the its limits of applicability (Einstein, 1916;Einstein, 1905). Of course, the limits of the applicability of gravity law in an open-ended system can easily be demonstrated also in everyday surroundings by introducing e.g. resistance of air to the system under study. One law (or model) applies only until one overlooked starts influencing the system. Nevertheless, the need for model 'validation' is as clear as is the impossibility of the task. This contradiction should not lead to confusion, as discussed in more detail in **V**. There is a real need to quantify model reliability, and several techniques available, including modelling of uncertainty and analysis of model errors (**V**).

Building a valid model starts from a credible conceptual model (Law and Kelton, 1991). A model should include all phenomena that can be expected to be significant in the target system. The conceptual model should then be transformed to a mathematical form and often implemented in a computer environment without introducing errors.

Models describe how the changes in the input variables are reflected into the outputs. Model applications are linked to a larger picture, to human understanding about how the phenomena of interest affect the model inputs, and how others are affected by the model outputs (Figure 4). Models are useless in assessing events that are not related to the model inputs or outputs. Thus the model input and output variables define the main domain of the model.

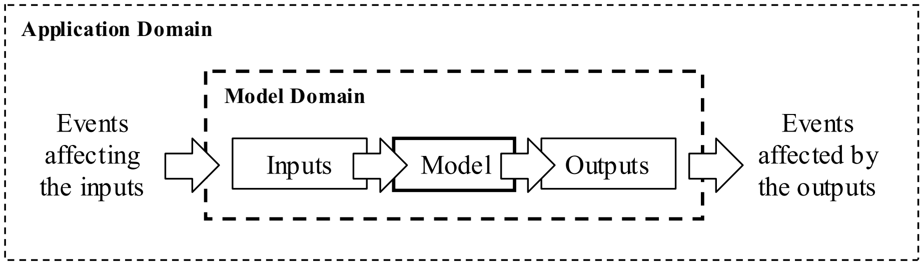

Figure 4. A model quantifies the relationships of input and output variables. Model is applied using understanding of these variables and the rest of the world that is related to them.

Law and Kelton (1991, p. 299) define model validation as determining whether the conceptual model is an accurate representation of the system under study. Model implementation transforms equations of the conceptual model into formulas that specify how the values of input variables are used to calculate intermediate and output variables. Algorithms define the computational sequences in which these calculations are performed. Comparison of the conceptual and implemented model is called by Law and Kelton 'model verification'. Implementation of even the simplest conceptual model adds another layer of complexity to the system, because valid equations produce nonsense results, if not applied in a proper sequence, or the formulas do not handle missing and out-of-range values properly.

When the model is ready, its outputs can be compared to observed values in a known system to further confirm the model (Oreskes et al., 1994). Accuracy in prediction can be tested only in a selected, existing target system. Leijnse and Hassanizaded (1994) called comparison of model predictions to observations 'strong validation'. They point out that even a conceptually bad model might by change seem to work well in a limited set of test data. Therefore, final trust or distrust on model applicability on a given problem must be based solely on our belief that our question concerns a target system similar to what the 'validated' model describes.

Two earlier works have been published on validation of probabilistic population exposure models (Law et al., 1997;Ott et al., 1988). Both of these are based on personal CO data collected in Denver, CO, in winter 1982-83. Microenvironment concentrations were estimated from the personal time-series data using time-activity diaries. The main result from both models was that the overall level of population exposures was captured well, but the variability was underestimated for reasons discussed in more detail in **III**.

4. MATERIAL AND METHODS

Model development requires both theoretical background and input data, as depicted in Figure 5. A model may be developed based on theory with literature and expert judgement for model inputs, but such an approach leaves open the uncertainties concerning the model validity and reliability. More detailed model evaluation requires data also on model output variables, exposures in the current case.

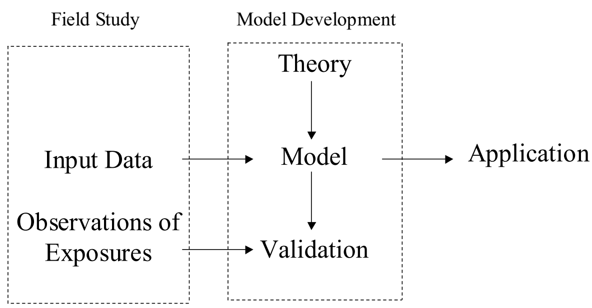

Figure 5. Relationships of the main elements of the current work.

The field data for the current work was collected in the *EXPOLIS* study (Air Pollution Exposure Distributions within Adult Urban Populations in Europe) conducted in Helsinki in 1996-97 and in six other European cities in 1997-2000. The following describes the study features relevant for the simulation of population exposures to $PM_{2.5}$ in Helsinki. The main design features of *EXPOLIS* are described in detail the original article **I** and compared to earlier and parallel PM studies in Table 1. The main study objectives were:

1 Assessment of exposures of European populations to major air pollutants
2 Analysis of personal and environmental determinants of these exposures
3 Development of a European database for simulation of air pollution exposures

The results for objective number one, the actual measured personal exposures, are used in the current work in the validation of the modelling results by comparing simulated and observed exposure distributions. Objective number two covers the measurements of microenvironment concentrations, time-activities, and personal exposure-related characteristics of the subjects that are used as model inputs. These inputs were accessed using the exposure database created according to the objective number three.

4.1. Designing the Field Study for Collecting Modelling Data (I)

The modelling approach developed as one of the main goals of the *EXPOLIS* study is not specific to $PM_{2.5}$ or Helsinki. The field study included other pollutants and cities, as shortly described below, and the modelling framework can quite well be utilized to modelling of other pollutants as well, as demonstrated by e.g. Kruize *et al*. (2003) and Bruinen de Bruin *et al*. (2004b). The current work is focused on $PM_{2.5}$ exposures in Helsinki to set a reasonable scope for a doctoral dissertation.

4.1.1. Multi-pollutant approach

While recent air pollution health studies has mainly focused on particulate matter, other pollutants have also been associated with various health effects, including mortality and morbidity, or have been shown to be irritating or carcinogenic. Many exposure-related factors correlate, causing subjects to be exposed to elevated levels of several air pollutants at the same time. Therefore it is important to be able to assess the exposures to multiple pollutants.

Exposure measurements are intruding and demanding for the subjects, including carrying the monitoring equipment with them for the study period and filling in lengthy questionnaires, taking their time and attention. In case of microenvironment measurements the subjects have to let the researcher in their homes and workplaces, installing noise-making and space-reserving monitoring devices. The subjects have to provide personal information regarding their social and occupational status, time-activities, and personal habits.

In a population-based approach a random sample of subjects must be drawn and recruited to the study. A significant load of resources are needed in the visits to the subjects' residences and workplaces, installing the monitoring equipment and instructing the subject. Several monitors can be easily installed during a single visit, and exposure related questionnaire data can be used to assess determinants of many exposures. Therefore maximal utility of the resources can be achieved by combining measuring several air pollutants together. For these reasons, the main air pollutants included in the *EXPOLIS* study were $PM_{2.5}$, a selected set of 30 VOCs, and CO. In some centres, including Helsinki, NO_2 exposures and concentrations were included with a separate funding. Additionally for a subset of the collected samples, the elemental composition of the $PM_{2.5}$ samples was analyzed separately.

4.1.2. Multi-centre study

The population-based urban, working age, multi-pollutant inhalation exposure research effort that became the *EXPOLIS* study, was originally designed as a national project for the Helsinki Metropolitan area in 1994. As such, the study was to be expensive due to the labour intensive nature of exposure measurements and other tasks, including the development of the measurement methods and quality assurance procedures, training of the research personnel, recruiting procedures for the population samples, and the computer software and databases to store and manage the collected data. Therefore the Academy of Finland redirected us to international sources. At that time Finland was just about to be a new member of the European Union, and two years after the study planning had started, the EU Directorate General (DG) for Research granted the funding for the study as part of the fourth framework program for European research. Before that, the research plan was transformed to a multi-centre approach and reviewed by European and American scientists including M. Lebowitz, B. Seifert, W. Ott, D. Mage, W. Wilson, J. Spengler, B. Leaderer, and D. Moschandreas.

Personal exposure and microenvironment monitoring techniques were not in wide use in Europe in the mid 1990's. Reliable simultaneous measurement of multiple pollutant concentrations in varying field conditions, including indoor and outdoor locations and private and semi-public places like workplaces and offices set high demands on the reliability, robustness, repeatability and user friendliness of the measurement methods. Moreover, handling of diverse sets of personal population based questionnaire data and data from physical measurements involving airflows, sample weights, temperatures, air pressures, and sample and equipment identification codes, required extensive data management procedures to maintain integrity and reliability of the collected data. Besides the development of study protocols, a written documentation and a training program for the personnel were needed to optimize data quality. All these tasks take the same effort whether done for only a single centre or for many. Air quality in Finland is known to be clean in comparison to many central European locations. All these points made it reasonable to extend the study from Helsinki to several other cities within Europe.

The multi-centre approach made the training events international and required a careful cross-translation program for the questionnaires and support materials that needed to be in the national language in each centre. The researcher-training program was conducted by having international workshops in Prague (April, 1996), Helsinki (September, 1996), Grenoble

(March, 1997), Bilthoven (February, 1998), Paris (May, 1999), and Bern (November, 1999). The study materials were developed in English, which was not the national language for any of the original partners.

4.1.3. Expedition for Exposure Determinants

The field study was designed to produce a large database on exposures and exposure related characteristics – potential exposure determinants. Detailed analysis of the potential exposure determinants was conducted. Many statistically significant determinants were found, but only factors with significant influence on exposures needed to enter the exposure model. Therefore from the point of view of the current work, the exposure determinant analyses of the collected data (Koistinen et al., 2004;Götschi et al., 2002;Kousa et al., 2002b;Koistinen et al., 2001;Rotko et al., 2000a) formed the basis for the exposure model structure, including the selection of microenvironments and population groups.

Targeting many air pollutants with different – but unknown – determinants, and multiple centres had many implications on development of the questionnaires. A good example is the use of double gazing in apartments: in Helsinki double glazing is the minimum requirement and is giving room for triple glazing, while on the other hand in Athens it stands for advanced insulation. This kind of research setup is different from the traditional experimental research, where a hypothesis is created before designing the experiment for testing the hypothesis. Therefore this type of exposure studies can be called "fishing expeditions" – large sets of presumably related data are collected with stated ideas about the needs and future use, but without definition of specific hypotheses to be tested. The actual statistical methods and variables used in the analysis of the data were to be selected later. Therefore the focus in the study design is in selection of a wide range of variables that are both (i) measurable and (ii) have causal or interesting statistical connections with the exposures. Such variables include naturally time-activities and microenvironment concentrations, but also variables related to personal, residential and occupational characteristics, including socio-economic status, smoking habits, exposure to ETS, etc.

4.1.4. Population sampling in Helsinki

Random population sample is not an absolute requirement for a model development study. The model could be developed using data from a selected group of volunteers. However, using a random population sample makes the results on model inputs and outputs representative of the general population from which the sample was drawn. Therefore the random population sample approach significantly increases the generalisability of the results.

In Helsinki the *EXPOLIS* sample was drawn by the Finnish Population Register Centre and consisted of 2523 Finnish speaking citizens of the Helsinki metropolitan area (including cities of Helsinki, Espoo, Kauniainen, and Vantaa) born in 1940-1970, inclusive (Table 2).The Helsinki population sample data was first received from the Civil Register on May 14th, 1996, with a correction file on May 21st, 1996.

Table 2. The random sample of the Helsinki working age population.

City	Male	Female	Total	%
Espoo	264	270	534	21.2
Helsinki	683	781	1464	58.0
Kauniainen	13	11	24	1.0
Vantaa	231	270	501	19.9
Total	1191	1332	2523	100.0
	47.2 %	52.8 %	100.0 %	

A mailed questionnaire was sent to this random population sample. After a mailed remainder a final attempt to reach the non-respondents was done using a telephone interview, resulting a final response rate of 74 % (Table 3).

Table 3. Response rate to the mailed questionnaire and telephone interview.

Questionnaire response rate		%
Questionnaires sent	2523	100 %
No response	650	26 %
Response, total	1873	74 %

One of the questions regarded the subject's willingness to participate in the whole study, including exposure and microenvironment measurements, or questionnaires only (Table 4).

Table 4. Responses to the willingness to participate in the study.

Willingness to participate the study		%
Respondents	1873	100.0
No answer to this question	32	1.7
Yes	1368	73.0
Yes, questionnaires only	56	3.0
No, travelling	376	20.1
No	41	2.2

Two sub-samples were randomly drawn from the respondents willing to participate in the exposure measurements or the questionnaires-only study. A running selection code was allocated to these subjects randomly to ensure random sampling over the one-year study period, integrating over seasonal variations. Computer forms supporting telephone contacts to the subjects during the study field phase were created into the local *EXPOLIS* Access Database (EADB) used in each of the study centres.

At a later stage, eleven participants of the simultaneous ULTRA-study (Vallius et al., 2003;Pekkanen et al., 2002;Ruuskanen et al., 2001) were recruited for the *EXPOLIS* exposure measurements. These subjects, being patients with cardiovascular diseases, lived in the Vallila-Kallio –area (zip codes 00500, 00510, 00520, 00530, 00550 and 00610) a few kilometres from the Helsinki downtown. Table 5 lists the relative effect of these additional subjects on the collected data used in the analyses.

Table 5. Sizes of the Exposure and Diary sub samples.

Data set	Random sample	%	Ultra subjects	%[2]	Total[1] subjects
Civil register data	2523	100.0	-	-	-
Short questionnaire	1871	74.2	11	0.6	1882
Questionnaires and diaries	423	16.8	11	2.5	434
Exposure measurements	190	7.5	11	5.5	201

[1] Data from these subjects have been used in the analysis. [2] Percentage from the total subjects

EXPOLIS subjects were aged 25-55 years at the time of sample formation. This population group forms a significant fraction of the total population, is legally and physically capable of participating in this kind of study. The selected age category includes working and non-working subjects with a large variety of leisure time activities and thus their time-activity is variable. Therefore, in terms of exposure characterization, this group is more heterogeneous than the susceptible sub populations like infants and elderly, which spend more of their time in and around their residences. Especially time spent in traffic, one of the most important exposure modifiers of the active population, has presumably much smaller effect on the exposures of the susceptible groups.

In Finland only Finnish speaking population was included in the sampling to avoid error prone and time-consuming translations of the questionnaires and other written support materials. The biggest non-Finnish speaking minority in the Helsinki metropolitan area consists of Swedish speaking Finnish citizens (9.3%, 6.5% and 3.5% in Espoo, Helsinki and Vantaa, respectively, in 2000-2002[1]). The fraction of other language minorities has been increasing constantly, being approximately 5% in Espoo and Vantaa, and almost 6% in Helsinki at the same time. Therefore the total percentage of minorities excluded from the study by the language limitation is approximately 11-12 %. Although there is no specific reason to assume that the time-activities, living or working areas, or other exposure modifiers of the language minority groups would be significantly different from the Finnish speaking majority, this limitation should be kept in mind when interpreting the results.

The population sampling and sample quality are described in detail in Rotko *et al.* (2000b). Rotko *et al*. found that the biggest loss of representativity occurred in the first contact phase, answering the short questionnaire. In general, women and individuals with higher education were overrepresented in the exposure and diary samples, and men, younger subjects (defined as 25-34 years) and unmarried individuals were somewhat underrepresented. In comparison to the other *EXPOLIS* cities the Helsinki response rates were good. From the model development point of view the population sampling can be considered successful and the results from the modelling representative of the general working age population in Helsinki metropolitan area.

---

[1] Tilastokatsaus 2003:5. Vantaan kaupunki, B6, ISSN 0786-7832. (In Finnish)

4.2. Time-Activity Measurements

One of the main exposure modifiers is the mobility of subjects. People spend their time in various types of environments in different locations within the metropolitan area. Time-activity measurements were conducted using a structured 15-minute resolution diary with eleven microenvironments and three activities. The microenvironments were grouped into transportation (five categories) and stationary microenvironments (residence, workplace and other, each subdivided into indoors and outdoors). The subjects classified their locations into these categories for approximately 48 hours, the same period when their microenvironment concentrations and personal exposures were monitored.

The diaries were entered into EADB and transformed into fractions of time using the duration of the subject's diary. Time fractions for the elementary diary microenvironments were further combined to create aggregate microenvironments for the simulation models (Table 6).

Table 6. Microenvironment categories used in the simulations.

	Number of microenvironments in the model			
μE	2	3	4	5
1	Residence	Residence	Residence	Residence
2	Workplace	Workplace	Workplace	Workplace
3		Other	Traffic	Traffic
4			Other	Other indoors
5				Other outdoors

The three activities were smoking, exposure to ETS, and cooking. The two tobacco activities were combined to ETS-exposure yes/no indicator also for active smokers, because only exposure to ETS was sampled. Cooking was recorded without more detailed specification of the type of cooking (e.g. boiling water versus frying or toasting). Moreover, the effects of cooking were diluted into the 48-hour sampling period and therefore cooking was found not to have a notable effect on concentrations and was not included in the exposure models.

4.3. $PM_{2.5}$ Measurements

In Helsinki a full set of personal workday and leisure time exposures, and residential indoor, outdoor and workplace concentrations were successfully obtained from 194 subjects (total number of exposure measurement participants was 201 with 7 subjects with various failures). The number of non-ETS exposed subjects was 126. The $PM_{2.5}$ measurement techniques and quality assurance results are described in Koistinen *et al.* (1999)and Hänninen *et al.* (2002b) and primary analysis of the data in (Koistinen et al., 2004;Götschi et al., 2002;Kousa et al., 2002b;Koistinen et al., 2001;Rotko et al., 2000a).

The measurements were carried out in a random sequence during an approximately 12-month field survey period (three final subjects were measured after a 2-month pause at 14 months from the beginning of the field phase). Each subject was measured for two consecutive working days, from Monday to Wednesday or from Wednesday to Friday. National holidays were excluded and during the holiday seasons only subjects not on vacation were measured.

Residential indoor and outdoor air was sampled from evening to morning, approximately at times when the study subject was expected to be at home according to the subject interview. The workplace air was sampled during the normal working hours. Personal samples were taken on two filters; one was taken into use in the morning, just before the subject left home or started the daily activities at home. Second filter was changed to when the subject returned home in the afternoon. Thus filter one corresponds to the daytime exposures, including workday and commuting, and filter two to leisure time (including night) exposures. The elemental composition of the filters of 98 subjects was analyzed using Energy Dispersive X-ray Fluorescence technique (ED-XRF) in the University of Basle (Mathys et al., 2001). Sulphur data was used in the current work to apportion indoor concentrations into ambient and indoor generated fractions (**IV, V, VI**).

4.4. Data Management

Data management for this work was integrated with the data management of the whole *EXPOLIS* study. This included managing data for over 300 measured compounds (i.e. selected target VOC compounds (30) and other compounds observed (290), and elemental composition of the PM samples (37 elements)), questionnaires, and time-activity diaries. Only $PM_{2.5}$ and sulphur data, time-activity diaries, and some questionnaire variables were used in the current work (**II** – **VI**), but the database was designed to support corresponding simulation of exposures to any of the measured pollutants (e.g. Bruinen de Bruin et al., 2004b).

The original objectives underestimated the role of the exposure database by putting it into being merely an aid for the simulation. As later summarised in the article **VII**, the combined international database (CIDB) turned out to be a major outcome of the project by itself. The database has been used for data analyses producing over thirty original articles with only few relevant ones for exposure simulation, and besides the current work, over ten doctoral dissertations in seven countries, involving nine universities and four other research institutions have been based on the data.

EXPOLIS data management goals were specified as: (i) all data items affecting the final calculated results are stored, (ii) data from all centres are stored, (iii) data storage structure is flexible, allowing later any analyses necessary, (iv) correctness of the data is maximized, (v) data entry tools and procedures are provided, and (vi) privacy of study subjects is protected. The data management procedures were developed as the second phase of the current work in integration and partly overlapping with the first one, the field phase.

Database design. A project database (*EXPOLIS* Access Database, EADB) was developed using Microsoft (Seattle, WA) Access 7.0 (a.k.a. version 95). Relational database model was selected to allow maximum flexibility. Microsoft Access with a powerful, visual, and user-friendly environment, low software cost, and easy availability as part of the most abundant office software package was selected as the platform. The database used in the European CESAR project served as a model in designing the *EXPOLIS* approach (Fletcher et al., 1999)

A local database was created for each centre. The local database consisted of several Access database files, containing data from local Civil Register and other national registers, *EXPOLIS* time activity diaries, questionnaires, monitors, laboratory analyses, calibration procedures and environmental conditions as well as urban air quality network and meteorological data covering the field study periods. All data was stored in its primary form and calculations were performed using the primary data dynamically.

The local data was grouped to be stored in separate database files. Population sample management, questionnaire data, and concentration sampling were stored into the local main database. Time-activity diaries were stored in a 15-minute resolution time series database, CO data in 1-minute resolution time series database, meteorological data in 3-hour resolution time series database, and ambient air quality fixed station data in one-hour database. Averages of environmental variables from the meteorological and fixed station databases were calculated into the *Fixedruns* database for the microenvironment and personal sampling periods.

Table 7. Local database files in Helsinki. Corresponding files were used in all centres.

Data files	Tool file	Description
HELSINKI.MDB	EADBTOOL.MDB	Main local database: Questionnaires, exposures, concentrations etc.
TMAD15min.MDB	TMAD15minTOOL.MDB	Time-activity diaries (15-min resolution), 15-min avg. personal CO data
CO1min.MDB	CO1minTOOL.MDB	1-min CO exposures and TMAD data
FIXED.MDB	AmbientTOOL.MDB	Hourly ambient air quality data
MET.MDB		3-hourly meteorological data
FIXEDRUNS.MDB		sampling period averages of ambient and met data; all stations

The local database files were split into two functional groups. (i) Data files contain all data tables; (ii) the data processing tool elements, queries, forms and Visual basic modules, were stored in tool files (Table 7). The tool databases were then linked to the data files using Access Linked Table Manager, allowing for development and upgrading of the tools without changes to the data files in continuous use. Finally after the field phase and local data cleaning were completed in each centre, the local database files were collected and combined into the Combined International *EXPOLIS* Database (CIDB). The database structures are described in detail elsewhere (Hänninen et al., 2002a).

A data integrity protocol was established according to the data security requirements of EU Directive on Protection of Individuals with Regard to Processing Personal Data (Directive 95/46/EC). Persons were labelled using codes, and personification information (names, addresses) was removed after the field phase. The database files were secured with user identification and password control and the staff working with the databases in all centres were specifically trained in several common workshops.

4.5. Simulation Framework (II)

Eighteen simulation models are presented in the original papers (**II**, **III**, **V**, **VI**). All these models were implemented using the microenvironment-based simulation framework developed originally in collaboration with **RIVM** (National Institute for Public Health and the Environment, Bilthoven, NL) as part of the *EXPOLIS* study (**II**). The development of the modelling framework was one of the main objectives of the *EXPOLIS* study to support exposure assessments for alternative policy options. The models based on the framework were to allow for assessing population exposure distributions of (i) selected sub populations and (ii) urban areas for (iii) different future scenarios (**I**).

The framework uses similar microenvironment approach like independently developed models by e.g. Burke *et al.* (2001) and Yeh and Small (2002) to calculate time weighted average exposure levels (Ryan et al., 1986;Letz et al., 1984). The framework allows for definition of sub populations, macro- and microenvironments, indoor sources and time activities. Population time is allocated to macro- and microenvironments selected by the user and modelled as fractions of time using 2-parameter beta-distributions (**II**, **III**, **V**, **VI**).

Microenvironment concentrations can be modelled in *direct* or *nested* mode. In the *direct mode* the concentration distribution is assumed lognormal and the probability distribution parameters are directly entered as inputs (**II**, **III, V**). In the *nested mode* the concentration of ambient origin is modelled from an ambient concentration distribution using an infiltration factor distribution (**V**, **VI**). In both modes indoor sources can be defined for a given fraction of each microenvironment type. The additional indoor source concentrations are defined as 2-parameter lognormal distributions. (**II**, **V**, **VI**). The framework was implemented as Microsoft (Seattle, WA) Excel workbook using the @Risk add-on software (Palisade, Newfield, NY).

The population exposure distribution is then simulated by applying probabilistic sampling to each of the input distributions. The partial exposure in each microenvironment is calculated by multiplying the microenvironment concentration (C) by the fraction of time spent in that microenvironment (f). The exposure level $\overline{E}$ of each iterated population member is calculated as the sum of the partial exposures over all microenvironments in the model (**II**).

The use of fractions of time to describe population time-activities implies that the microenvironment model in this equation must be complete for the equation to produce average exposure level, i.e. that $\sum f_i = 1$. When this condition is met (or the result is scaled to unity time fraction by dividing it by $\sum f_i$), the equation is applicable for any averaging time and any number of microenvironments and can in principle be used for any air pollutant. Repeating the calculation for a large number of hypothetical population members estimates the exposure distribution for the target population. The number of iterations in simulation runs ranges typically from hundreds to thousands.

The development of the framework was described and models based on it were demonstrated in **II** using two examples. The first example used *direct* mode models to simulate the annual distribution of 48-hour $PM_{2.5}$ exposures in Athens, Basle, Helsinki, and Prague. ETS and other indoor source exposures were not separately modelled, but were included in the microenvironment concentration distributions as observed in the *EXPOLIS* study. The second example demonstrated the *nested* mode to model the distribution of daily PM_{10} exposures in the general Dutch population, including all age groups and both rural and urban areas, for current situation and an alternative scenario, where ETS exposures were set to zero.

A more detailed evaluation of the *direct* mode was conducted for $PM_{2.5}$ exposures in Helsinki in **III**. The required number of microenvironments was studied by starting with the simplest possible models that take into account the mobility of the population, i.e. models with two and three microenvironments. Because in this stage (and in **II**) it turned out, that ETS exposures are a significant modifier of the exposures, the more detailed models in **III** were run excluding these to see how well the non-ETS exposures can be captured by the model. Population time-activity was modelled separately for the working and non-working adults.

Analysis of residential infiltration factors and indoor source strengths was conducted in **IV**. These data were used as inputs in the main paper of the current work (**V**), where validation of the *nested* modelling approach was completed. This paper elaborated on the theoretical aspects of terminology in model validation and uncertainty analysis, and quantified model errors for $PM_{2.5}$ models in Helsinki. The model was enhanced by handling exposures in traffic as a separate fourth microenvironment; the exposure levels while in traffic were estimated using separate traffic measurements conducted during the *EXPOLIS* field phase.

Finally, the use of the developed and evaluated modelling tool was demonstrated in **VI** by estimating the risk reduction potential achievable by using modern ventilation systems. The current situation was described using the *EXPOLIS* measurement data and a subset of the data was utilized in creating the future scenario. Occupational buildings built in and after 1990 all use a mechanical ventilation system with fine particle filtration according to the Finnish Building Code. The infiltration factors analysed for these buildings were used in the alternative scenario for all buildings, assuming that the whole Helsinki building stock would have been renewed to the condition currently required for new buildings.

5. MODEL AND EVALUATION RESULTS

Model development and model evaluation were conducted in two major steps. In the first phase a *direct* microenvironment-approach was used, where the parameters of concentration distributions in all microenvironments are entered directly into the model as inputs. These concentration distributions represent the total measurable $PM_{2.5}$ concentration in the microenvironments, making no difference on origin of the particles. In the second phase another layer of modelling was added to allow for *nested* modelling of concentrations in indoor microenvironments by using ambient concentrations, infiltration factors and indoor sources as inputs. This approach required analysis of the infiltration factors and contributions of indoor generated particles to the indoor concentrations from observed total concentrations and corresponding elemental compositions.

The *direct*-mode results from the first phase proved that the microenvironment-based modelling approach and the simulation technique can be applied to 48-hour $PM_{2.5}$ exposures without any significant problems (**III**). Starting with the simplest approach with only two microenvironments and no sub population divisions, and working towards more detailed models when a need was indicated by the previous step, ETS exposures were identified as the most significant modifier of personal exposures. Further division of the target population into two groups according to the working status improved the time-activity modelling, but still turned out not to be a very significant modifier for $PM_{2.5}$ exposure modelling.

Infiltration factors and indoor source strengths were analysed for Helsinki and three other *EXPOLIS* cities (**IV**). Buildings in Helsinki were better sealed than in the other cities, leading to slightly lower infiltration factors. Concentrations caused by non-ETS indoor sources were comparable in all cities. Similar finding was made in U.S. using a statistical estimation technique for PM_{10} (Ott et al., 2000). The *nested* model, based on ambient concentrations, infiltration, and indoor sources, produced equal results to the *direct* model, indicating that the additional layer of modelling did not significantly deteriorate the modelling results. From the model applicability point of view, however, the ability to use ambient levels instead of microenvironment measurements is a significant advantage.

After model validation, the model was applied to a hypothetical, but data based exposure reduction scenario (**VI**). The buildings sampled in the *EXPOLIS* study were classified into two categories according to the construction year, dividing line drawn to 1990. Mechanical ventilation is more common in the newer buildings, and in the occupational buildings built after 1990 mechanical ventilation system with efficient fine particle filtration is standard. Therefore the infiltration factors estimated for these buildings were used to define the hypothetical scenario representing a future building stock where all buildings utilize controlled ventilation and fine particle filters. The validated simulation model is used to estimate the exposure reduction potential for such a scenario that will, in fact, become reality, as the required standards have already been mandated in the National Building Code of Finland (section D2, 2003).

The main findings are summarized in the sub sections below. The reader is directed to the original articles for more detailed presentation.

5.1. Direct Microenvironment Model (III)

The simulation framework was applied on $PM_{2.5}$ exposures in Helsinki in the *direct* mode in **III**. The simulated exposure distributions matched the observed ones well, especially when the ETS exposures were excluded from the model.

Four simulation models were built; the first two crude models targeted the whole *EXPOLIS* population without using any sub groups. The refined models 3 and 4 excluded ETS-exposed subjects (Another option would have been to model the ETS-exposures as separate indoor sources, but this was done later as a part of the *nested* model in **V**). In the models 3 and 4 the time-activity of the working and non-working subjects were also modelled separately.

The distribution assumptions of lognormality of concentration distributions and beta-distribution for the time fractions were tested statistically and graphically. The concentrations followed lognormal distributions quite well. The goodness-of-fit of the beta distribution for the time fractions was worse.

5.1.1. Simulation of Population Time-Activity

Simulated three-microenvironment fractions of population time were compared to corresponding observed distributions in **III**, Figure 2. First, the whole *EXPOLIS* population was grouped together in the left column of charts labelled “whole population”. Simulated distributions are shown as lines and observed ones as histograms. X-axis represents the fraction of time spent in each microenvironment; y-axis shows a measure of the relative frequency of each value in the distribution (defined so that the area under each distribution is unity).

For the home microenvironment (topmost chart) the Figure 2 in **III** displays a clear underestimation of the relative frequency of the mode and other central percentiles. This underestimation is compensated in the tails of the distribution around time fractions 0.25 – 0.45 and 0.60 – 0.85 where the modelled frequencies are to high. The mode of the fraction of time distribution is somewhat shifted to the right (i.e. overestimated) by the model. Simulated frequencies for those that spent their time almost completely at home are underestimated.

For the workplace the most obvious discrepancy between the simulated and observed distribution in **III**, Figure 2 is the significant probability mass at zero, representing the subjects that did not spend any time at work. This might include some occupied subjects that happened to be off-duty for the measurement period and is called the “non-working” subpopulation for simplicity’s sake. The simulated beta distribution is shifted to left and the observed mode around fraction of time 0.35 is underestimated to be around 0.20. The main cause of this problem, the probability mass at zero, cannot be handled by the beta distribution.

The fitted beta distribution has a closest resemblance to the observed one for the “Other” microenvironment class (the bottom chart in **III**, Figure 2). The mode height is still somewhat underestimated. Kolmogorov-Smirnov test for the above comparison shows clearly that the fitted beta distributions are not statistically representative of the histograms.

In the second step the *EXPOLIS* population was divided into two main categories according to the major modifier of their time-activity: the working status. In the centre and right column of charts in **III**, Figure 2 (labelled “working” and “non working sub population”) the fitted beta distributions have much better resemblance with the observed ones. Still, for the working sub population the mode frequencies are underestimated for all three microenvironments.

Kolmogorov-Smirnov test still indicates statistically significant differences for all cases of the working subpopulation (p-values below 2%), but the two non-working population distributions are acceptable even in terms of statistical significance (p-values >0.25).

5.1.2. Microenvironment Concentration Distributions

Simulated and observed microenvironment concentration distributions for the homes and workplaces were compared visually in **III**, Figure 3. Visually all the five fits seem to capture the overall shape of the observed data. The main determinant of the microenvironment concentrations was clearly shown to be exposure to tobacco smoke. Both of the distributions on the left column of charts labelled “whole population” show slight indications of two modes, the higher mode being attributable to smoking. Because smoking in residences in Finland is becoming rare, the second mode in the home distribution is clearly weaker than the first one, attributable to other $PM_{2.5}$ sources than smoking. In the workplace case the smoking mode is more profound.

Shapiro-Wilk’s test indicated statistically significant deviations from the lognormal distribution fitted using method of matching moments (Small, 1990) (p-values < 0.00). Same result applied to the distribution of ambient 1-hour concentrations from Vallila monitoring station. In the Vallila case the cause for the statistical deviation from log-normality were negative measurement results close to zero that were coded as zeros for the analysis.

When the ETS-exposed microenvironments are excluded from the data, the lognormal fits become statistically acceptable (p-values 0.2 and 0.6 for homes and workplaces, respectively).

5.2. Nested Model: the Infiltration Approach (IV, V)

The next step, after the functionality of the *direct* simulation was affirmed, was to add the *nested* layer of modelling indoor concentrations using outdoor concentrations, infiltration factors, and indoor sources as inputs. The basic time-activity model remained the same, but the number of microenvironments was increased from 2-3 to 4 by splitting the aggregate group “Other” into “Traffic” and “Other-non-traffic”.

5.2.1. Infiltration Factors (IV, V)

Infiltration factors and fractional concentrations from indoor and outdoor sources cannot be directly measured in practical situations, where both indoor and outdoor sources are present. Therefore these terms have to be analysed from the observed total concentrations. In the current work sulphur was used as a particle bound marker element that seemed to have no indoor sources in Helsinki or the other cities included in the analysis.

Residential indoor $PM_{2.5}$ concentrations regressed well against corresponding outdoor concentrations in Helsinki (slope 0.64, p-value <0.000). Corresponding slope for sulphur were 0.76 (p-value <0.000), showing that the particles with high sulphur content, infiltrate indoors with a slightly higher rate. This was expected, because sulphur is mostly of secondary origin in air and is mostly present in submicron accumulation mode particles. A significant fraction of the mass-based $PM_{2.5}$ concentration, on the other hand, is in the largest particles. The larger particles have higher settling velocities and therefore are removed from the indoor air more rapidly, leading to a lower infiltration rate even in case when the penetration rate of both particles would be identical. However, in cases of tightly sealed buildings with coarse filtering in the air exchange system, the larger $PM_{2.5}$ particles are also removed more efficiently at entry. For these reasons, when using sulphur as a marker for particles of ambient origin, the sulphur infiltration rate should be corrected for these differences caused by the different size distributions. The ratio of the regression slopes (0.84) was used to scale sulphur infiltration factors for $PM_{2.5}$ in individual residences.

Concurrent outdoor measurements were not available for the workplace locations. Therefore the infiltration factor analysis for the workplaces was conducted using the residential nighttime outdoor sulphur concentrations, daytime workplace indoor sulphur concentrations, and daytime $PM_{2.5}$ concentrations from the Vallila fixed monitoring station. PM-bound sulphur, being a long-range transported pollutant, does not have a diurnal pattern or any significant spatial variation in the Helsinki metropolitan area. Consequently this replacement of missing observations should not introduce significant bias (i.e. systematic error) to the results. Naturally in individual cases the uncertainty of the infiltration rates is higher.

The resulting mean infiltration factor for the workplaces was significantly lower (mean 0.47) than that for residences (0.64). This could be expected and is presumably mainly due to the higher percentage of mechanical ventilation systems with PM filtering in office and other occupational buildings than in residential buildings.

5.2.2. Indoor sources (IV, V)

Estimation of the infiltration rates for individual indoor environments allowed, together with the observed outdoor concentrations, for calculation of the level of outdoor generated particles indoors. This, subtracted from the observed indoor concentration, is then an estimate for the indoor generated $PM_{2.5}$ level. Assuming a constant decay rate for $PM_{2.5}$ particles based on the PTEAM study in Riverside, U.S., also the ventilation rates (h^{-1}) and consequently the source strengths could be estimated for residences. Indoor source generated concentrations were 2.5 and 4.2 $\mu g\ m^{-3}$ in non-ETS exposed residences and workplaces, respectively. In the residences ventilation rate was 0.8 h^{-1} and mean indoor $PM_{2.5}$ source strength was 0.6 mg h^{-1}. Relative variability of the indoor generated particle levels was much higher than that of the infiltration factors.

The simulation of the indoor concentrations in the next step will show that the presented estimates for the infiltration factors and indoor source strengths produce reasonable total concentration distributions when compared to corresponding observations.

5.2.3. Simulation of Indoor Concentrations (V)

For simulation model component evaluation, the simulated indoor concentrations were compared against corresponding observed distributions (**V**). The comparison included both a *direct* model, where the indoor concentration model consisted of a lognormal distribution fitted to the observations, and a *nested* model where the distributions of infiltration factors and indoor source generated concentrations were used as inputs in the simulation model together with a distribution of ambient concentrations. Numerical results for the latter approach are shown for residences in Table 8

In the residences (**V**, Figure 7, left chart) the performance of both approaches was almost identical and matched the observations very well. In the case of workplaces (right chart in the same figure) both modelling approaches had a lower correspondence to the observed distribution. The *direct* model predicted the upper half of the distribution quite well with rather clear overestimation of the highest five percentiles, but somewhat underestimated the lower percentiles. In absolute terms the underestimation, however, was small (<1 µg m^{-3}). The *nested* model matched the lower tail quite well, but underestimated the percentiles between the 70th and the 95th. In relative terms the biggest underestimation for the 95th percentile was almost 30%.

Table 8. Comparison of simulated and observed residential indoor concentration distributions.

	n	First Moments mean [µg m-3]	sd [µg m-3]	Percentiles 5 % [µg m-3]	10 % [µg m-3]	25 % [µg m-3]	50 % [µg m-3]	75 % [µg m-3]	90 % [µg m-3]	95 % [µg m-3]
Simulated	2000	8.80	5.82	2.8	3.4	5.0	7.4	10.9	15.6	19.5
Observed	153	8.76	5.66	2.7	3.4	4.7	7.1	11.0	18.1	21.2
Difference:										
Sim - Obs		+0.0	+0.2	+0.1	+0.1	+0.2	+0.3	-0.0	-2.5	-1.7
Relative to Obs		**+0.5%**	**+2.9%**	**+4.8%**	**+1.7%**	**+4.7%**	**+4.8%**	**-0.4%**	**-13.6%**	**-8.1%**

An alternative approach to the indoor concentration simulation used by many modellers would have been a mass-balance approach (Yeh and Small, 2002;Burke et al., 2001). It requires more input data, some of which are not widely available or easy to measure. The infiltration approach selected here is based on the same overall equation, but only two probability distributions are estimated (for F_{INF} and C_{ig}, see symbol definitions in **IV**) instead of five (for P, a, k, Q and V). The more detailed mass-balance approach is more flexible in modelling various technical changes affecting ventilation patterns and indoor sources, but as it is based on more numerous inputs it is potentially more prone to parameter uncertainty induced errors than the infiltration model.

5.2.4. Model Evaluation: Characterisation of Model Errors (V)

Model evaluation can be attempted using different setups, some of which are depicted in Figure 6. An exposure model is based on a conceptual model and its implementation includes the definition of input variables used in the model calculations. These input values are typically estimated using measurements from a population sample. Even a randomly drawn sample gives imperfect information on the true values of the variables of interest in the whole population due to sampling error (response bias can be added due to imperfect sampling). The extent to which the sample represents the whole population is called "representativeness" and for a good random sample it is a function of the sample size. Case 1 in Figure 6 describes the calculation of the model error, which will be pursued in more detail shortly.

Case 2 in Figure 6 describes the use of an independent data set for model validation, partly utilized e.g. in (Ott et al., 1988). While this setup makes sure that any specific relationship of the model structure and the sample 1 are not driving the model results, and the model results really can describe another population sample as well, two separate sampling errors are added to the comparison. Case 3 adds another layer of sampling errors and representativity issues to the comparison by using input values created from multiple samples of the target population.

The model evaluation in the current work was done in **V** by quantifying the model errors using setup case 1 for the non-ETS exposed Finnish speaking working age Helsinki metropolitan area inhabitants. The model errors were quantified by comparing the observed and simulated distributions, and compared to the other error terms affecting population exposure assessments: the error in the observed exposure distribution caused by measurement error and to the sampling error in the observed distribution caused by the random sampling process. The latter represents the uncertainty in the field study results in representing the true underlying target population.

Graphical comparison of the simulation results and the observed distribution is shown in **V**, Figure 5. It can be seen that the overall match is similar for both the *direct* and *nested* models. For the upper half of the distribution the *direct* model performs slightly better, and both models somewhat underestimate the observed levels. In the lower half of the distribution the models perform identically. The same comparison is presented numerically in **V**, Table 3. The *direct* model overestimates population mean exposure by 1%, the *nested* model underestimates it by 5%. Both results can be considered at least satisfactory. The model errors are bigger for the standard deviation, which is underestimated by both models, by –9 and –23% by *direct* and *nested* models, respectively. In the 25th and 50th percentiles the relative error approaches 10%, but is well below 1 μg m^{-3} in absolute terms. Such an error is comparable to the measurement error in a single measurement. Highest model error occurs for the 99th percentile in the *nested* model – this level is underestimated by –18%. The corresponding absolute error is –6 μg m^{-3}.

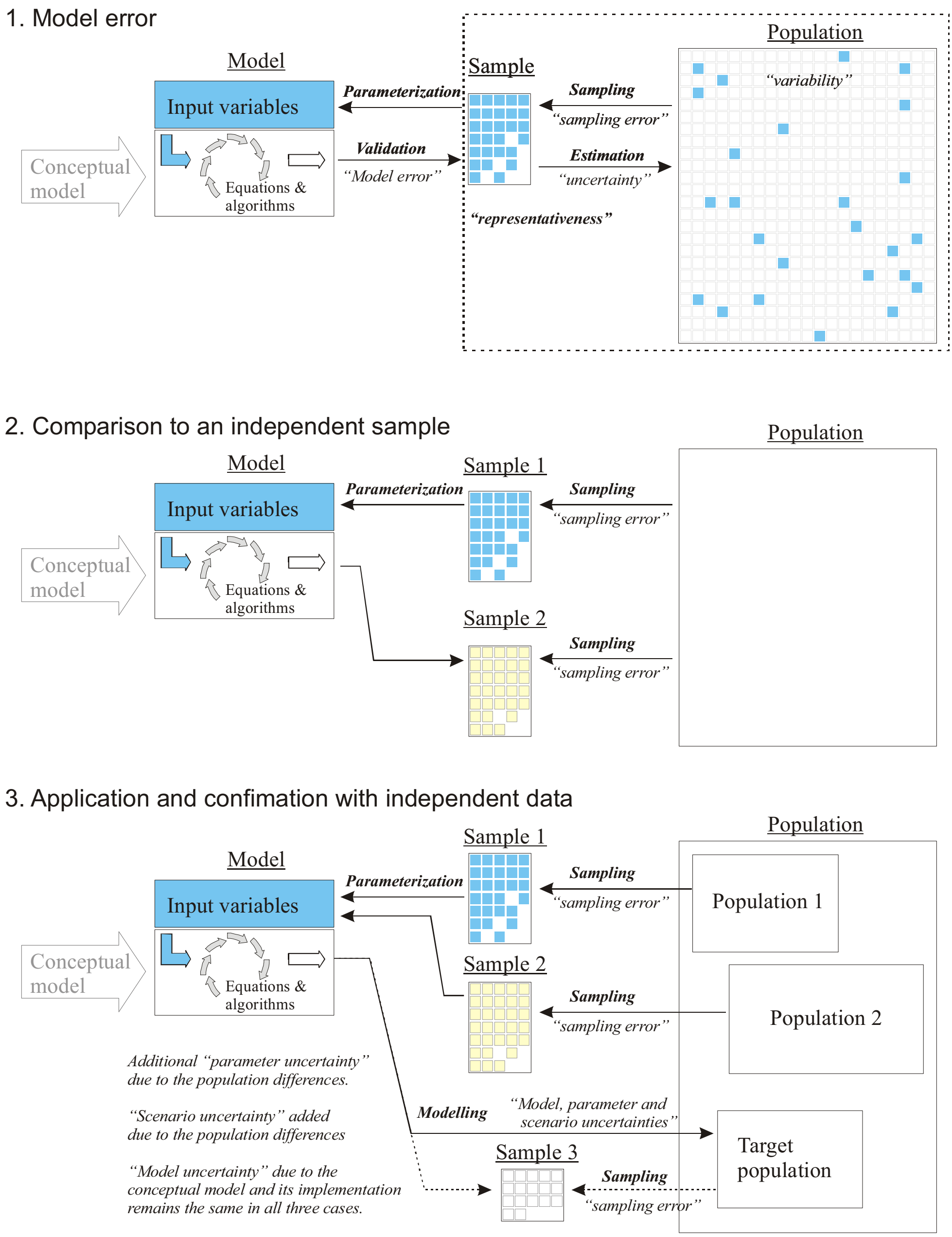

Figure 6. Different possible setups for model evaluation. Setup 1 allows for estimation of model error by excluding probabilistic sampling errors.

The different error terms affecting population exposure assessments are compared in **V**, Figure 6. The top chart displays the uncertainty caused by population sampling. The current study with its 201 exposure measurement subjects can be considered a medium-to-large sized exposure study, and yet the uncertainty in the exposure percentiles is notably large. In the percentiles above 90th the uncertainty increases above ±10 µg m^{-3}.

The middle chart displays the effect of measurement errors. The light grey area displays the measurement error in single personal exposure measurements. The dotted line displays the corresponding bias in the observed distribution. The dark grey area displays an estimate for the uncertainty in this bias by assuming 0.5 (the edge of the dark grey area that is closer to zero) and 2 x (the other edge) measurement error. It can be seen, that the measurement error biases the lower tail low and upper tail high, meaning that the ob served distribution is, in fact wider than the true underlying distribution. Because the measurement error adds a random variation component to the observations, this is natural.

The bottom chart in the **V**, Figure 6, displays the measurement error bias corrected model errors for the *direct* and *nested* models. These are comparable, the *direct* model being slightly more accurate. The model errors are somewhat smaller than – but comparable to – the uncertainty about the true population exposure distribution caused by the random sampling error. It should be noted that as this analysis of sampling error accounts only for the effect of random sampling; it does not include any effects of potential participation bias or subject modification of behaviour. Therefore the random sampling error represents a minimum estimate for the sampling error component.

5.3. Application: Risk Reduction Potential of Good Ventilation (VI)

$PM_{2.5}$ infiltration factor distribution for all residential buildings was 0.64±0.20 and 0.47±0.24 for occupational buildings. In the occupational buildings mechanical ventilation systems with at least coarse particle filters are more common than in residential buildings. However, in the newer buildings, which in the current study are represented by buildings built in or after 1990, the corresponding mean values are 0.58 and 0.35, respectively, indicating a clear lowering tendency. The difference is especially significant in relative terms for occupational buildings.

The simulation model developed and validated in the earlier part of the work was now applied for the estimation of the exposure reduction potential in a future scenario, where the infiltration efficiency of all buildings would follow the distribution of infiltration factors in the post 1990 occupational buildings in the *EXPOLIS* sample. It was assumed that all the other model parameters would be unchanged.

Because the infiltration efficiency affects mainly particles of ambient origin, the model was run without indoor sources. The health effects connected with the ambient PM levels in the epidemiological studies must be caused by ambient particles, because the indoor generated particle levels do not correlate with the ambient levels. Therefore, if the indoor generated particles have similar health effects than the ambient particles, they are additional to those observed in the epidemiological studies. Therefore the exposure reduction potential for the demonstration case was calculated mainly for the ambient particles.

The results (**VI**, Table 3) indicated that a 27% reduction could be achieved by the changes in the ventilation systems. Because the needed requirements have already been implemented in the Building Codes, it can be assumed that this reduction will be achieved along the natural renewal of the building stock in every case. When new understanding is generated on the risks caused by particles to susceptible population groups, special actions regarding building characteristics can be taken to target exposure reductions to those that will benefit most.

6. DISCUSSION

Comparison of deterministic and probabilistic approaches. In the strictest sense of the phrase, deterministic models are based on physical equations describing causal relationships and target identifiable individuals and events. A single outcome of such a model could at least in principle be validated by comparing the model estimate to a corresponding observation. Deterministic models are, of course, not limited to modelling ‘specific individuals or events’. Large populations may be modelled by including all population members individually into the model. A common objection to deterministic models is that the collection of the input data needed for such an attempt would be impossible. But there is no need to include every member of the target population in a deterministic model, similarly as no one would suggest this for a personal exposure monitoring study. A statistical sampling scheme can be employed to create a random sample of the target population, to collect the required input data for this more limited number of subjects, and to run the model for them.

Strengths of deterministic exposure models include presentation of exposures in geographical scales (using GIS), short-term forecasting, and modelling of alternative future scenarios. Practical challenges of the deterministic modeller may be solved using probabilistic approaches. It is obvious that in fact we do not need exposure data for specific individuals to manage exposures in a city or for a specific sub population. Relevant are the general exposure characteristics of the target population, including estimates for the mean exposure, exposure variability, and perhaps some idea of the levels of the highest exposures. For a model to be useful, it should help answering questions like “How could we best reduce these exposures?” and “How much would the exposures be reduced if we implemented these management options?” Of course, a model can replace neither the exposure analyst nor the decision maker in this process, but the model should be usable as a tool for comparing alternative options and scenarios for them.

Because it is very difficult or practically impossible to collect individual data for anything more than small samples of selected populations, the deterministic modeller is drifted towards estimating input variables with point values more or less representative for the target population. Such point values are in the best case not biased, but they always lead to ignoring some of the variability of the values within the target population. Therefore such model can in the best case estimate the population mean exposures well, but the estimates of variability will be compromised. This is exactly the main issue that a probabilistic modeller tries to solve. Probabilistic input variables are described as distributions that intend to capture the true variability of the input values.

Another point related more to risk than exposure modelling, is the use of conservative point estimates in the models to create a safety margin. In such a context, instead of using conservative point estimates in a deterministic model, the probabilistic modeller tries to capture the true variability (and sometimes also separately the uncertainty) in the input variables, and to create a best estimate for the whole range of variability of the exposure in the target population. Then, it is on the responsibility of the risk manager to apply a required margin of safety on top of the exposure assessment representing our best knowledge (with explicitly expressed uncertainty) on the true exposures.

Model development and data acquisition. In the current work a population exposure model was developed in the context of a large European multi-centre study with extensive fieldwork in seven metropolitan areas. This directed resources towards the data collection, including personal exposure and microenvironment concentration measurements with the accompanying work related to development of measurement methods, quality assurance, multi-centre collaboration, data management, and data analysis, and it is difficult to avoid the question whether such a large field study gives the best environment for model development. The current work would have benefited more from an environment focusing on model development with support for theoretical aspects, computer based modelling, and statistical and mathematical expertise.

On the other hand, a major limitation in many deterministic and probabilistic modelling attempts is the implicit uncertainty in the model inputs and outputs. When model inputs are estimated from various sources, including literature and pilot studies to mention few examples, the only way to assess the applicability and representability of the data for the purpose at hand is expert judgement. Here, at best, science enters to the round table of experts, where the peer review of the presented models and results judges the validity of selections and assumptions made by the model authors.

On the other hand, when the model input data are collected using population-based random sampling, it is ensured that the data entering to the model are representative of the underlying population. Traditional statistical techniques can then be used to assess the uncertainty about the underlying population caused by the random sampling process. Collecting observations of the model output variables at the same time and from the same subjects makes it possible to compare the observed and predicted values to calculate the model errors as the difference of these two.

Estimation of model parameters from observations. In an ideal world a good model would use easily observable variables as inputs and calculate the desired outputs from those using physical equations completely capturing causal relationships between the inputs and outputs.

Unfortunately we do not live in an ideal world. Taking ventilation as an example, it is operated by individuals, affected by e.g. ambient temperature and stochastic events like burning a toast, with a great personal variability, suitable for probabilistic characterization at best. A modeller could attempt to use questionnaire data specific to the day and apartment in case, or a typical value (perhaps classified more specifically to the type of day and apartment and other factors perhaps affecting the outcome). The first option becomes soon too detailed and demanding when the target population size increases. The second option in the simplest case uses population average as a point value for a specific individual, or uses statistical modelling to estimate it from other variables. This is not far from full-fledged probabilistic modelling, where uncertain statistical determinants can be left out of the model and replaced by a description of the variability of each variable.

Attempts to model validation. As Oreskes *et al*. (1994) point out in their rather philosophical study, it is it is always impossible to 'validate' a model in an open system in a pre-emptive way. This is similar to ideas presented much earlier by Karl Popper (Popper, 1935) about falsification of a scientific theory: even what we considered the laws of nature are subject to falsification they are applied in a new environment, where new forces became effective. Any success in model evaluations may only increase gradually on our trust on the model. When the model fails in a new setting, limits of the model applicability become clearer. A classic example from physics were the measurements of the speed of light in late nineteenth century that led to the birth of the theory of relativity few years later and changed our understanding on the nature of gravity. In exposure modelling similar limits of model applicability may be associated with interactions of relatively simple phenomena like air exchange of an unoccupied room interacting with its complete environment including human behaviour in the rest of the building, ambient wind, temperature, radiation balance, etc.

Popper and Oreskes *et al*. are, of course, right in principle. On the other hand also the need for different kind of models and the evaluation of their accuracy are very real. Therefore Oreskes et al.'s point should not be taken as discouragement for evaluation of model accuracy. Decision makers, for example, need to be aware of the uncertainties in the model predictions that they rely on when making expensive or restrictive decisions to protect the safety of the public. This very well illustrated by the benzene exposure reduction case in California (Jantunen, 1998;Ott, 1995) where expensive requirements were set on industry to reduce their benzene emissions. Later it turned out that a simple evaluation of population exposures to benzene would have saved all the trouble, as the controlled industrial emissions had only marginal impact on population exposures, which were driven by tobacco smoke and traffic. The underlying model that the population risk is a straightforward function of emission tons was false and the decision makers should not have counted on it in the first place.

Ott *et al*. (1988) and others have argued that the model validation data set has to be independent of the one used for the model development. Ott *et al*. used the personal CO monitoring data from Denver, Colorado, to develop the SHAPE (Simulation of Human Air Pollution Exposures) model. In the monitoring study the exposures were logged with 1-minute resolution for two days per subject. Ott *et al*. used the first day data to create concentration distributions for the 22 microenvironments included in the model. Then they combined these distributions with the time-activity diaries for the day 2 and compared the model outputs with the observed day 2 exposures, claiming that now the model development data (day 1) and the model validation data (day 2) were independent. However, this approach can be expected to work only if the true day 2 concentration distributions were similar to the day 1 ones. If, e.g. different mixing conditions, or ambient temperature that would affect the use of indoor heaters and ventilation patterns, would be different for the second day, there would be no reason to expect the day 1 distributions to represent the day 2 ones.

The above example demonstrates that the requirement for independent data for model validation is problematic; the input data used in validation must be representative of the target system from where the corresponding observations are collected. If this is not the case, then similarities or differences in the input values may drive our comparison and conclusions, and this of course makes no sense. Therefore in the current study the model input values and the personal exposures used in the model validation were specifically collected from the same population sample.

Underestimation of variance. In the two validation studies for probabilistic population exposure models one common finding has been the underestimation of exposure variability (Law et al., 1997;Ott et al., 1988). One factor not mentioned by the authors is the use of 1-minute concentration data in combination with time-activity diaries. Individual entries in time activity diaries may have significant timing errors due to watches, recall errors, errors in filling the diary, and errors in the data entry into the database. These dilute the estimated concentrations in all microenvironments towards the overall average concentration, i.e. the concentration variation is underestimated. Moreover, in those microenvironments, where the concentrations are especially high, like in the case under study focusing on CO exposures, parking garages, highways, street traffic, tunnels, gasoline stations etc., the time spent is very short. Even a minor error of few minutes in the timing of the visit to such a microenvironment will have a significant effect on the observed average concentration for the visit. Minor timing errors do not have remarkable effects on microenvironments where the time spent is hours.

Time-activity modelling. The most common approach to time activity modelling is to use a database of actual time-activity diaries. Such a database is sampled in the simulation; individuals with the correct gender, age, and ethnic, socioeconomic, and other characteristics for the current simulated population group are randomly selected and used in the simulation (AirPex, SHAPE, pNEM etc.). The main strength of this approach is that the actual sequential dependences between visits to various microenvironments are completely maintained. The actual diaries are also very suitable for tying the visits to specific times of the day.

On the other hand, if the model is used for future scenarios, it must either be assumed that the time-activity of the population does not change, setting a limit to the scenarios that can be studied, or the change in time activity must be implemented on each of the used diaries in the database. Both alternatives are limiting from the point of view of model application. Therefore a different approach was selected in the current work. The time of day and sequential nature of visits to microenvironments are merely ignored, and the total daily fraction of time used in each microenvironment is used instead. This way the time-activity inputs are very easily documented and hypothetical changes can be easily applied to them.

Correlations. One new feature that seems to have been added to probabilistic exposure models in the current work is the statistical modelling of correlations of values sampled from various input distributions. The sampling used in basic probabilistic model simulations assumes independent input distributions. In such a model the dependencies between model variables should be causally modelled as far as possible, but those phenomena for which no causal relationships are specified, are assumed independent. This is not true in the real world. A good example is exposure to tobacco smoke. Smoking subjects are more likely to be exposed to ETS in all of the microenvironments they visit, and subjects sensitive to tobacco smoke will try to avoid all contact with it.

Correlations of microenvironment concentrations can be partly traced back to correlations of ambient concentrations. Ott *et al*. (1988) used this in their SHAPE model, where the microenvironment concentrations were split into an ambient background component and microenvironment specific component. On the other hand, also other factors may affect the correlations of microenvironment concentrations. For example a smoking subject is likely to be exposed to higher levels at both home and workplace – and even the restaurants he or she visits. Also, daily ambient temperatures and the season affect the ventilation patterns and thus modify the infiltration rates in a way that will increase the correlation between the different microenvironments. As a conclusion, the factors leading to correlations can partly be traced back to causal issues (e.g. the general ambient background level), but partly are merely statistical phenomena. In this sense it can be said that ultimately it might be impossible to capture the full range of variability of exposures using purely deterministic models.

Simulation of exposures to other air pollutants. The original goals set for the simulation model development presented in this thesis were not limited to $PM_{2.5}$. In principal the simulation framework, and the conceptual exposure model behind it, are generic and can be applied to different pollutants, as demonstrated e.g. by the simulations run for PM_{10} (**II**) and CO (Bruinen de Bruin et al., 2004b). In models for other pollutants than $PM_{2.5}$, the role of different microenvironments and population groups must be considered separately. Exposures to benzene are driven by different microenvironments than exposures to particles, and even when looking at different size fractions of particles, or particles from specific sources, the microenvironments to be included in the model must be carefully considered.

Exposure-response relationships. During the past decade of intensive research on health effects of particulate matter it has become evident that not all particles are equally toxic, nor are all people equally sensitive to the toxicity of the particles. It is clear that there are many toxic components in particulate matter and that the toxicity is mediated via numerous mechanisms. As the epidemiological and toxicological studies bring more light to the subject, the question about environmental health protection and particles becomes increasingly complex. For each mechanism affecting health there are susceptible population groups, and particles from different sources affect different health mechanisms differently. Therefore the answer to the question: “How can we reduce these health effects most effectively?” requires population group level assessment of exposures to a multitude of PM fractions.

The development of the current modelling approach towards this direction has already begun. In the national HEAT study we have specifically modelled exposures to traffic generated combustion particles (a.k.a. tailpipe particles) (Tainio et al., 2005). In the EU-funded FUMAPEX study we have looked at $PM_{2.5}$ exposures of the most important general population groups that are considered susceptible to particles: elderly and infants (unpublished work). Much remains, however, to be done in this area.

Exposure Modelling and Air Pollution Risk Management. Risk management policies cost money and restrict the alternatives available for individuals and institutions. The justification for such policies is the reduced mortality and disease burden. Therefore the public health achievements of the implemented policies should be evaluated against the set risk reduction objectives. The achieved mortality and morbidity reductions due to implementation of an air pollution policy, however, are in most cases practically impossible to measure. Implementation takes years, and other simultaneous changes in diseases, treatments, demography, and other environmental parameters will inevitably, and in many unknown ways, change the population mortality and morbidity – with all likelihood more than air pollution reduction. Options as dramatic, instantaneous, and effective, as the banning of coal sale for domestic use in Dublin in 1990, are rarely identified and even more rarely successfully implemented (Clancy et al., 2002). While the ultimate goal of urban air pollution abatement policy is to reduce the avoidable disease burden, the targets must be set on intermediate goal, reduction of air pollution exposures, because this can be planned, modelled, managed, measured, and verified independently from other developments in the society. When alternative future policies are being compared, exposure modelling is the only means to perform this important comparison.

Exposure to some pollutants may concern only a small minority of the public. This may be the case, if this minority has much higher exposures than the rest (occupational or vicinity to a source), or if the minority is exceptionally sensitive to this pollutant (e.g. allergic). In these cases, the target population must, of course, be selected accordingly.

7. CONCLUSIONS

The developed probabilistic modelling techniques can be successfully used for modelling of population exposures to $PM_{2.5}$, capturing the population variability of exposures (**II, III, V**). The model is suitable for comparison of alternative future scenarios (**VI**). Such analysis should be conducted regularly for optimization of environmental policies (**VII**). The following paragraphs list the main conclusions associated with the detailed study aims.

7.1. Study design (I, II, III, V, VII)

◆ Integrated population-based measurement of exposures and affecting factors (i.e. microenvironment concentrations, time-activity, etc.) allows for detailed analysis of exposure determinants and development of exposure models with detailed evaluation.

◆ Population-based sampling of subjects ensures that the observations, and thus exposure analysis and modelling based on them, are representative for the general population from which the random sample was drawn.

7.2. Simulation Framework (II, III, V)

◆ Implementation of the modelling system using a pre-structured framework makes model development faster, easier and more reliable.

◆ Inclusion of correlation structures is much easier using a pre-structured approach.

7.3. Model input estimation methods (III, IV, V, VI)

◆ Some model parameters (best examples being infiltration factors and indoor source strengths) are not directly measurable, but can be estimated from observed variables using state of art numerical analysis techniques.

◆ Correlations are population level features that can be estimated from population sampled data. The causal dependencies between model variables should be modelled as such as far as possible; however, in some cases this may not be possible. Probabilistic model with the rank correlation feature is one solution to the modelling of these features that are not easily included in physical models.

◆ In model application many parameters must be estimated based on assumptions on local conditions etc., or values measured elsewhere must be used in lack of local data; heterogeneity of correlation structures, infiltration factors and other input values remains an interesting and potentially important research area relevant to future applications.

◆ Goodness-of-fit evaluation methods for probabilistic exposure modelling are not very well established. Some methods based on p-values indicate statistically highly significant differences for distributions that are for all practical purposes identical. On the other hand in some cases (especially time-activity modelling) even visually obvious discrepancies have only minor effects in simulation results. Evaluation of GOF should not be excluded in data-based modelling studies, but care should be taken in interpretation of the results.

7.4. Model Accuracy (II, III, V, VI)

◆ Model errors were found to be relatively small; comparative or smaller than population sampling uncertainties.

◆ Measurement error is typically smaller in microenvironment monitoring than in personal exposure measurement (in case of $PM_{2.5}$ due to the larger flow rates and consequent sample sizes) and therefore modelling based on the microenvironment monitoring can produce more accurate results than personal exposure monitoring.

◆ Simulation models can be used to estimate population variances (unlike deterministic models without proper population sampling schemes), but as found also in previous studies, tend to underestimate exposure variances. With inclusion of correlations and by taking into account the measurement error bias in the observed exposure distribution the underestimation can be alleviated but not completely removed.

◆ Model predicts $PM_{2.5}$ exposure percentiles from 5^{th} to 95^{th} very well; in the tails the model errors become relatively (lower tail) or absolutely (upper tail) larger. Only the upper tail underestimation has practical significance for exposure management.

7.5. Model error, uncertainty and need for independent data (V)

◆ Uncertainty concerns probabilistic evaluation of possible errors in model estimates; more precise and not probabilistic model error may be estimated using a observations of the model output variables together with carefully designed setup that removes other error terms.

◆ Quantification of model error must be based on model inputs and comparison data from the same population sample and times, because otherwise sampling errors obscure them.

◆ Requirement of independent data for model evaluation applies for evaluating model equations and algorithms in alternative setups. In such tests the input data used must describe the alternative target system.

7.6. Model application for a policy-relevant scenario (VI)

◆ Successful model application demonstrated that the developed modelling environment can be used to estimate reductions in exposures for given exposure scenarios.

◆ Population-based exposure studies allow for data based development of exposure scenarios.

◆ The model itself can be applied for hypothetical scenarios (with increased uncertainty).

7.7. Development of efficient environmental policies (II, V, VI, VII)

◆ Policy decisions must be based on reliable quantitative estimates of the expected benefits.

◆ The model was validated for the current exposure scenario and applied successfully for a data based future scenario.

◆ Preliminary scenarios may be created using theoretical assumptions about model inputs, but a data based approach, as demonstrated, ties the scenarios more tightly to reality.

◆ Limitations in obtaining model parameters concern alternative modelling approaches, too.

◆ Exposure assessment using this kind of models allows for realistic and quantitative risk assessment and management.

8. Impacts on Environmental Policy and Public Health (VII)

Exposure analysis is a crucial part of effective management of public health risks caused by pollutants and chemicals in our environment. Development of science-based policies for promotion of public health requires careful analysis of exposures within the population, including identification of emission sources, exposure routes, behavioural determinants, and population groups at risk. Comparison of alternative future policies in terms of environmental health is possible only by using exposure models. One such model was developed and evaluated in the current work with encouraging results.

Exposures to specific pollutants vary from subpopulation to another, and various policy options affect these exposures with largely different efficacies. Therefore future exposure and risk analyses should be carried out in population group level. Optimal benefits can be achieved by reducing exposures specifically in those subpopulations where the burden of adverse health effects is the highest.

In the case of particulate matter, the pollutant itself consists of different fractions, with presumably different toxicities, and thus in this case the dose–response factors should be determined for each of these fractions. If analysis of population exposures is based on only centrally monitored ambient air quality data and dose-response factors obtained for the general population, non-optimal policies may be selected.

EC pursues to develop guidelines for new pollutants, including $PM_{2.5}$, and methodologies to control exposures to pollutants and chemicals with significant indoor sources. The collected exposure data in the *EXPOLIS* database and the models developed as part of the current work should, can and will be used to support these processes among other available tools and exposure analysis techniques.

REFERENCES

1. Alm, S., Mukala, K., Jantunen, M.J., 2000. Personal carbon monoxide exposures of preschool children in Helsinki, Finland: levels and determinants. Atmospheric Environment 34: 277-285.

2. Alm, S., Mukala, K., Pasanen, P., Tiittanen, P., Ruuskanen, J., Tuomisto, J., Jantunen, M.J., 1998. Personal NO2 exposures of preschool children in Helsinki. Journal of Exposure Analysis and Environmental Epidemiology 8(1): 79-100.

3. Alm, S., Mukala, K., Tiittanen, P., Jantunen, M.J., 2001. Personal carbon monoxide exposures of preschool children in Helsinki, Finland - comparison to ambient air concentrations. Atmospheric Environment 35: 6259-6266.

4. Alm, S., Reponen, A., Mukala, K., Pasanen, P., Tuomisto, J., Jantunen, M.J., 1994. Personal exposures of preschool children to carbon monoxide: roles of ambient air quality and gas stoves. Atmospheric Environment 28: 3577-3580.

5. Basunia, M.S., Landsberger, S., Yli-Tuomi, T., Hopke, P.K., Wishinski, P., Paatero, J., Viisanen, Y., 2003. Ambient silver concentration anomaly in the Finnish arctic lower atmosphere. Environmental Science & Technology 37(24): 5537-5544.

6. Bell, M., Davis, D.L., 2001. Reassessment of the Lethal London Fog of 1952: Novel Indicators of Acute and Chronic Consequences of Acute Exposure to Air Pollution. Env Health Pers 109 (S3): 389-394.

7. Briggs, D.J., Denman, A.R., Gulliver, J., Marley, R.F., Kennedy, C.A., Philips, P.S., Field, K., Crockett, R.M., 2003. Time activity modelling of domestic exposures to radon. Journal of Environmental Management 67(2): 107-120.

8. Bruinen de Bruin, Y., Carrer, P., Jantunen, M., Hänninen, O., Scotto di Marco, G., Kephalopoulos, S., Cavallo, D., Maroni, M., 2004a. Personal carbon monoxide exposure levels: contribution of local sources to exposures and microevironment concentrations in Milan. Journal of Exposure Analysis and Environmental Epidemiology 14: 312-322.

9. Bruinen de Bruin, Y., Hänninen, O., Carrer, P., Maroni, M., Kephalopoulos, S., Scotto di Marco, G., Jantunen, M.,2004b. Simulation of working population exposures to carbon monoxide using EXPOLIS-Milan microenvironment concentration and time activity data. Journal of Exposure Analysis and Environmental Epidemiology 14(2), 154-163.

10. Buck, R.J., Hammerstrom, K.A., Ryan, P.B., 1995. Estimating long-term exposures from short-term measurements. Journal of Exposure Analysis and Environmental Epidemiology 5 (3): 359-373.

11. Burke, J.M., Zufall, M.J., Özkaynak, H., 2001. A population exposure model for particulate matter: case study results for PM2.5 in Philadelphia, PA. Journal of Exposure Analysis and Environmental Epidemiology 11(6): 470-489.

12. Burke, T., Anderson, H., Beach, N., Colome, S., Drew, R., Firestone, M., Hauchman, F., Miller, T., Wagener, D., Zeise, L., Tran, N., 1992. Role of Exposure Databases in Risk Management. Arch Env Health 47(6): 421-429.

13. Callahan, M.A., Clickner, R.P., Whitmore, R.W., Kalton, G., Sexton, K., 1995. Overview of important design issues for a National Human Exposure Assessment Survey. Journal of Exposure Analysis and Environmental Epidemiology 5 (3): 257-282.

14. Clancy, L., Goodman, P., Sinclair, H., Dockery, D.W., 2002. Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. Lancet 360(9341): 1210-1214.

15. Clayton, C.A., Pellizzari, E.D., Quackenboss, J.J., 2002. National Human Exposure Assessment Survey: Analysis of exposure pathways and routes for arsenic and lead in EPA region 5. Journal of Exposure Analysis and Environmental Epidemiology 12(1): 29-43.

16. Clayton, C.A., Pellizzari, E.D., Rodes, C.E., Mason, R.E., Piper, L.L., 1999a. Estimating distributions of long-term particulate matter and manganese exposures for residents of Toronto, Canada. Atmospheric Environment 33: 2515-2526.

17. Clayton, C.A., Pellizzari, E.D., Whitmore, R.W., Perrit, R.L., Quackenboss, J.J., 1999b. National Human Exposure Assessment Survey (NHEXAS): distributions and associations of lead, arsenic and volatile organic compounds in EPA region 5. Journal of Exposure Analysis and Environmental Epidemiology 9(5): 381-392.

18. Clayton, C.A., Pellizzari, E.D., Wiener, R.W., 1991. Use of a pilot study for designing a large scale probability study of personal exposure to aerosols. Journal of Exposure Analysis and Environmental Epidemiology 1(4): 407-421.

19. Clayton, C.A., Perrit, R.L., Pellizzari, E.D., Thomas, K.W., Whitmore, R.W., Wallace, L.A., Özkaynak, H., Spengler, J.D., 1993. Particle total exposure assessment methodology (PTEAM) study: distributions of aerosol and elemental concentrations in personal, indoor and outdoor air samples in an southern California com. Journal of Exposure Analysis and Environmental Epidemiology 3: 227-250.

20. Cox, L.A.J., 1999. Internal dose, uncertainty analysis, and complexity of risk models. Environment International 25(6/7): 841-852.

21. Cullen, A. C. and H. C. Frey 1999 Probabilistic techniques in exposure assessment.New York: Plenum Press.

22. Dockery, D.W., Pope, I.C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, J.B.G., Speizer, F.E., 1993. An association between air pollution and mortality in six US cities. The New England J of Medicine 329: 1753-1759.

23. Dockery, D.W., Spengler, J.D., 1981. Personal exposure to respirable particulates and sulfates. JAPCA 31: 153-159.

24. Duan, N., 1982. Models for human exposure to air pollution. Environment International 8: 305-309.

25. Edwards, R.D., Jurvelin, J., Koistinen, K., Saarela, K., Jantunen, M., 2001. VOC source identification from personal and residential indoor, outdoor and workplace microenvironment samples in EXPOLIS-Helsinki, Finland. Atmospheric Environment 35: 4829-4841.

26. Einstein, A., 1905. Zur Elektrodynamik bewegter Körper. Annalen der Physik 17: 891-921.

27. Einstein, A., 1916. Die Grundlage der allgemeinen Relativitätstheorie. Annalen der Physik 49: 769-822.

28. Ezzati, M., Lopez, A.D., Rodgers, A., Vander Hoorn, S., Murray, C.J.L., 2002. Selected major risk factors and global and regional burden of disease. Lancet 360(9343): 1347-1360.

29. Fletcher, T., Brunekreef, B., Fabianova, E., Gurzau, E.S., Houthuijs, D., Lebret, E., Nikiforov, B., Pinter, A., Rudnai, P., Volf, J., van de Wiel, H., Zejda, J.E., 1999. Central European study on air pollution and respiratory health (CESAR): An introduction. Epidemiology 10(4): S29.

30. Freijer, J.I., Bloemen, H.J.T., deLoos, S., Marra, M., Rombout, P.J.A., Steentjes, G.M., van Veen, M.P., 1998. Modelling exposure of the Dutch population to air pollution. Journal of Hazardous Materials 61: 107-114.

31. Frey, H.C., Rhodes, D.S., 1996. Characterizing, simulating and analyzing variability and uncertainty, An Illustration of methods using an air toxics emissions example. Human and ecological risk assessment: an international journal 2(4): 762-797.

32. Fugas, M.,1975. Assessment of total exposure to an air pollutant. Proceedings of the International Conference on Environmental Sensing and Assessment.Paper No.38-5, Vol.2, IEEE #75-CH 1004-1, ICESA, 1975, Las Vegas.

33. Gangnon, R.E., Clayton, M.K., 2004. Likelihood-based tests for localized spatial clustering of disease. Environmetrics 15(8): 797-810.

34. Georgoulis, L.B., Hänninen, O.O., Samoli, E., Katsouyanni, K., Künzli, N., Srám, R.J., Bruinen de Bruin, Y., Alm, S., Jantunen, M., 2002. Personal carbon monoxide exposures in five european cities and its determinants. Atmospheric Environment 36: 963-974.

35. Girman, J.R., Jenkins, P.L., Wesolowski, J.J., 1989. The role of total exposure in air pollution control strategies. Environmental International 15: 511-515.

36. Gordon, S.M., Callahan, P.J., Nishioka, M.G., Brinkman, M.C., O'Rourke, M.K., Lebowitz, M.D., Moschandreas, D.M., 1999. Residential environmental measurements in the National Human Exposure Assessment Survey (NHEXAS) pilot study in Arizona: preliminary results for pesticides and VOCs. Journal of Exposure Analysis and Environmental Epidemiology 9(5): 456-470.

37. Götschi, T., Oglesby, L., Mathys, P., Monn, C., Manalis, N., Koistinen, K., Jantunen, M., Hänninen, O.O., Polanska, L., Künzli, N., 2002. Comparison of Black Smoke and PM2.5 Levels in Indoor and Outdoor Environments of Four European Cities. EST 36(6): 1191-1197.

38. Graham, J., Walker, K., Berry, M., Bryan, E., Callahan, M., Fan, A., Finley, B., Lynch, J., McKone, T., Özkaynak, H., Sexton, K., 1992. Role of exposure databases in risk assessment. Arch Env Health 47(6): 408-420.

39. Gulliver, J., Briggs, D.J., 2004. Personal exposure to particulate air pollution in transport microenvironments. Atmospheric Environment 38(1): 1-8.

40. Hamey, P.Y., 2001. An example to illustrate the potential use of probabilistic modelling to estimate operator exposure to pesticides. Annals of Occupational Hygiene 45(90001): 55S-564.

41. Hänninen, O.O., Alm, S., Kaarakainen, E., Jantunen, M.,2002a. The EXPOLIS Databases. B13/2002. Kuopio, KTL. Publications of the National Public Health Institute.

42. Hänninen, O.O., Koistinen, K.J., Kousa, A., Keski-Karhu, J., Jantunen, M.J., 2002b. Quantitative analysis of environmental factors in differential weighing of blank Teflon filters. Journal of the Air & Waste Management Association 52(2): 134-139.

43. Hartwell, T.D., Pellizzari, E.D., Perritt, R.L., Whitmore, R.W., Zelon, H.S., Wallace, L., 1987. Comparison of volatile organic levels between sites and seasons for the total exposure assessment methodology (team) study. Atmospheric Environment 21: 2413-2424.

44. Hattis, D.B., Burmaster, D.E., 1994. Assessment of Variability and Uncertainty Distributions for Practical Risk Analyses. Risk Analysis 14(5): 713-730.

45. Hertel, O., Jensen, S.S., Andersen, H.V., Palmgren, F., Wahlin, P., Skov, H., Nielsen, I.V., Sorensen, M., Loft, S., Raaschou-Nielsen, O., 2001. Human exposure to traffic pollution. Experience from Danish studies. Pure and Applied Chemistry 73(1): 137-145.

46. Hoffmann, K., Becker, K., Friedrich, C., Helm, D., Krause, C., Seifert, B., 2000a. The German Environmental Survey 1990/92 (GerES II): cadmium in blood, urine and hair of adults and children. Journal of Exposure Analysis and Environmental Epidemiology 10(2): 126-135.

47. Hoffmann, K., Krause, C., Seifert, B., Ullrich, D., 2000b. The German Environmental Survey 1990/92 (GerES II): Sources of personal exposure to volatile organic compounds. Journal of Exposure Analysis and Environmental Epidemiology 10(2): 115-125.

48. Hopke, P.K., Ramadan, Z., Paatero, P., Norris, G.A., Landis, M.S., Williams, R.W., Lewis, C.W., 2003. Receptor modeling of ambient and personal exposure samples: 1998 Baltimore Particulate Matter Epidemiology-Exposure Study. Atmospheric Environment 37(23): 3289-3302.

49. Hunter Youngren, S., Youngren, M.A., Barraj, L., 2001. Challenges of probabilistic assessment of operator and residential non-dietary exposure. Annals of Occupational Hygiene 45(90001): 49S-54.

50. Janssen, N.A.H., de Hartog, J.J., Hoek, G., Harssema, H., Brunekreef, B., Lanki, T., Timonen, K.L., Pekkanen, J., 2000. Personal exposure to fine particulate matter in elderly subjects: relation between personal, indoor, and outdoor concentrations. Journal of the Air & Waste Management Association 50: 1133-1143.

51. Janssen, N.A.H., Hoek, G., Brunegreef, B., Harssema, H., Mensink, I., Zuidhof, A., 1998a. Personal sampling of particles in adults: Relation among personal, indoor, and outdoor air concentrations. American Journal of Epidemiology 147(6): 537-547.

52. Janssen, N.A.H., Hoek, G., Brunekreef, B., Harssema, H., 1999a. Mass concentration and elemental composition of PM10 in classrooms. Occup Environ Med 56: 482-487.

53. Janssen, N.A.H., Hoek, G., Harssema, H., Brunekreef, B., 1997a. Childhood exposure to PM10: relationship between personal, classroom and outdoor concentrations. Occup Environ Med 54: 888-894.

54. Janssen, N.A.H., Hoek, G., Harssema, H., Brunekreef, B., 1998b. Personal sampling of airborne particles: method performance and data quality. Journal of Exposure Analysis and Environmental Epidemiology 8: 37-49.

55. Janssen, N.A.H., Hoek, G., Harssema, H., Brunekreef, B., 1999b. Personal exposure to fine particles in children correlates closely with ambient fine particles. Arch Env Health 54(2): 95-101.

56. Janssen, N.A.H., van Mansom, D.F.M., van der, J.K., Harssema, H., Hoek, G., 1997b. Mass concentration and elemental composition of airborne particulate matter at street and background locations. Atmospheric Environment 31: 1185-1194.

57. Jantunen, M., 1998. Assessing the benefits and costs of air pollution research: Benzene exposure in the San Francisco Bay area. Atmospheric Environment 32(6): 1135-1136.

58. Jensen, S.S., 1998. Mapping human exposure to traffic air pollution using GIS. Journal of Hazardous Materials 61(1-3): 385-392.

59. Karppinen, A., Harkonen, J., Kukkonen, J., Aarnio, P., Koskentalo, T., 2004a. Statistical model for assessing the portion of fine particulate matter transported regionally and long range to urban air. Scandinavian Journal of Work Environment & Health 30: 47-53.

60. Karppinen, A., Härkönen, J., Kukkonen, J., Aarnio, P., Koskentalo, T., 2004b. Statistical model for assessing the portion of fine particulate matter transported regionally and long range to urban air. Scandinavian Journal of Work, Environment & Health 30(S2): 47-53.

61. Kashiwagi, N., 2004. Chemical mass balance when an unknown source exists. Environmetrics 15(8): 777-796.

62. Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Le Tertre, A., Monopolis, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., Anderson, H.R., Wojtyniak, B., Paldy, A., Braunstein, R., Pekkanen, J., Schindler, C., Schwartz, J., 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.

63. Katsouyanni, K., Touloumi, G., Spix, C., Schwartz, J., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharove, L., Schouten, J.P., Pönkä, A., Anderson, H.R., 1997. Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. British Medical J 314 (7095): 1658.

64. Koistinen, K., Edwards, R., Mathys, P., Ruuskanen, J., Künzli, N., Jantunen, M., 2004. Sources of PM2.5 in personal exposures and residential indoor, outdoor and workplace microenvironments in EXPOLIS-HELSINKI, FINLAND. Scandinavian Journal of Work, Environment & Health 30(Suppl 2): 36-46.

65. Koistinen, K.J., Hänninen, O.O., Rotko, T., Edwards, R.D., Moschandreas, D., Jantunen, M.J., 2001. Behavioral And Environmental Determinants Of Personal Exposures To PM2.5 In EXPOLIS-Helsinki, Finland. Atmospheric Environment 35(14): 2473-2481.

66. Koistinen, K.J., Kousa, A., Tenhola, V., Hänninen, O.O., Jantunen, M.J., Oglesby, L., Künzli, N., Georgoulis, L., 1999. Fine particle (PM2.5) measurement methodology, quality assurance procedures, and pilot results of the EXPOLIS study. Journal of the Air & Waste Management Association 49: 1212-1220.

67. Kousa, A., Kukkonen, J., Karppinen, A., Aarnio, P., Koskentalo, T., 2001a. Statistical and diagnostic evaluation of a new-generation urban dispersion modelling system against an extensive dataset in the Helsinki area. Atmospheric Environment 35 (27): 4617-4628.

68. Kousa, A., Kukkonen, J., Karppinen, A., Aarnio, P., Koskentalo, T., 2002a. A model for evaluating the population exposure to ambient air pollution in an urban area. Atmospheric Environment 36: 2109-2119.

69. Kousa, A., Monn, C., Rotko, T., Alm, S., Oglesby, L., Jantunen, M., 2001b. Personal exposures to NO2 in the Expolis-study: Relation to residential indoor, outdoor and workplace concentrations in Basel, Helsinki and Prague. Atmospheric Environment 35(20): 3405-3412.

70. Kousa, A., Oglesby, L., Koistinen, K., Kunzli, N., Jantunen, M., 2002b. Exposure chain of urban air PM2.5 - associations between ambient fixed site, residential outdoor, indoor, workplace and personal exposures in four European cities in the EXPOLIS-study. Atmospheric Environment 36: 3031-3039.

71. Koutrakis, P., Briggs, S.L.K., Leaderer, B.P., 1992. Source Apportionment of Indoor Aerosols in Suffolk and Onondaga Counties, New-York. Environmental Science & Technology 26(3): 521-527.

72. Kruize, H., Hänninen, O.O., Breugelmans, O., Lebret, E., Jantunen, M., 2003. Description and demonstration of the EXPOLIS simulation model: Two examples of modeling population exposure to particulate matter. Journal of Exposure Analysis and Environmental Epidemiology 13(2): 87-99.

73. Kukkonen, J., Härkönen, J., Karppinen, A., Pohjola, M., Pietarila, H., Koskentalo, T., 2001a. A semi-empirical model for urban PM10 concentrations, and its evaluation against data from an urban measurement network. Atmospheric Environment 35 (26): 4433-4442.

74. Kukkonen, J., Härkönen, J., Walden, J., Karppinen, A., Lusa, K., 2001b. Evaluation of the CAR-FMI model against measurements near a major road. Atmospheric Environment 35: 949-960.

75. Kukkonen, J., Valkonen, E., Walden, J., Koskentalo, T., Aarnio, P., Karppinen, A., Berkowicz, R., Kartastenpää, R., 2001c. A measurement campaign in a street canyon in Helsinki and comparison of results with predictions of the OSPM model. Atmospheric Environment 35: 231-243.

76. Law, A. M. and W. D. Kelton 1991 Simulation modeling & analysis.Singapore: McGraw-Hill.

77. Law, P.L., Lioy, P.J., Zelenka, M.P., Huber, A.H., McCurdy, T.R., 1997. Evaluation of a probabilistic exposure model applied to carbon monoxide (pNEM/CO) using Denver personal exposure monitoring data. Journal of the Air & Waste Management Association 47: 491-500.

78. Lebowitz, M.D., O'Rourke, M.K., Gordon, S., Moschandreas, D.J., Buckley, T., Nishioka, M., 1995. Population-based exposure measurements in Arizona: A phase I field study in support of the national

human exposure assessment survey. Journal of Exposure Analysis and Environmental Epidemiology 5: 297-325.

79. Lefohn, A.S., Husar, J.D., Husar, R.B., 1999. Estimating historical anthropogenic global sulfur emission patterns for the period 1850-1990. Atmospheric Environment 33(21): 3435-3444.

80. Leijnse, A., Hassanizaded, S.M., 1994. Model definition and model validation. Advances in Water Resources 17: 197-200.

81. Letz, R., Ryan, P.B., Spengler, J.D., 1984. Estimated distributions of personal exposure to respirable particles. Environmental Monitoring and Assessment 4: 351-359.

82. Lioy, P.J., 1990. Assessing total human exposure to contaminants. EST 24: 938-945.

83. Lioy, P.J., 1991. Assessing human exposure to airborne pollutants. EST 25: 1361-1362.

84. Lioy, P.J., Pellizzari, E., 1995. Conceptual framework for designing a national survey of human exposure. Journal of Exposure Analysis and Environmental Epidemiology 5: 425-444.

85. Lioy, P.J., Waldman, J.M., Buckley, T., Butler, J., Pietarinen, C., 1990. The personal, indoor and outdoor concentrations of PM10 measured in an industrial community during the winter. Atmospheric Environment 24B: 57-66.

86. Lunchick, C., 2001. Probabilistic exposure assessment of operator and residential non-dietary exposure. Annals of Occupational Hygiene 45(90001): 29S-42.

87. Macintosh, D.L., Kabiru, C., Echols, S.L., Ryan, P.B., 2001. Dietary exposure to chlorpyrifos and levels of 3,5,6-trichloro-2-pyridinol in urine. Journal of Exposure Analysis and Environmental Epidemiology 11(4): 279-285.

88. Mage, D.T., 2001. A procedure for use in estimating human exposure to particulate matter of ambient origin. Journal of the Air & Waste Management Association 51(1): 7-10.

89. Mathys, P., Stern, W.B., Oglesby, L., Braun-Fahrländer, C., Ackermann-Liebrich, U., Jantunen, M.J., Künzli, N., 2001. Elemental Analysis of Airborne Particulate Matter by ED-XRF within the European EXPOLIS Study. ICP Information Newsletter 27 (3): 29-34.

90. McCurdy, T., Glen, G., Smith, L., Lakkadi, Y., 2000. The National Exposure Research Laboratory's Consolidated Human Activity Database. Journal of Exposure Analysis and Environmental Epidemiology 10: 566-578.

91. Mekel, O.C.L., Fehr, R., 2001. Use of probabilistic methods in exposure assessment in Germany. Annals of Occupational Hygiene 45(90001): 65S-667.

92. Meng, Q.Y., Turpin, B.J., Korn, L., Weisel, C.P., Morandi, M., Colome, S., Zhang, J., Stock, T., Spektor, D., Winer, A., Zhang, L., Lee, J.H., Giovanetti, R., Cui, W., Kwon, J., Alimokhtari, S., Shendell, D., Jones, J., Farrar, C., Maberti, S., 2005. Infuence of ambient (outdoor) sources on residential indoor and personal PM2.5 concentrations: Analyses of RIOPA data. Journal of Exposure Analysis and Environmental Epidemiology 15: 17-28.

93. Mitchell, M., Campbell, C., 2001. Probabilistic exposure assessment of operator and residential exposure; a Canadian regulatory perspective. Annals of Occupational Hygiene 45(90001): 43S-447.

94. Morandi, M., Stock, T.H., Contant, C.F., 1988. A comparative study of respirable particulate microenvironmental concentrations and personal exposures. Environmental Monitoring and Assessment 10: 105-122.

95. Morgan, M. G. and M Henrion 1990 Uncertainty.New York: Cambridge University Press.

96. Mukala, K., Alm, S., Tiittanen, P., Salonen, R.O., Jantunen, M., Pekkanen, J., 2000. Nitrogen dioxide exposure assessment and cough among preschool children. Arch Env Health 55(6): 431-438.

97. Mukala, K., Pekkanen, J., Tiittanen, P., Alm, S., Salonen, R.O., Jantunen, M., Tuomisto, J., 1996. Seasonal exposure to NO2 and respiratory symptoms in preschool children. Journal of Exposure Analysis and Environmental Epidemiology 6 (2): 197-210.

98. Murray, D., 1997. Residential house and zone volumes in the United States: Empirical and estimated parametric distributions. Risk Analysis 17(4): 439-445.

99. Murray, D., Burmaster, D., 1995. Residential air exchange rates in the United States: Empirical and estimated parametric distributions by season and climatic region. Risk Analysis 15(4): 459-465.

100. Newton, I. S. 1687 Philosophiae naturalis principia mathematica.London: Royal Society.

101. O'Rourke, M.K., Rogan, S.P., Jin, S., Robertson, G.L., 1999a. Spatial distributions of arsenic exposure and mining communities from NHEXAS Arizona. Journal of Exposure Analysis and Environmental Epidemiology 9(5): 446-455.

102. O'Rourke, M.K., Van de, W.P.K., Jin, S., Rogan, S.P., Weiss, A.D., Gordon, S.M., Moschandreas, D.M., Lebowitz, M.D., 1999b. Evaluations of primary metals from NHEXAS Arizona: distributions and preliminary exposures. Journal of Exposure Analysis and Environmental Epidemiology 9(5): 435-445.

103. Oglesby, L., Künzli, N., Röösli, M., Braun-Fahrländer, C., Mathys, P., Stern, W., Jantunen, M., Kousa, A., 2000. Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution. Journal of the Air & Waste Management Association 50: 1251-1261.

104. Oreskes, N., Shraderfrechette, K., Belitz, K., 1994. Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences. Science 263(5147): 641-646.

105. Ott, W., Thomas, J., Mage, D., Wallace, L., 1988. Validation of the simulation of human activity and pollutant exposure (Shape) model using paired days from the Denver, CO, carbon monoxide study. Atmospheric Environment 22: 2101-2113.

106. Ott, W., Wallace, L., Mage, D., 2000. Predicting particulate (PM10) personal exposure distributions using a random component superposition statistical model. Journal of the Air & Waste Management Association 50: 1390-1406.

107. Ott, W.R., 1982. Concepts of human exposure to air pollution. Environmental International 7: 179-196.

108. Ott, W.R., 1984. Exposure estimates based on computer generated activity patterns. J Toxicol Clin Toxicol 21: 97-128.

109. Ott, W.R., 1985. Total human exposure-An emerging science focuses on humans as receptors of environmental pollution. EST 19: 880-886.

110. Ott, W.R., 1995. Human exposure assessment: the birth of a new science. Journal of Exposure Analysis and Environmental Epidemiology 5: 449-472.

111. Özkaynak, H., Xue, J., Spengler, J., Wallace, L., Pellizzari, E., Jenkins, P., 1996. Personal exposure to airborne particles and metals: results from the particle TEAM study in Riverside, California. Journal of Exposure Analysis and Environmental Epidemiology 6: 57-78.

112. Paatero, P., Hopke, P.K., 2003. Discarding or downweighting high-noise variables in factor analytic models. Analytica Chimica Acta 490(1-2): 277-289.

113. Pandian, M.D., Bradford, J., Behar, J.V.,1990. THERDBASE: Total Human Exposure Relational Database. In: TEAM: Total Exposure Assessment Methodology, A New Horizon.1990: Proceedings of the International Conference of Air & Waste Management Association and U.S.Environmental Protection Agency, Las Vegas, Nevada, November 1989.

114. Pang, Y.H., Macintosh, D.L., Camann, D.E., Ryan, B., 2002. Analysis of aggregate exposure to chlorpyrifos in the NHEXAS-Maryland investigation. Env Health Pers 110(3): 235-240.

115. Pekkanen, J., Peters, A., Hoek, G., Tiittanen, P., Brunekreef, B., de Hartog, J., Heinrich, J., Ibald-Mulli, A., Kreyling, W.G., Lanki, T., Timonen, K.L., Vanninen, E., 2002. Particulate air pollution and risk of ST-segment depression during repeated submaximal exercise tests among subjects with coronary heart disease - The exposure and risk assessment for fine and ultrafine particles in ambient air (ULTRA) study. Circulation 106(8): 933-938.

116. Pekkanen, J., Timonen, K.L., Ruuskanen, J., Reponen, A., Mirme, A., 1997. Effects of ultra-fine and fine particles in urban air on peak expiratory flow among children with asthmatic symptoms. Environmental Research 74: 24-33.

117. Pellizzari, E., Lioy, P., Quackenboss, J., Whitmore, R., Clayton, A., Freeman, N., Waldman, J., Thomas, K., Rodes, C., Wilcosky, T., 1995. Population-based exposure measurements in EPA region 5: a phase I field study in support of the National Human Exposure Assessment Survey. Journal of Exposure Analysis and Environmental Epidemiology 5(3): 327-358.

118. Pellizzari, E.D., Clayton, C.A., Rodes, C.E., Mason, R.E., Piper, L.L., Fort, B., Pfeifer, G., Lynam, D., 1999. Particulate matter and manganese exposures in Toronto, Canada. Atmospheric Environment 33: 721-734.

119. Pellizzari, E.D., Clayton, C.A., Rodes, C.E., Mason, R.E., Piper, L.L., Fort, B., Pfeifer, G., Lynam, D., 2001a. Particulate matter and manganese exposures in Indianapolis, Indiana. Journal of Exposure Analysis and Environmental Epidemiology 11(6): 230-440.

120. Pellizzari, E.D., Smith, D.J., Clayton, C.A., Michael, L.C., Quackenboss, J.J., 2001b. An assessment of the data quality for NHEXAS - Part I: exposure to metals and volatile organic chemicals in Region 5. Journal of Exposure Analysis and Environmental Epidemiology 11(2): 140-154.

121. Penttinen, P., Timonen, K., Tiittanen, P., Mirme, A., Ruuskanen, J., Pekkanen, J., 2001. Ultrafine particles in urban air and respiratory health among adult asthmatics. European Respiratory Journal 17: 428-435.

122. Phillips, K., Bentley, M.C., Howard, D.A., Alvan, G., 1996. Assessment of air quality in Stockholm by personal monitoring of nonsmokers for respirable suspended particles and environmental tobacco smoke. Scandinavian Journal of Work, Environment & Health 22: 1-24.

123. Phillips, K., Bentley, M.C., Howard, D.A., Alván, G., 1998a. Assessment of air quality in Paris by personal monitoring of nonsmokers for respirable suspended particles and environmental tobacco smoke. Environment International 24(4): 405-425.

124. Phillips, K., Bentley, M.C., Howard, D.A., Alván, G., Huici, A., 1997a. Assessment of air quality in Barcelona by personal monitoring of nonsmokers for respirable suspended particles and environmental tobacco smoke. Environment International 23: 173-196.

125. Phillips, K., Howard, D.A., Bentley, M.C., Alvan, G., 1997b. Assessment of air quality in Turin by personal monitoring of nonsmokers for respirable suspended particles and environmental tobacco smoke. Environment International 23: 851-871.

126. Phillips, K., Howard, D.A., Bentley, M.C., Alvan, G., 1999. Assessments of environmental tobacco smoke and respirable suspended particle exposures for nonsmokers in Basel by personal monitoring. Atmospheric Environment 33: 1889-1904.

127. Phillips, K., Howard, D.A., Bntley, M.C., Alvan, G., 1998b. Assessment of environmetal tobacco smoke and respirable suspended particle exposures for nonsmokers in Hong Kong using personal monitroing. Atmospheric Environment 24: 851-870.

128. Phillips, K., Howard, D.A., Browne, D., Lewsley, J.M., 1994. Assessment of personal exposures to environmental tobacco smoke in British nonsmokers. Environment International 30(6): 693-712.

129. Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. The Journal of the American Medical Association 287 (9): 1132-1141.

130. Pope, C.A., Thun, M.J., Namboodiri, M.M., Dockery, D.W., Evans, J.S., Speizer, F.E., Heath, C.W., 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Critical Care Med 151: 669-674.

131. Popper, K. R. 1935 Logik der forschung.Wien: Julius Springer Verlag.

132. Price, P.S., Young, J.S., Chaisson, C.F., 2001. Assessing aggregate and cumulative pesticide risks using a probabilistic model. Annals of Occupational Hygiene 45(90001): 131S-142.

133. Robertson, G.L., Lebowitz, M.D., O'Rourke, M.K., Gordon, S., Moscandreas, D., 1999. The National Human Exposure Assessment Survey (NHEXAS) study in Arizona -introduction and preliminary results. Journal of Exposure Analysis and Environmental Epidemiology 9(5): 427-434.

134. Robertson, G.L., Quackenboss, J.J., Hern, S.C., Thompson, C.V.,2001. The human exposure database system (HEDS) - putting the NHEXAS data on-line. ISEA 11th annual meeting, Charleston (SC), 2001: Nov 4-8, W-IF p.291.

135. Roemer, W., Hoek, G., Brunegreef, B., Haluszka, J., Kalandidi, A., Pekkanen, J., for, t.P.E.A., 1998. Daily variations in air pollution and respiratory health in a multicentre study: the PEACE project. European Respiratory Journal 12: 1354-1361.

136. Rotko, T., Koistinen, K., Hänninen, O.O., Jantunen, M., 2000a. Sociodemographic descriptors of personal exposure to fine particles (PM2.5) in EXPOLIS-Helsinki. Journal of Exposure Analysis and Environmental Epidemiology 10 (4): 385-393.

137. Rotko, T., Kousa, A., Alm, S., Jantunen, M., 2001. Exposures to Nitrogen Dioxide (NO2) in Expolis-Helsinki: Microenvironment, Behavioral and Sociodemographic Factors. Journal of Exposure Analysis and Environmental Epidemiology 11: 216-223.

138. Rotko, T., Oglesby, L., Künzli, N., Jantunen, M., 2000b. Population sampling in European air pollution exposure study, EXPOLIS: comparisons between the cities and representativity of the samples. Journal of Exposure Analysis and Environmental Epidemiology 10 (4): 355-364.

139. Rovers, M.M., van der Wilt, G.J., van der Bij, S., Straatman, H., Ingels, K., Zielhuis, G.A., 2005. Bayes' theorem: A negative example of a RCT on grommets in children with glue ear. European Journal of Epidemiology 20(1): 23-28.

140. Ruuskanen, J., Tuch, T., Ten Brink, H., Peters, A., Khlystov, A., Mirme, A., Kos, G.P.A., Brunekreef, B., Wichmann, H.E., Buzorius, G., Vallius, M., Kreyling, W.G., Pekkanen, J., 2001. Concentrations of ultrafine, fine and PM2.5 particles in three European cities. Atmospheric Environment 35(21): 3729-3738.

141. Ryan, P.B., 1991. An overview of human exposure modeling. Journal of Exposure Analysis and Environmental Epidemiology 1(4): 453-474.

142. Ryan, P.B., Spengler, J.D., Letz, R., 1986. Estimating personal exposures to NO2. Environment International 12: 395-400.

143. Samoli, E., Analitis, A., Touloumi, G., Schwartz, J., Anderson, H.R., Sunyer, J., Bisanti, L., Zmirou, D., Vonk, J.M., Pekkanen, J., Goodman, P., Paldy, A., Schindler, C., Katsouyanni, K., 2005. Estimating the Exposure-Response Relationships between Particulate Matter and Mortality within the APHEA Multicity Project. Env Health Pers 113(1): 88-95.

144. Seifert, B., 1995. Validity criteria for exposure assessment methods. Sci Tot Env 168: 101-107.

145. Seifert, B., Becker, K., Helm, D., Krause, C., Schulz, C., Seiwert, M., 2000a. The German Environmental Survey 1990/1992 (GerES II): reference concentrations of selected environmental pollutants in blood, urine, hair, house dust, drinking water and indoor air. Journal of Exposure Analysis and Environmental Epidemiology 10(6): 552-565.

146. Seifert, B., Becker, K., Hoffmann, K., Krause, C., Schulz, C., 2000b. The German Environmental Survey 1990/1992 (GerES II): a representative population study. Journal of Exposure Analysis and Environmental Epidemiology 10 (2): 103-114.

147. Sexton, K., Callahan, M.A., Bryan, E.F., Saint, C.G., Wood, W.P., 1995a. Informed decisions about protecting and promoting public health: rationale for a national human exposure assessment survey. Journal of Exposure Analysis and Environmental Epidemiology 5: 233-256.

148. Sexton, K., Kleffman, D.E., Callahan, M.A., 1995b. An introduction to the national human exposure assessment survey (NHEXAS) and related phase I field studies. Journal of Exposure Analysis and Environmental Epidemiology 5(3): 229-232.

149. Sexton, K., Spengler, J.D., Treitman, D., 1984. Personal exposure to respirable particles: a case study in Waterbury, Vermont. Atmospheric Environment 18: 1385-1398.

150. Sexton, K., Wagener, D., Selevan, S., Miller, T., Lybarger, J., 1994. An inventory of human exposure-related data bases. Journal of Exposure Analysis and Environmental Epidemiology 4(1): 95-109.

151. Sexton, K.S., Selovan, G., Wagner, D.K., Lybarger, J.A., 1992. Estimating human exposures to environmental pollutants: availability and utility of existing databases. Arch Env Health 47: 398-407.

152. Small, M.1990 Probability distributions and statistical estimation. *In* Uncertainty. M. G. Morgan and M Henrion, eds. New York: Cambridge University Press.

153. Spengler, J.D., Treitman, R.D., Tosteson, T.D., Mage, D.T., Soczek, M.L., 1985. Personal exposures to respirable particulates and implications for air pollutionepidemiology. EST 19: 700-707.

154. Spengler, J.D., Soczek, M.L., 1984. Evidence for improved ambient air quality and the need for personal exposure research. EST 18: 268A-280A.

155. Suh, H.H., Koutrakis, P., Spengler, J.D., 1993. Validation of Personal Exposure Models for Sulfate and Aerosol Strong Acidity. Journal of the Air & Waste Management Association 43: 845-850.

156. Tainio, M., Tuomisto, J.T., Hänninen, O.O., Aarnio, P., Koistinen, K., Jantunen, M., Pekkanen, J., 2005. Health effects caused by primary fine particulate matter (PM2.5) emitted from buses in the Helsinki metropolitan area, Finland. Risk Analysis 25(1): 149-158.

157. Taylor, A.C., 1993. Using Objective and Subjective Information to Develop Distributions for Probabilistic Exposure Assessment. Journal of Exposure Analysis and Environmental Epidemiology 3: 285-298.

158. Thomas, K.W., Pellizzari, E.D., Clayton, C.A., Whitaker, D.A., Shores, R.C., Spengler, J., Ozkaynak, H., Froehlich, S.E., Wallace, L.A., 1993. Particle Total Exposure Assessment Methodology (PTEAM) 1990 study: method performance and data quality for personal, indoor, and outdoor monitoring. Journal of Exposure Analysis and Environmental Epidemiology 3(2): 203-226.

159. Vallius, M., Lanki, T., Tiittanen, P., Koistinen, K., Ruuskanen, J., Pekkanen, J., 2003. Source apportionment of urban ambient PM2.5 in two successive measurement campaigns in Helsinki, Finland. Atmospheric Environment 37(5): 615-623.

160. Wallace, L., 1993. A decade of studies of human exposure:what have we learned? Risk Analysis 13: 135-139.

161. Wallace, L., 1996. Indoor particles: a review. Journal of the Air & Waste Management Association 46: 98-126.

162. Weisel, C.P., Zhang, J., Turpin, B.J., Morandi, M., Colome, S., Stock, T., Spektor, D., Korn, L., Winer, A., Alimokhtari, S., Kwon, J., Mohan, K., Harrington, R., Giovanetti, R., Cui, W., Afshar, M., Maberti, S., Shendell, D., 2005. Relationship of Indoor, Outdoor and Personal Air (RIOPA) study: study design, methods and quality assurance/control results. Journal of Exposure Analysis and Environmental Epidemiology advance online June 2004.

163. WHO,2000. Air Quality Guidelines for Europe. 2. WHO Regional Publications, European Series No. 91.

164. WHO,2002. The World health report : 2002 : Reducing risks, promoting healthy life. Geneva, World Health Organization.

165. Wikle, C.K., Berliner, L.M., 2005. Combining information across spatial scales. Technometrics 47(1): 80-91.

166. Wilson, W.E., Mage, D.T., Grant, L.D., 2000. Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: Why and how. Journal of the Air & Waste Management Association 50: 1167-1183.

167. Yeh, S., Small, M.J., 2002. Incorporating exposure models in probabilistic assessment of the risks of premature mortality from particulate matter. Journal of Exposure Analysis and Environmental Epidemiology 12(6): 389-403.

168. Yli-Tuomi, T., Hopke, P.K., Paatero, P., Basunia, M.S., Landsberger, S., Viisanen, Y., Paatero, J., 2003a. Atmospheric aerosol over Finnish Arctic: source analysis by the multilinear engine and the potential source contribution function. Atmospheric Environment 37(31): 4381-4392.

169. Yli-Tuomi, T., Venditte, L., Hopke, P.K., Basunia, M.S., Landsberger, S., Viisanen, Y., Paatero, J., 2003b. Composition of the Finnish Arctic aerosol: collection and analysis of historic filter samples. Atmospheric Environment 37(17): 2355-2364.

170. Zhang, X.L., Shu, X.O., Yang, G., Li, H.L., Xiang, Y.B., Gao, Y.T., Li, Q., Zheng, W., 2005. Association of passive smoking by husbands with prevalence of stroke among Chinese women nonsmokers. American Journal of Epidemiology 161(3): 213-218.

I

Jantunen, M.J., Hänninen, O.O., Katsouyanni, K., Knöppel, H., Künzli, N., Lebret, E., Maroni, M., Saarela, K., Srám, R., Zmirou, D., 1998. **Air pollution exposure in European cities: The *EXPOLIS*-study**. *Journal of Exposure Analysis and Environmental Epidemiology* 8 (4): 495-518.

AIR POLLUTION EXPOSURE IN EUROPEAN CITIES: THE "*EXPOLIS*" STUDY

MATTI J. JANTUNEN,[*] OTTO HÄNNINEN,[*] KLEA KATSOUYANNI,[†] HELMUT KNÖPPEL,[‡] NINO KUENZLI,[§] ERIK LEBRET,[¶] MARCO MARONI,[**] KRISTINA SAARELA,[††] RADIM SRÁM,[‡‡] AND DENIS ZMIROU,[§§]

[*]KTL
Department of Environmental Hygiene
Kuopio, Finland

[†]University of Athens
Medical School
Athens, Greece

[‡]EC JRC
Institute of the Environment
Ispra, Italy

[§]University of Basel
Institute of Social and Preventive Medicine
Basel, Switzerland

[¶]RIVM
Department of Chronic Diseases and Environmental Epidemiology
Bilthoven, The Netherlands

[**]University of Milan
Institute of Occupational Health
Milan, Italy

[††]VTT, Chemical Technology
Espoo, Finland

[‡‡]Reg. Institute of Hygiene of Central Bohemia
Lab. of Genetic Ecotoxicology,
Prague, The Czech Republic

[§§]University Joseph Fourier
Medical School
Grenoble, France

1. Address all correspondence to: Matti J. Jantunen, KTL— Division of Environmental Health, P.O. Box 95/FIN-70701 Kuopio, FINLAND. Tel.:+358 400 587 816. Fax:+358 17 201 265. E-mail: matti.jantunen@ktl.fi.

2. Abbreviations: CO, carbon monoxide; exposure sample, subsample for exposure and microenvironmental monitoring plus TMAD and questionnaire application (direct exposure monitoring sample), Diary sample, subsample for TMAD and questionnaire application without exposure or microenvironmental monitoring (indirect exposure assessment sample), KTL, (Finnish) National Public Health Institute; ME, microenvironment; MEM, microenvironmental monitor; NO_2, nitrogen dioxide; PEM, personal exposure monitor; PM, particulate matter; $PM_{2.5}$ (PM_{10}), particles smaller than 2.5 (10) μm in aerodynamic diameter; QA, quality assurance; QC, quality control; RIVM, (Dutch) National Public Health Institute; SOP, standard operating procedure; TMAD, time-microenvironment-activity diary; VOC, volatile organic compound; VTT, (Finnish) State Technical research Centre.

3. Key words: air pollution, CO, European cities, exposure determinants, NO_2, $PM_{2.5}$, population exposure, VOC.

1053-4245/98 $12.00

Epidemiological literature of the 1990's has revealed surprisingly large public health impacts associated with present common air pollution levels in North American and European cities. Any causal explanation of the health effects of air pollutants must go through exposure, yet no large, population based air pollution exposure studies have been conducted in Europe, and consequently no European database of air pollution exposures of urban populations has existed until now. EXPOLIS *is a European multicenter study for measurement of air pollution exposures of working age urban populations during workdays. The selected urban areas are Athens, Basel, Grenoble, Helsinki, Milan and Prague. The main objectives of* EXPOLIS *are:*

- *To assess the exposures of European urban populations to major air pollutants.*
- *To analyse the personal and environmental determinants and interrelationships to these exposures.*
- *To develop an European database for simulation of air pollution exposures.*

These objectives were pursued by measuring the personal exposures, home indoor and outdoor and workplace levels of particles smaller than 2.5 µm in aerodynamic diameter ($PM_{2.5}$), volatile organic compounds (VOCs) and carbon monoxide (CO) of approximately 500 subjects representing the adult populations of the selected cities.

The field work continued from fall of 1996 to winter of 1997–1998. Identical sampling equipment, operating procedures, time-microenvironment-activity diaries, questionnaires, database and data entry tools were used in each center. To assure comparability of the data from the 6 cities in 6 countries, a strict quality assurance/ quality control (QA/QC) protocol was established and the field work was supervised by the QA Unit of KTL. Standard operating procedures SOPs were prepared for all subject, laboratory and field procedures, and the EXPOLIS *field teams were trained in three joint workshops. VOC laboratory analyses were intercalibrated by the European Commission/Joint Research Centre (EC/JRC) Environment Institute in Ispra. Other techniques were intercalibrated between the teams.*

*This paper describes the main design features of the European Union 4th Framework RTD Programme funded multicenter study; Air Pollution Exposure Distributions of Adult Urban Populations in Europe (*EXPOLIS*). The* EXPOLIS *Centers are KTL- (coordinating center) in Helsinki, University of Athens, University of Basel, University Joseph Fourier in Grenoble, University of Milan, Regional Institute of Hygiene of Central Bohemia in Prague, VTT (Finnish State Technical Research Center) in Helsinki, and RIVM (Dutch National Public Health Institute) in Bilthoven. More detailed descriptions of the materials, methods and results of this large, multicenter and multidimensional study will be published later in more focused articles.*

INTRODUCTION

Recent investigations by American epidemiologists, Dockery et al. (1992, 1993 and 1994), Schwartz et al. (1992) and Pope et al. (1995), reanalysis of the Six-Cities-Study data by the Health Effects Institute (HEI Oversight Committee, 1995), and the European APHEA project (Katsouyanni et al., 1995; Dab et al., 1996; Katsouyanni et al., 1996; Ponce de Leon et al., 1996; Pönkä and Virtanen, 1996; Schouten et al., 1996; Schwartz et al., 1996; Spix and Wichmann, 1996; Sunyer et al., 1996; Touloumi et al., 1996; Vigotti et al., 1996; Zmirou et al., 1996) have radically changed our understanding of the health effects of air pollutants. Ten years ago, most experts would have agreed that severe health effects of the present air pollution levels in North America and Western Europe are rare. We now estimate that differences of air pollution levels, especially fine particulate matter (PM), in time and space are associated with tens of thousands of cases of respiratory and cardiovascular mortality in Europe annually, and significant reduction in the length of life of large populations (WHO, 1995).

However, although the above mentioned time-series and cohort studies are based on ambient air data from urban air quality monitoring networks, the harmful health effects of urban air pollutants may not be caused by the levels of air pollutants at those fixed monitoring sites alone, but instead by the personal exposures of the millions of individuals in their daily activities in indoor and outdoor urban environments and in commuting between them.

A number of air pollution studies where personal exposures have been monitored have been done, but rather few on representative population samples. Table 1 introduces the main design features of such already published exposure studies. Most personal exposure studies have been done on nitrogen dioxin (NO_2), because it is a significant air pollutant, has both outdoor and indoor sources, and can be easily monitored with cheap passive samplers (Hoek et al., 1984; Fischer et al., 1986; Quackenboss et al., 1986; Ryan et al., 1989; Özkaynak et al., 1993; Song et al., 1993; Xue et al., 1993; Spengler et al., 1994; and Alm et al., 1998). Personal exposures to ozone have been studied in two small scale studies in Switzerland and the Netherlands (Monn et al., 1993; Fischer et al., 1993). The Washington-Denver CO (carbon monoxide) study covered one pollutant and two cities (Ackland et al., 1985; Jungers et al., 1985; Ott et al., 1988; Wallace et al., 1989; Mage et al., 1989). VOC exposures have been studied in one population based study in California (Hartwell et al., 1987), and in another large indoor air and exposure study in Germany (Hoffmann et al., 1996). Nicotine as an indicator of passive tobacco smoke exposure has been monitored with passive personal samplers on a random sample of American nonsmoking women (O'Connor et al., 1993). Lioy et al. (1990) were the first to collect personal PM_{10} exposure samples. The Particle-TEAM study collected both PM_{10} and nicotine exposures of residents of Riverside, California (Wallace et al., 1993; Thomas et al., 1993; Clayton et al., 1993; Özkaynak et al., 1996). Personal exposures to PAH were studied by Waldman et al. (1991) and both PAH and organic mutagens were analyzed in the Czech-USEPA health study in the Teplice area (Watts et al., 1994). Reported multicomponent exposure studies are few, the most prominent one being The National Human Exposure Assessment Survey (NHEXAS) (Lioy et al., 1995; Sexton et al., 1995). The LIILA study in Helsinki is the only one with personal exposure sampling of preschool children, and multicomponent gaseous (CO and NO_2) exposures (Alm et al., 1994; Alm et al.,

TABLE 1. Key Design Features of Selected Population Exposure Studies

Pollutant	Exposure time frame	Sampling time(s)	Sources of data	Target population	Sampling frame	Population sample	Reference
CO	short-term	continuous	PEM SAM TAD	nonsmoking residents (18-70 y) Washington DC and Denver CO	stratified probability sample	712+808	Ackland *et al.* 1985 Jungers *et al.* 1985
NO2	long-term	7 d 7 d 7 d 7 d	PPS PMSS POSS TAD MoPE	families of Portage WI with school aged children	stratified cluster sample	350	Quackenboss *et al.* 1986
VOC	full-year	2 x 12 h	PAS BME EI BI	population of California over 7y	stratified probability sample	188	Hartwell *et al.* 1987
CO	long-term	2 x 1 d	PEM SAM SPE	urban non-smoking population of Denver, CO	random sample	336	Ott *et al.* 1988
CO	short-term	continuous	PEM SAM BME	non-smoking populations of Denver CO and Washington DC	stratified probability sample	454+625	Wallace et el. 1988
CO	short-term	continuous	PEM SAM BME TAD SPE	non-smoking populations of Denver CO and Washington DC	random sample	555	Mage *et al.* 1989
NO2	short-term	2 x 0-24 h	PISS PPS TAD	population of Boston, MA	stratified probability sample	313	Ryan *et al.* 1989
PM10	short-term	14 x 24 h	PAS AMSS AOSS REQ	14 non-smoking adult individuals in Phillipsburg NJ	selected sample	14	Lioy *et al.* 1990
CO NO2	full-year	20 h cont. 1 wk (*13)	PEM PPS SAM REQ HD TQ	day care center children in Helsinki	all children in 8 day care centers	250	Alm *et al.* 1994 Alm *et al.* 1997
NO2	full-year	8 h 24 h (?)	PPS SPE TAD REQ	people in gas range homes in Los Angeles, CA	see Spengler *et al.* 1992	400	Özkaynak *et al.* 1993

TABLE 1. Key Design Features of Selected Population Exposure Studies (cont'd)

Pollutant	exposure time frame	sampling time(s)	exposure data	target population	sampling frame	population sample	reference
NO2	full-year	24 h -"- -"- -"-	PPS PMSS POSS TAD REQ	30-60 year old housewives in Beijing	cluster random sampling	59	Song *et al.* 1993
NO2	full-year	48 h 48 h&2wk 48 h&2wk 48 h	PPS POSS PMSS TAD REQ MePE MoPE MIAQ	residents of Los Angeles, CA	random representation sample	700	Xue *et al.* 1993
ETS Nicotine	cross-sect	7 d	EQ PPS BME	(self reported) non-smoking pregnant women in the U.S.	random sample	415	O'Connor *et al.* 1993
PM10 PM 2.5 Nicotine	48 d	2 x 12 h 2 x 12 h 2 x 12 h 48 x 1 d 2 x 12 h 2 x 12 h	PAS POSS PISS SAM TAD EQ	Non-smoking resid. of Riverside, CA	stratified probability sample	178	Wallace *et al.*1993 Thomas *et al.* 1993 Clayton *et al.* 1993 Özkaynak *et al.* 1996
NO2	short-term	48 h	PPS PISS POSS TAD	population of the Los Angeles Basin	population representative sample	682	Spengler *et al.* 1994

Abreviations used

AMSS = active microenvironmental stationary sampler
AOSS = active outdoor stationary sampler
ASD = active stationary sampler
ASD = active sampling device
BI = behavior interview
BME = biomarker of exposure
EI = exposure interview
EQ = exposure questionnaire
HD = health diary
HSI = health status interview
MAQM = mobile air quality monitor
MEM = microenvironmental monitor
MePE = measurement of personal exposure
MoIAQ = modeling of indoor air quality
MoMAQ = modeling of microenvironmental air quality
MoOAQ = modeling of outdoor air quality
MoPE = modeling of personal exposures
MoTPE = modeling of total personal exposure
OEI = occupational exposure interview
OEQ = occupational exposure questionnaire
PAS = personal active sampler
PEM = personal exposure monitor
PISS = passive indoor stationary sampler
PMSS = passive microenvironm stationary sampler
POSS = passive outdoor stationary sampler
PPS = personal passive sampler
PSD = passive sampling device
PSS = passive stationary sampler
REQ = residential environmental questionnaire
SAM = stationary ambient air monitor
SEI = socioeconomic interview
SIM = stationary indoor air monitor
SPE = simulation of population exposure
TAD = time activity diary
TI = transportation interview
TQ = transportation questionnaire

1998). In addition there have been a few studies where personal exposures to multiple air pollutants have been monitored in traffic situations (Bevan et al., 1991; Wijnen et al., 1995).

Most of these data are American, or collected from nonrepresentative and often small numbers of subjects. Clearly missing are European representative and comparable air pollution exposure data, which could be used to assess air pollution exposure distributions in populations, to search for the factors that are associated with high exposures or to evaluate exposure distributions within specific subpopulations.

In this paper we describe the general goals, design and an overview of the materials and methodologies of the *EXPOLIS* study. No results are presented here. More detailed descriptions of the materials, methods and results of this large, multicenter and multidimensional study will be published later in more focused articles.

SCOPE AND OBJECTIVES OF *EXPOLIS*

The EXPOLIS (Air Pollution Exposure Distributions within Adult Urban Populations in Europe) study focuses on working age urban populations in Europe, exposed to air pollutants in their homes, workplaces and other common urban microenvironments (streets, shopping, etc), and commuting between them during the workdays of the week. The urban areas selected for *EXPOLIS* study are Athens, Basel, Grenoble, Helsinki, Milan and Prague, to represent different European regions, city sizes and air pollution situations.

For selected urban air pollutants *EXPOLIS* directly determines:

- exposure distributions of target populations
- concentration distributions of the most important microenvironments
- time-activity distributions of target populations

The *EXPOLIS* data are stored in an international database for further exposure assessments and simulations. With this database, using statistical methods the following will be analyzed:

- statistical associations and logical links between exposures to different air pollutants
- the contributions of different air pollution sources to air pollution exposures
- the relationships of geographic, housing, occupation and commuting related, behavioral and socioeconomic factors to air pollution exposures

Using the *EXPOLIS* database, a probabilistic simulation model will be developed to assess the population exposure distributions

- of selected subpopulations,
- of selected urban areas and, or
- for selected future scenarios.

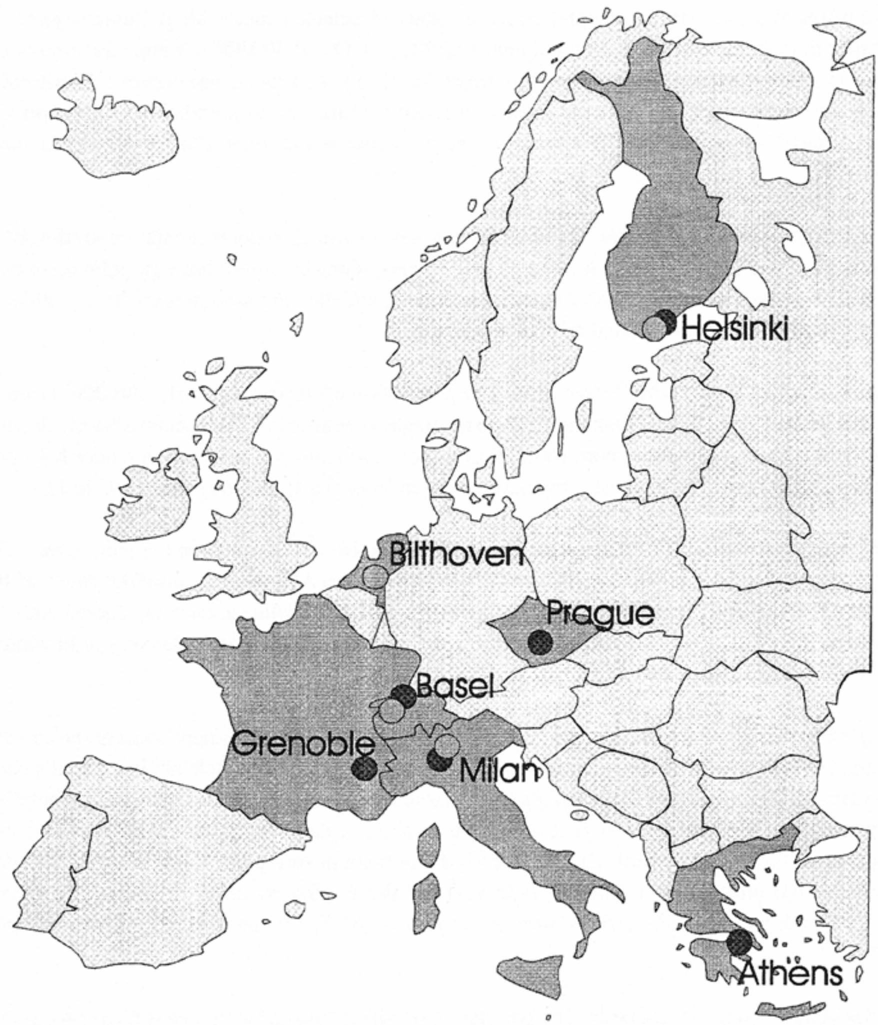

FIGURE 1. The study locations. The dark dots mark the exposure measurement cities. The light dots mark the sample analysis (VTT, Espoo, Finland, Carbotech, Basel, Switzerland, EC/JRC Ispra, Italy) and data analysis (RIVM, Bilthoven, The Netherlands) locations.

METHODS

Overall Structure of EXPOLIS

Exposures and microenvironmental concentrations of selected major air pollutants, particles smaller than 2.5 μm in aerodynamic diameter ($PM_{2.5}$), CO and 30 VOCs, were measured in six European cities: Athens, Basel, Grenoble, Helsinki, Milan and Prague, see Figure 1. These cities were selected to represent different European regions, climates and populations. Selection was also dictated by the presence of a research facility capable and willing to carry out this study protocol.

Athens is the capital and largest city of Greece. It lies on a small plain that extends southward to the Aegean Sea. The city center is 11 km from the coast. Greater Athens has a population of over 3 million. Athens has a typical Mediterranean climate, with hot, dry summers (> 25 °C) and mild winters (10 °C). Average annual rainfall is 400 mm.

Basel is located in northern Switzerland. The population of the city is nearly 400,000. Located on the Rhine River, Basel is a major industrial center (chemicals, pharmaceuticals, machinery, and textiles) and commercial port. Rainfall averages 1,000 mm per year, and in winter fog often enshrouds the area. The average temperature ranges from 0.6°C in January to 18°C in July.

Grenoble is located in the French Alps about 217 km north of Marseille on the Isere River. The city's population is over 400,000. Hydroelectric power from Alpine rivers provides much of the energy for the production of electrical machinery, electrometallurgy, cement, chemicals, and plastics. The climate is characterized by warm, dry summers (20 °C) and relatively mild winters (0 °C), with an average annual rainfall of 1000 mm.

Helsinki is the capital and largest city of Finland. It is located on the southern coast of the country on the Gulf of Finland. The population of Helsinki is about 1 million. Helsinki is Finland's chief port and handles more than half of all its foreign trade. Engineering, electronics and shipbuilding industries and food and timber processing are important. Climate exhibits both maritime and continental influences. Surrounding seas cool the climate in spring but warm it in fall. Rainfall averages 700 mm per year. The sea is frozen and the ground covered with snow for several months each winter. Mean temperature in January is -6 °C, but the summer months are mild (17°C).

Milan is the capital of Lombardy. The population of Milan metropolitan area is nearly 4 million. Milan is located in the basin of the Po River about 480 km northwest of Rome. Most industrial development has taken place in Milan's suburbs, far from the central city. Milan is Italy's chief commercial, financial, and industrial center manufacturing steel, textiles (particularly silk), clothing, machine tools, aircraft, automobiles, railroad equipment, agricultural machinery, chemicals, printed materials, pharmaceuticals, furniture, and foodstuffs. Milan has a continental climate. Seasonal temperatures average 24 °C for July and 5 °C for January.

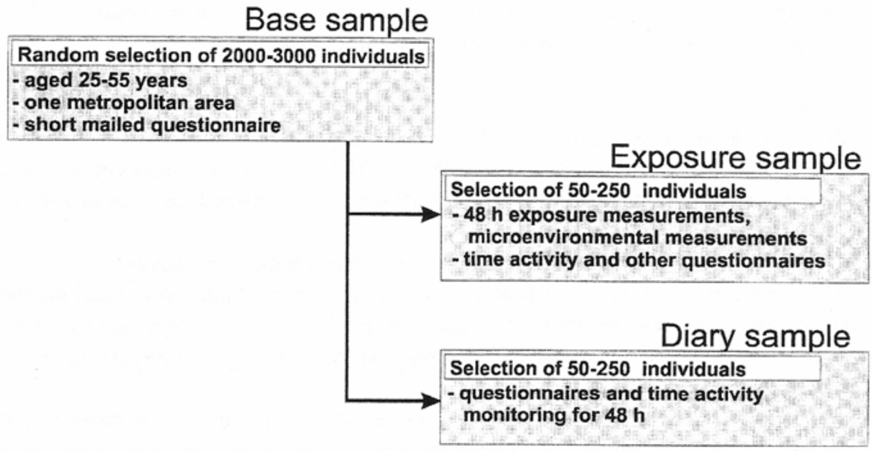

FIGURE 2. Population sampling scheme in each center (with the exception of Grenoble, see text).

TABLE 2. **Target Sizes for the Base Population Samples. *Exposure* Samples in Each *EXPOLIS* Center**

EXPOLIS ceneter	Primary sample	*Exposure* sample	*Diary* sample
Athens	2.000	50	50
Basel	3.000	50	250
Grenoble	-	40	40
Helsinki	2.523	240	240
Milan	3.000	50	250
Prague	-	50	50
Total	10.523	480	880

Prague is the capital and largest city of the Czech Republic. Its population is 1,200,000. The city is situated along both banks of the Vltava River. It is an industrial city, producing goods ranging from machinery, rolling stock, and chemicals to textiles, furniture, foodstuffs, and beer. Winters are generally cold, with many days of subfreezing weather (January average -1 °C). Summers are moderately warm with average July temperature about 19°C.

In each city, a population sample of 25–55 year old persons was formed and subsamples for exposure measurements and questionnaire applications were drawn. Population sample sizes are summarized in Figure 2 and Table 2.

Identical time activity and background questionnaires were used for *Exposure* sample and Diary sample. Personal exposures as well as the most important microenvironmental concentrations were measured for the *Exposure* sample. The microenvironments investigated were home indoors, home outdoors and main work place.

In each city, also a selected group of public microenvironments were measured. These microenvironments include shops, restaurants and public transport.

Air Pollutants

In each center, the personal exposures and personal microenvironmental concentrations were measured for $PM_{2.5}$, CO and 30 VOCs (see Table 3). The major air pollutants common to all cities were selected based on their health effects and their environmental concerns as follows:

- CO to represent exposure to traffic exhausts and indoor combustion sources,
- VOCs (see Table 3) because of health and welfare concerns both indoors and outdoors (carcinogenic, odorous and irritating compounds, precursors for tropospheric O_3), because many VOCs are useful source markers, and because the presently available data are of very variable quality, and
- $PM_{2.5}$ because inhalable particles are presently the air pollutants of greatest health concern and interest, and because no $PM_{2.5}$ exposure studies on representative population samples have been reported so far.

In addition 48 h NO_2 samples were collected from Basel, Helsinki and Prague. In Helsinki also carbonyl compounds and air exchange rate, and in Milan aldehydes were measured. Table 4 summarizes the air pollutants, microenvironments and measurement techniques of *EXPOLIS*.

Microenvironments and Activities

For the purpose of the study, a microenvironment (ME) is a location where the air pollutant concentrations at any given time can be considered relatively homogenous. For population exposure distribution simulations, all individual microenvironments that fall into the same category are grouped and processed as one microenvironment, and the concentrations measured or modeled for this microenvironment are presented in the form of a frequency distribution. In air pollution exposure modeling and simulation, concentration information is needed from the microenvironments contributing significantly to the population exposure. The microenvironments selected for *the EXPOLIS* time activity diaries were *home indoors, home outdoors, workplace indoors, workplace outdoors, other outdoor* and *other indoor,* and *traffic* (with subcategories), see Figure 3.

Work environments differ more than home environments from the viewpoint of exposure to air pollution. Public services, shops, offices, industrial work, transportation all have different characteristics. Heavy occupational exposures are excluded from the analysis, because they are too uncommon to be adequately represented in our population samples.

The microenvironments/activities, about which information was separately collected, were transportation (with subcategories), supermarkets and restaurants that may expose individuals to elevated levels of $PM_{2.5}$, VOCs or CO. Microenvironmental concentrations in traffic - inside automobiles, busses, trams, trains, metros and while walking or biking - were measured separately during the most active traffic hours. Microenvironmental levels in supermarkets and restaurants were measured during their active opening hours. Exposures related to specific activities were measured by the field team members.

TABLE 3. EXPOLIS VOC Target Compound List Based on Health and Irritation Concerns, and Environmental Significance

VOC	CAS-number	Mucous irritant	Airway hypersens initiator(*	Skin (* contact allergen	IARC carcinogen	USA 1990 CAAA HAP (**
Alkanes						
nonane	111-84-2					
decane	124-18-5					
undecane	1120-21-4					
cyclohexane	110-82-7					
Aromatics						
benzene	71-43-2				I	x
toluene	108-88-3					x
ethylbenzene	100-41-4					x
m&p-xylene	108-38-3					x
o-xylene	95-47-6					x
styrene	100-42-5	Yes	II B		II B	x
naphtalene	91-20-3					x
propylbenzene	103-65-1					
trimethylbenzenes	95-63-6					
Alcohols						
2-methyl-1-propanol	78-83-1					
1-butanol	71-42-0					
2-ethylhexanol	104-76-7	Yes				
phenol	108-95-2	Yes				x
1-octanol	111-87-5					
Esters						
2-buthoxyethanol	111-76-2	Yes		III		
Alkanals						
hexanal	66-25-1	Yes				
benzaldehyde	100-52-7	Yes				
octanal	124-13-0					
Halogenated						
trichloroethene	79-01-6					x
tetrachloroethene	127-18-4					x
1,1,2-trichloroethane	79-00-5					x
Miscellaneous						
d-limonene	138-86-3			II B		
1-methyl-2-pyrrolidinone	872-50-4					
3-carene	13466-78-9					
alpha-pinene	80-56-8					

*) NKB (1994), **) Clean Air Act Amendments of 1990 (1990)

Date:			LOCATION											ACTIVITIES		
Time		Briefly Describe Activity	IN TRANSFER					NOT IN TRANSFER						COOK-ING	SMOKING	
			walk bike	motor-cycle	car taxi	bus tram	metro train	home in	home out	work in	work out	other in	other out		self	same room
8	0		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	15		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	30		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	45		o	o	o	o	o	o	o	o	o	o	o	o	o	o
9	0		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	15		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	30		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	45		o	o	o	o	o	o	o	o	o	o	o	o	o	o
10	0		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	15		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	30		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	45		o	o	o	o	o	o	o	o	o	o	o	o	o	o
11	0		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	15		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	30		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	45		o	o	o	o	o	o	o	o	o	o	o	o	o	o
12	0		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	15		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	30		o	o	o	o	o	o	o	o	o	o	o	o	o	o
	45		o	o	o	o	o	o	o	o	o	o	o	o	o	o

FIGURE 3. The first hours of the 48 h time-microenvironment-activity diary (TMAD) for marking each 15 min of the day at the appropriate microenvironment-activity category/ies. Multiple entires are accepted for each 15 minutes.

TABLE 4. **Summary of the Meaured Compounds, Microenvironments, and Measurement Techniques**

Measurement	$PM_{2.5}$	VOC	CO	NO_2 (*
Personal exposures	gravimetric	active integrated	continuous	passive
Subject's microenvironments				
Home indoors	gravimetric	active integrated		passive
Home outdoors	gravimetric	active integrated		passive
Main workplace	gravimetric	active		passive
Public microenvironments	gravimetric+ optical continuous	active integrated	continuous	

*) Additional measurements in some centers only.

Population Samples

The target populations of this study are the adult, urban populations of Europe. *EXPOLIS* focuses on 25 to 55 year old individuals, because their exposures are most affected by urban traffic planning, zoning and occupational conditions. WHO (1991) estimates that a probability sample of a minimum of 50 subjects are needed for the sample to represent any target population. Larger samples are needed if the target population is divided into subpopulations for quantitative estimation of how the exposures relate to, e.g., home location, indoor sources, commuting, work, and socioeconomic parameters.

Too small subsamples produce poor estimates about exposure frequency distributions in the respective subpopulations. On the other hand, ensuring that all interesting subpopulations would have at least 50 representatives in our probability samples in each of the six *EXPOLIS* cities would result in a prohibitively expensive study.

While personal exposure and microenvironmental sampling/monitoring is laborious, questionnaire and time-microenvironment-activity diary (TMAD) application is much simpler. These two methods for acquiring personal exposure information can be combined by sample pooling; drawing one subsample for exposure and microenvironmental monitoring plus TMAD and questionnaire application (direct exposure monitoring sample or *Exposure* sample for short), and another subsample for TMAD and questionnaire application without exposure or microenvironmental monitoring (indirect exposure assessment sample or *Diary* sample for short). As the unit costs of the *Diary* sample are much lower than those of the *Exposure* sample, a pooled sample may give a smaller variance for population exposure estimate than an *Exposure* sample for the same investment, and the division between the two subsamples can be optimized (Duan and Mage, 1993). Because numerous different pollutants with different costs and presumably different correlations between modeled and monitored exposures were sampled in *EXPOLIS*, no one optimum could be determined for the division between the *Exposure* and *Diary* subsamples.

EXPOLIS includes a large population in only one city (Helsinki) whereas the other centers had a smaller *Exposure* sample to participate in the full assessment, and a larger *Diary* sample to contribute time- microenvironment-activity diary and questionnaire data only. Thus in one center, Helsinki, the aim was to estimate both population exposure distributions and exposure differences between different subpopulations as well as the relative roles of different determinants of exposure, and 240 subjects were drawn for the *Exposure* sample. In the other centers, the aim was to estimate population exposure levels and distributions for comparison between the centers and combined analysis of pooled data. The *Exposure* samples consisted of 50 subjects in the other centers. In addition, samples of another 50–250 subjects, depending on the sampling logistics in each center, formed the less laborious *Diary* samples. Grenoble was an exception from this design, as described later.

In Helsinki, a base sample of the target population was formed by a random draw of 2523 adults (25–55 years of age) of the Helsinki Metropolitan Area from the population census. A short questionnaire about home environment, occupation, socioeconomic status, commuting, some personal characteristics and willingness to participate in the study was mailed to this primary population sample. A response rate of 75% was the target, and 1881 subjects did return the completed short questionnaire. The final *Exposure* and *Diary* subsamples were drawn at random from the primary sample subjects, who had answered the short questionnaire, after having excluded the clearly unwilling or unqualified (e.g. work outside of the area) individuals. Similar procedures were applied in other *EXPOLIS* centers.

In Athens, Basel, Milan and Prague the primary samples were also based on a random draw from the city inhabitants. However, in Milan and Prague the *Exposure* and *Diary* samples were drawn from the municipality employees. These samples will be compared to the larger primary population samples, and the results will be statistically corrected as necessary to represent the more general population. In Grenoble an ongoing study on the $PM_{2.5}$ exposures and daily symptoms of 40 volunteering asthmatics, 20 to 60 years of age, was adapted to yield $PM_{2.5}$ exposure results which can be related to the data from other *EXPOLIS* centers. Results of the population sampling process and descriptions of the primary population samples and the *Exposure* and *Diary* samples will be published separately. The primary, *Exposure* and *Diary* sample sizes in each *EXPOLIS* center are listed in Table 2.

A data integrity protocol was established according to the data security requirements of the *EU Directive on Protection of Individuals with Regard to Processing Personal Data in Medical and Epidemiological Research.* This protocol includes the contents and security of the *EXPOLIS* databases, use of person code numbers which cannot be translated back to identity, and training for the whole staff.

Measurement Scheme

The personal exposure and microenvironmental concentration data were collected from the *Exposure* subjects during one year from fall of 1996 to winter of 1997–1998. Each subject carried a personal exposure monitoring case, and her/his home, inside and outside, and workplace were equipped with microenvironmental measuring equipment for a period of 48 h. The workplace concentrations were measured for the normal working hours at the actual workplace of the subject, or if the subject moved from place to place during work, at a typical workplace. The home inside and outside concentrations were monitored from the time when the subject would normally return from work to the time when she/he would normally leave home for work. The measurements were made during the work weeks, mostly from Monday morning to Wednesday morning, and Wednesday evening to Friday evening. The *Diary* subject's data collection covers the same periods.

Weekend exposures were considered either (i) to be simpler than workday exposures, and (ii) to occur often outside of the urban area of interest, or (iii) to be too uncommon (e.g., moth spraying, painting , motorbike maintenance), which on the one hand would require much larger population samples for representative coverage and on the other hand are outside the main scope of urban environmental management.

The common weekend exposures can be simulated using the database, the less common ones would require separate focused sampling programs.

Personal and Microenvironmental Measurements

The purpose of the following description of the sampling and analysis procedures is just to shortly list the equipment and name the methods. Detailed descriptions of the $PM_{2.5}$ and VOC sampling and sample analysis techniques, together with VOC methods intercalibration data, QA methods and QC data will be published separately.

The personal exposure monitoring equipment (PEM), (sampling pump, 2.5 μm cyclone, 37 mm holders with filters, VOC sampling tube, CO monitor, and a battery pack) was packed into a 5.2 kg (total) aluminum briefcase carried by each subject for 48 h. The modified Buck IH (A.P. Buck Inc., Orlando Florida) pump is silent, lightweight and after modification capable of sampling 48 h with a single set of batteries and therefore suitable for personal measurements. It was adjusted to draw air at 4 L/min using a simple volumetric flow control. Small $PM_{2.5}$ GK2.05 cyclones for personal $PM_{2.5}$ sampling at 4 L/min were designed and constructed for the *EXPOLIS* study (BGI Inc., Waltham, MA). With this design the filters are handled from pre- to post-weighing in standard 37 mm plastic filter holders which minimizes the risk of filter contamination and damage in the field. In the laboratory the flow rate was adjusted to 4 L/min with a bubble flow meter (Mini BUCK Calibrator M-30) before and controlled after the sampling period. Two filter holders with 2 μm pore Gelman Teflo (Gelman Sciences, Ann Arbor, Michigan) filters were provided for each subject: one 'day filter' for two sampling periods beginning at leaving home for work and ending at return home from work, and one 'night filter' for the remaining times. The subjects changed the PEM filter holders according to personal instructions.

VOCs were sampled into a Perkin Elmer Tenax-TA tube (VOC-tube) by vacuum of the same pump that sampled the $PM_{2.5}$. The target sample size was 2 to 3 L, the VOC-tube flow rate was restricted to about 0.5–1.0 mL/min, and VOC diffusion to the tube before and after timed sampling was prevented by drawing the sample air into and from the VOC-tube through 200 mm long stainless steel capillary tubes. VOC-tube flow rate was measured before and after each sampling by a bubble flowmeter (Mini BUCK Calibrator M-1). In Basel the VOCs were collected using Carbotrap tubes instead of Tenax-TA. The target sample sizes and flow rates for Carbotrap sampling were about 10 times higher than with Tenax-TA.

The CO-PEM used was the CO Enhanced Measurer T15 (Langan Products Inc., San Francisco California) based on diffusion air flow to a CO specific electrochemical detector. The unit records the CO concentration (0.1–12.8 ppm, and 1–128 ppm ranges) as well as internal and external temperature in short, user-selectable intervals. 1 minute interval was used in *EXPOLIS* measurements. The measured values together with date/time were internally stored in memory for later downloading to a computer.

Workplace and home indoor and outdoor microenvironmental monitors (MEM) (sampling pump, 2.5 μm impactor, 47 mm filter holder with filters and a VOC tube packed into a portable sound absorbing container) were programmed to run inside and outside of the home for the expected nonworking hours and in the workplace for the expected working hours of each subject. The MEM sampler contained a WINS $PM_{2.5}$ (EPA Well Impactor Ninety Six) impactor (BGI), a 47 mm filter holder (BGI) with a Gelman Teflo filter and a PQ100 pump (BGI). The WINS $PM_{2.5}$

is a single jet well impactor designed to remove particles with a 50% cut size at 2.5 μm at 16.7 L/min. A Graseby-Andersen PM_{10} inlet (Sierra- Andersen, Inc.) preceding the WINS $PM_{2.5}$ impactor was used in outdoor measurements during bad weather to avoid wind and rain effects. The PQ100 pump is weatherproof, equipped with a microprocessor-controlled timing and mass flow adjustment system, and capable of operating up to 36 h on an internal lead-acid battery. The pump is designed to pull in a sample of air at a constant flow rate of 1.0–25 L/min (mass flow rate accuracy ± 5 %). The flow rate was measured/adjusted before each sampling and controlled after sampling with a bubble flow meter (Buck M-30).

The VOC-tube arrangement was identical to the PEM case, except that the flow rate was adjusted to about 2 mL/min with Tenax TA sampling and 20 mL/min with Carbotrap sampling.

Sample Analysis
The VOCs were thermally desorbed from the tubes and subsequently analyzed at VTT, Chemical Technology, Finland, by GC separation and simultaneous detection by MSD and FID. The VOC samples collected on Carbotrap, were analyzed by Carbotech, SA in Switzerland using GC/FID technique. The $PM_{2.5}$ sample filters were weighed before and after sampling in each center using a microbalance, and archived in a refrigerator for later elemental/chemical analyses. These analytical procedures together with QA/QC data will be described in detail in later articles.

The NO_2 samples were collected using Palmes passive tubes (Palmes et al., 1976). The tubes were prepared and analyzed spectrophotometrically in Zürich University of Technology. This technique has been used for personal sampling in numerous studies (e.g., Alm et al., 1998) and proven reliable.

Questionnaires and Time-activity Monitoring
EXPOLIS used four questionnaire-based data collection tools:

1) A Short Screening Questionnaire,
2) a Core Questionnaire,
3) a Time-Microenvironment-Activity-Diary (TMAD), and
4) a Retrospective 48 h Exposure Questionnaire.

The purpose of the *Short Screening Questionnaire* was particularly to evaluate the subjects' intention for participation.

The *Core Questionnaire* covered the indoor air quality related characteristics of each subject's home and workplace, as well as commuting, socioeconomic and some exposure related personal characteristics, such as smoking.

The *TMAD* was needed to assess the times that subjects spent in each microenvironment and activity while their personal exposures and the microenvironmental concentrations were measured. The TMAD asked the subjects to mark each 15 min of the day at the appropriate microenvironment-activity category (see Figure 3). The microenvironment categories in this TMAD are *in transfer* (*walk/bike, motor cycle, car/taxi, bus/tram,* and *metro/train*) and *not in transfer* (*home in* and *out, work in* and *out, other in* and *out*), and activities are *cooking, smoking self* and *smoking in same room.* Multiple entries (e.g. home indoor, home outdoor, car) are allowed for each 15 min. In the analysis each 15 min is divided evenly between all entries

The *Retrospective Short-Term Exposure Questionnaire* referred to 'the last 48 hours' and was requested to be completed at the end of the 48-hour PEM/MEM measurement period of each subject. The 48-hour recall questions addressed specific activities which may influence personal exposure, particularly to VOCs (cleaning, gluing, etc.).

All questionnaires were originally prepared in English, and translated to the 6 *EXPOLIS* languages and back-translated independently to control for a common meaning and understanding of each question. As some TMAD locations (e.g. metro in Basel) and some questions (e.g. gas fired hot water heaters in Helsinki) are irrelevant in some centers, such locations and questions were omitted in local translations. Centers also added some questions of local research interest to the questionnaire. However, for the need of combining the data the coding of the common locations and questions remained the same in all centers.

EXPOLIS *Database (EADB)*

A common relational database (*EXPOLIS* Access DataBase, EADB) was developed using Microsoft Access 7.0 to contain all *EXPOLIS* data from the local Civil Register or other local registers, TMADs, questionnaires, monitors, laboratory analyses, calibration procedures and environmental conditions (see Figure 4). The idea was to store all data in their primary form, if possible by direct downloading from pumps, microbalances and monitors, and to perform all physical and statistical calculations on the primary data in this database. EADB contains 36 data tables for data storage and over 100 queries and forms to facilitate the data entry, concentration calculations and other database usage. For privacy protection purposes all information identifying a particular subject is removed, when the questionnaire, TMAD and monitoring data for the subject have been entered and quality controlled.

The centrally developed database together with data entry tools was distributed to all *EXPOLIS* centers, and these databases were transferred to the central EADB at KTL, Helsinki. After the field work has been completed, all local EADBs are quality controlled, cleaned and combined into a European EADB. This final *EXPOLIS* database is then distributed to each *EXPOLIS* center for data queries and analyses of local as well as European *EXPOLIS* data.

To maximize the utility of the complete EADB, its structure will be published in 1999, and the EADB will also be made available to research teams outside of the *EXPOLIS* team for specified analyses.

Pilot Phase

The selected measuring equipment were tested in Milan and Kuopio prior to the pilot. Prior to the survey all equipment, techniques, training, instructions, questionnaires, standard operating procedures (SOPs) and general information materials were tested with volunteer subjects. The pilot samples were analyzed and the pilot experiences collected from all EXPOLIS centers. This material was discussed and assessed in a common workshop, and the SOPs, questionnaires and TMADs were edited according to the pilot experiences. The database structure and data entry (questionnaires and TMAD) and downloading tools (pumps, CO monitors and microbalances) were developed and tested during and after the pilot phase.

Team Organization

A complicated multicenter (and multilingual) protocol like *EXPOLIS*, where multiple compounds are monitored in multiple microenvironments, needs a great deal of practical everyday problem-solving and other communication to ensure on the one hand a common practice and comparable study results, and on the other hand minimum data losses. The junior researchers were trained at the different phases of the study together in *EXPOLIS*-Workshops in Prague (April 21–24, 1996), Helsinki (September 9–13, 1996), Grenoble (March 23–26, 1997), and Bilthoven (February 5–8, 1998). These opportunities were also used for equipment intercalibrations. In each center one researcher was assigned to one or more of the following contact groups: *Equipment*, *Database* and *VOCs*. *QA/QC* and *Privacy Protection* responsibilities lay within the principal investigators. For example the Database Contact Group members collected all database-related problems, ideas and experiences in each center, and communicated them to other centers for distribution there. Communication occurred mostly via E-mail and faxes, but each junior researcher was also assigned a GSM telephone (Nokia 2110 or 1610) with the GSM numbers of all other *EXPOLIS* junior researchers and principal investigators programmed to ensure fast access when and where problems/questions were encountered in the field or laboratory.

Field Survey

The base sample population first received an information letter about *EXPOLIS* and a Short Screening Questionnaire, which they were asked to complete and send back to the local *EXPOLIS* center, including the response card indicating the intention to participate.

The *Exposure* subjects were drawn from the database, which contained all subjects in the primary sample, who responded to the short questionnaire and were not excluded from exposure monitoring. The subjects were then contacted by telephone to remind about the study and to agree about the exact timing of the measurement period, time and place for meeting, driving instructions, etc. To start the measurements a junior researcher went to the home and—where the employers accepted—to the workplace of the subject, positioned the MEMs, gave the PEM and instructed the subject with regard to the filter change procedure, the core questionnaire and TMAD use. The subject was also offered a GSM telephone for the 48 hours to easily reach the respective junior researcher in case of any problems or questions.

The *Diary* subjects were drawn and contacted similarly, and invited to a meeting where TMADs and questionnaires were distributed and their use was instructed to a small group at a time. Those *Diary* subjects who could not come to the meetings, were contacted at home or workplace. The TMADs and questionnaires that they completed are almost identical to the *Exposure* subjects'. The *Diary* subjects returned these materials in prepaid and addressed envelopes.

Quality Assurance

The performance criteria of the quality assurance program in *EXPOLIS* were in general to minimize any differences between the centers which would affect the comparability of the results, and specifically to ensure quantified data for all $PM_{2.5}$, CO and NO_2 exposures and microenvironmental concentrations. A maximum detection limit of 1 µg/m^3 was requested for all VOCs in the target compound list (see Table 3).

The first performance criteria were pursued by using identical sampling equipment, questionnaires, time-activity diaries and work procedures in all centers (except for VOCs in Basel), by training the junior researchers together in common workshops and by encouraging daily communication between them between the workshops.

The QA was based on the principles that (i) all procedures must be carefully planned, tested and performed according to standard operating procedures (SOPs) approved by the study director, (ii) each unit of data must be traceable as to who produced it, when, with what equipment and according to which SOP(s), and (iii) if any deviations or irregularities occur they must be recorded. Before the pilot phase, the preparation work for preliminary standard operating procedures (pSOP) was distributed among the centers and individual researchers. A SOP for preparation of SOPs was applied for guidance as to the structure and contents of each pSOP. Preliminary SOPs were prepared for the pilot stage for all subject, field and laboratory procedures, accepted by the local principal investigators, and distributed to all *EXPOLIS* centers. They covered contacting and instructing the subjects, the use, maintenance and calibration of the measuring equipment, preparation and positioning of the samplers, collection, identification and handling of the samples, collection and handling of data, as well as recording and archiving of all data from the study. These pSOPs were tested in the pilot phase, corrected and retested until they were approved as real SOPs for the field work by the study coordinator and the KTL QA Unit. The approved SOPs were distributed to all *EXPOLIS* centers and filed by the coordinator. Table 5 lists the standard operating procedures of the *EXPOLIS* study.

TABLE 5. The Common Standard Operating Procedures Developed for the *EXPOLIS* Study

Preparation of standard operating procedures (SOPs)	SOP Expolis/KTL-G-1.0
Customer procedure, *Exposure* subjects	SOP Expolis/KTL-I-1.0
Customer procedure, *Diary* subjects	SOP Expolis/KTL-I-2.0
MEM sampler positioning and PEM sampler carrying instructions	SOP Expolis/KTL-F-1.0
PM2.5 PEM sampling	SOP Expolis/KTL-F-1.0
PM2.5 MEM sampling	SOP Expolis/KTL-F-3.0
PM2.5 Teflon filter analysis	SOP Expolis/ETHZ-L-5.0
VOC sampling	SOP Expolis/VTT-F-4.0
VOC sample analysis	SOP Expolis/VTT-L-3.0
CO monitoring	SOP Expolis/KTL-F-5.0
Core questionnaire & application	SOP Expolis/UoA-I-3.0
Time-location-activity diary (TMAD) & application	SOP Expolis/UoA-I-4.0

The QA Unit also monitored the study in the different *EXPOLIS* centers according to the study protocol and the respective SOPs. The QA Unit inspected the protocol, and the critical phases of the study in the different *EXPOLIS* centers to ensure that the research plan and SOPs were followed, and reviews the reports and report to the study director.

MICROENVIRONMENT ORIENTED EXPOSURE SIMULATION (MOSES)

The *EXPOLIS* project will substantially expand the European exposure database, yet the number of individuals for which full data sets on microenvironmental and personal exposure data will be obtained is modest and restricted to the work week of urban adult working-age population. Additional sources of information will be needed for indirect exposure assessment. Some of this information has been collected in the *EXPOLIS* framework, but most will be collected from other unrelated studies; information on pollutant concentrations in other microenvironments; data on other age groups, results from indoor air quality surveys etc. The Monte Carlo technique uses probabilistic information about simultaneous or sequential steps to predict the probability distribution of a wanted outcome. In the case of population exposure simulation, input data consist of time-activity distributions and the microenvironmental concentration distributions and the desired result is the probability distribution of personal exposures in the population.

The Dutch Public Health Institute, RIVM, in Bilthoven will together with KTL develop, test and validate a model system for the assessment of integrated (sub-)population exposure distributions for selected air pollutants. The model should be applicable for scenario studies, assessing the public health gains of environmental policy options (in terms of population exposure and/or population attributive risk). The modeling tool should facilitate the evaluation of policy scenarios, e.g., traffic policies. In combination with information on exposure-response relations for the pollutants, it could potentially estimate public health gain/loss under changing policy scenarios.

EXPECTED RESULTS

Based on the monitored population samples in 6 European cities, the *EXPOLIS* project will produce and report the following new information (for the work weeks), which is currently not available from Europe:

- Levels and distributions of population exposures to $PM_{2.5}$, VOCs, CO, and for some cities also NO_2.
- Levels and distributions of microenvironmental concentrations of $PM_{2.5}$, VOCs and for some cities also NO_2 in workplaces and inside and outside of homes.
- Distributions of time and activities in different microenvironments of European urban populations.
- Statistical associations between different air pollutant levels in personal exposures and different microenvironments.
- Key determinants—housing, occupation and commuting related, geographic, behavioral and socioeconomic—affecting personal exposure and microenvironmental pollution levels.

ACKNOWLEDGMENTS

EXPOLIS has been supported by EU contracts ENV4-CT96-0202 and ERB IC20-CT96-0061 (Prague), Academy of Finland contract No: 36586 (Helsinki, Finland), Swiss Ministry for Education and Science contract BBW No: 95.0894 (Basel, Switzerland), French National Environment Agency (ADEME), Union Routière de France and Grenoble Communauté de Communes (Grenoble, France), other national research funds, and intramural funding from the participating institutes.

REFERENCES

ACKLAND, G., HARTWELL, T., JOHNSON, T., and WHITMORE, R. (1985). "Measuring human exposure to carbon monoxide in Washington, D.C., and Denver, Colorado, during the winter of 1982–1983." Environ. Sci. Technol. 19:911–918.

ALM, S., REPONEN, A., MUKALA, K., PASANEN, P., TUOMISTO, J., and JANTUNEN, M.J. (1994). "Personal exposures of preschool children to carbon monoxide: Roles of ambient air quality and gas stoves." Atmos. Environ. 28:3577–3580.

ALM, S., MUKALA, K., JANTUNEN, M.J., PASANEN, P., TIITTANEN, P., RUUSKANEN, J., and TUOMISTO, J. (1998). "Personal NO_2 Exposures of Preschool Children in Helsinki." J. Expos. Anal. Environ. Epidem. 8(1):79–100.

BEVAN, M.A.J., PROCTOR, C.J., BAKER-ROGERS, J., and WARREN, N.D. (1991). "Exposure to carbon monoxide, respirable suspended particulates, and volatile organic compounds while commuting by bicycle." Environ. Sci. Technol. 25:788–791.

CLAYTON, A., PERRIT, R.L., PELLIZZARI, E.D., THOMAS, K.W., WHITMORE, R.W., WALLACE, L.A., ÖZKAYNAK, H., and SPENGLER, J.D. (1993). "Particle total exposure assessment methodology (PTEAM) 1990 study: Distributions of aerosol and elemental concentrations in personal, indoor, and outdoor air samples in a southern California community." J. Expos. Anal. Environ. Epidem. 3(2):227–250.

CLEAN AIR ACT AMENDMENTS (1990). Conference report to accompany S. 1630, Report No. 101-952; U.S. Govt. Printing Office: Washington, DC. pp. 139–162.

DAB, W., MEDINA, S., QUÉNEL, P., LE MOULLEC, Y., LE TERTRE, A., THELOT, B., MONTEIL, C., LAMELOISE, P., PIRARD, P., MOMAS, I., FERRY, R., and FESTY, B. (1996). "Short term respiratory health effects of ambient air pollution: Results of the APHEA project in Paris." J. Epidem. Community Health 50:s42–s46.

DOCKERY, D.W., SCHWARTZ, J., and SPENGLER, J. (1992). "Air pollution and daily mortality: Associations with particulates and acid aerosols." Environ. Res. 59:362–373.

DOCKERY, D.W., POPE, C.A., XU, X., SPENGLER, J., WARE, J., FAY, M., FERRIS, B., and SPEIZER, F. (1993). "An association between air pollution and morality in six U.S. cities." New Engl. J. Med. 329:1753–1759.

DOCKERY, D.W. and POPE, C.A. (1994). "Acute respiratory effects of particulate air pollution." Ann. Rev. Public Health 15:107–132.

DUAN, N. and MAGE, D.T. (1997). "Combination of direct and indirect approaches for exposure assessment." J. Expos. Anal. Environ. Epidem. 7(4):493–504.

FISCHER, P., BRUNEKREEF, B., and BOLEIJ, J.S.M. (1986). "Indoor NO_2 pollution and personal exposure to NO_2 in two areas with different outdoor NO_2 pollution." Environ. Monitoring Assess. 6:221–229.

FISHER, P.H., LEBRET, E., and VAN DE POLL, W. (1993). "Personal monitoring of exposure to ozone; Results of a pilot study." In: Proceedings of the Fourth European Meeting of Environmental Hygiene. June 9–11, 1993. Wageningen, The Netherlands.

HARTWELL, T.D., PELLIZZARI, E.D., PERRIT, R.L., WHITMORE, R.W., ZELON, H.S., SHEELDON, L.S., and SPARACINE, C.M. (1987). "Results from the total exposure assessment methodology (TEAM) study in selected communities in northern and southern California." Atmos. Environ. 21:1995–2004.

HEI Oversight Committee (1995). "Particulate air pollution and daily mortality: Replication and validation of selected studies: Phase I report of the particle epidemiology evaluation project." August, 1995.

HOECK, G., MEIJER, R., SCHOLTEN, A., NOIJ, D., and LEBRET, E. (1984). " The relationship between indoor nitrogen dioxide concentration levels and personal exposure: A pilot study." Intl. Arch. Occup. Environ. Health 55:73–78.

HOFFMANN, K., SCHWABE, R., KRAUSE, C., SCHULTZ, C., SEIFERT, B., and ULLRICH, D. (1996). Umwelt-Survey 1990/91, Band IV: Personengebundene Exposition gegenüber flüchtigen organischen Verbindungen in den alten Bundesländern. Institut für Wasser-, Boden-, und Lufthygiene des Umweltbundesamtes, Berlin, Germany.

LIOY, P. J., WALDMAN, J. M., BUCKLEY, T., BUTLER, J., and PIETARINEN, C. (1990). "The personal, Indoor and outdoor concentrations of PM-10 measured in an industrial community during the winter." Atmos. Environ. 24B:57–66.

LIOY, P. J. and PELLIZZARI, E. (1995). "Conceptual framework for designing a national survey of human exposure." J. Expos. Anal. Environ. Epidem. 5(3):425–444.

JUNGERS, R.H., AKLAND, G., and SAULS, H. (1985). "A human exposure assessment study using personal monitoring of carbon monoxide—direct and indirect methods." Environ. Int. 11:413–418.

KATSOUYANNI, K., ZMIROU, D., SPIX, C., SUNYER, J., SCHOUTEN, J.P., PÖNKÄ, A., ANDERSON, H.R., LEMOULEC, Y., WOJTYNIAK, B., VIGOTTI, M.A., and BACHAROVE, L. (1995). " Short-term effects of air pollution on health: A European approach using epidemiological time-series data." Eur. Respir. J. 8:1030–1038.

MONN, C., FRAUCHIGER, P., and WANNER, H. (1993). "Assessment of personal exposure to nitrogen dioxide and ozone." In: Indoor Air '93: Proceedings of the 6th International Conference on Indoor Air Quality and Climate (M.J. Jantunen, P. Kalliokoski, E. Kukkonen, K. Saarela, O. Seppänen, and H. Vuorelma, eds.). Helsinki, Finland, July 4–8, 1993. Vol. 3. Combustion Products, Risk Assessment, Policies." Gummerus, Jyväskylä, Finland, 1993. pp. 319–323.

NKB (1994). "Testing and discussion of the proposed criteria for the ability of chemical substances to cause allergy and hypersensitivity in the skin and lower airways." NKB Committee and Work Reports 1994:03E, Nordic Committee on Building Regulations, NKB, Indoor Climate Committee. Painatuskeskus OY, Helsinki, Finland. 58 pp.

O'CONNOR, T., LEADERER, B., HODFORD, T., and BRACKEN, M. (1993). "Measurement of exposure to environmental tobacco smoke in pregnant women using questionnaire, personal monitor and urine cotinine: A problem in exposure modelling." In: Indoor Air '93: Proceedings of the 6th International Conference on Indoor Air Quality and Climate (M.J. Jantunen, P. Kalliokoski, E. Kukkonen, K. Saarela, O. Seppänen, and H. Vuorelma, eds.). Helsinki, Finland, July 4–8, 1993. Vol. 3. Combustion Products, Risk Assessment, Policies." Gummerus, Jyväskylä, Finland, 1993. pp. 373–378.

OTT, W., THOMAS, J., MAGE, D., and WALLACE, L. (1988). "Validation of the simulation of human activity and pollutant exposure (SHAPE) model using paired days from the Denver, CO, carbon monoxide field study." Atmos. Environ. 22:2101–2113.

PALMES, E.D., GUNNISON, A.F., DIMATTIO, J., and TOMCZYK, C. (1976). "Personal sampler for nitrogen dioxide." Am. Ind. Hyg. Assoc. J. 37:570–577.

ÖZKAYNAK, H., SPENGLER, J., BUTLER, D. and BILLICK, I. (1993). "Predicting the distribution of population exposures to NO_2 in a large urban area." In: Indoor Air '93: Proceedings of the 6th International Conference on Indoor Air Quality and Climate (M.J. Jantunen, P. Kalliokoski, E. Kukkonen, K. Saarela, O. Seppänen, and H. Vuorelma, eds.). Helsinki, Finland, July 4–8, 1993. Vol. 3. Combustion Products, Risk Assessment, Policies." Gummerus, Jyväskylä, Finland, 1993. pp. 295–300.

ÖZKAYNAK, H., XUE, J., SPENGLER, J., WALLACE, L., PELLIZZARI, E., and JENKINS, P. (1996). "Personal Exposure to Airborne Particles and Metals: Results from the Particle TEAM Study in Riverside, California." J. Expos. Anal. Environ. Epidem. 6(1):57–78.

PONCE DE LEON, A., ANDERSON, H.R., BLAND, J.M., STRACHAN, D.P., and BOWER, J. (1996). "Effects of air pollution on daily hospitals admissions for respiratory disease in London between 1987–1988 and 1991–1992." J. Epidem. Community Health 50:s63–s70.

PÖNKÄ, A. and VIRTANEN, M. (1996). "Asthma and ambient air pollution in Helsinki." J. Epidem. Community Health 50:s59–s62.

POPE, C.A., THUN, M.J., NAMBOODIRI, M.M., DOCKERY, D.W., EVANS, J.S., SPEIZER, F.E., and HEATH, C.W. (1995). "Particulate air pollution as predictor of mortality in a prospective study of U.S. adults." Am. J. Respir. Crit. Care Med. 151:669–674.

QUACKENBOSS, J., SPENGLER, J., KANAREK, M., LETZ, R., and DUFFY, C. (1986). "Personal exposure to nitrogen dioxide: Relationship to indoor/outdoor air quality and activity patterns." Environ. Sci. Technol 20:775–783.

RYAN, P.B., SPENGLER, J.D., SCHWAB, M., SOCZEK, M.L. and BILLICK, I.H. (1989). "Nitrogen Dioxide Exposure Studies: I. The Boston Personal Monitoring Study." In: Proceedings of the First International Symposium on Total Exposure Assessment Methodology: A New Horizon, Las Vegas, NV. November 26–28, 1989. pp. 38–65.

SCHOUTEN, J.P., VONK, J.M., and DE GRAAF, A. (1996). "Short term effects of air pollution on emergency hospital admissions for respiratory disease: results of the APHEA project in two major cities in the Netherlands." J. Epidem. Community Health 50:s22–s29.

SCHWARTZ, J., SPIX, C., TOULOUMI, G., BACHÁROVÁ, L., BARUMAMDZADEH, T., LE TERTRE, A., PIEKARKSI, T., PONCE DE LEON, A., PÖNKÄ, A., ROSSI, G., SAEZ, M., and SCHOUTEN, J.P. (1996). "Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions." J. Epidem. Community Health 50:s3–s11.

SCHWARTZ, J. and DOCKERY, D.W. (1992). "Particulate air pollution and daily mortality in Steubenville, Ohio." Am. J. Epidem. 135:12–19.

SEXTON, K., CALLAHAN, M.A., BRYANT, E.F., SAINT, C.G., and WOOD, W.P. (1995). "Information decisions about protecting and promoting human health: The rationale for the national human exposure assessment survey." J. Expos. Anal. Environ. Epidem. 5(3):233–256.

SONG, R., WANG, G., and ZHOU, J. (1993). "Study on the personal exposure level to nitrogen dioxide for housewives in Beijing." In: Indoor Air '93: Proceedings of the 6th International Conference on Indoor Air Quality and Climate (M.J. Jantunen, P. Kalliokoski, E. Kukkonen, K. Saarela, O. Seppänen, and H. Vuorelma, eds.). Helsinki, Finland, July 4–8, 1993. Vol. 3. Combustion Products, Risk Assessment, Policies." Gummerus, Jyväskylä, Finland, 1993. pp. 337–342.

SPENGLER, J., SCHWAB, M., RYAN, P.B., BILLICK, I.H., COLOME, S., WILSON, A.L., and BECKER, E. (1994). "Personal Exposure to Nitrogen Dioxide in the Los Angeles Basin." J. Air Waste Manag. Assoc. 44:39–47.

SPIX, C. and WICHMANN, H.E. (1996). "Daily mortality and air pollutants: findings from Köln, Germany." J. Epidem. Community Health 50:s52–s58.

SUNYER, J., CASTELLSAGUÉ, J., SÁEZ, M., TOBIAS, A., and ANTÓ, J.M. (1996). "Air pollution and mortality in Barcelona." J. Epidem. Community Health 50:s76–s80.

THOMAS, K.W., PELLIZZARI, E.D., CLAYTON, A., WHITAKER, D.A., SHORES, R.C., SPENGLER, J., ÖZKAYNAK, H., FROEHLICH, S.E., and WALLACE, L.A. (1993). "Particle total exposure assessment methodology (PTEAM) 1990 study: Method performance and data quality for personal, indoor, and outdoor monitoring." J. Expos. Anal. Environ. Epidem. 3(2):203–226

TOULOUMI, G., SAMOLI, E., and KATSOUYANNI, K. (1996). "Daily mortality and "winter type" air pollution in Athens, Greece—a time series analysis within the APHEA project." J. Epidem. Community Health 50:s47–s51.

VIGOTTI, M.A., ROSSI, G., BISANTI, L., ZANOBETTI, A., and SCHWARTZ, J. (1996). "Short term effects of urban air pollution on respiratory health in Milan, Italy, 1980–1989." J. Epidem. Community Health 50:s63–s70.

WALDMAN, J.M., LIOY, P.J., GREENBERG, A., and BUTLER J.P. (1991) "Analysis of human exposure to benzo(a)pyrene via inhalation and food ingestion in the Total Human Exposure Study (THEES)." J. Expos. Anal. Environ. Epidem. 1(2):193–225.

WALLACE, L.A., PELLIZZARI, E.D., HARTWELL, T.D., WHITMORE, R., and SPARACINO, H.Z. (1986). "Total Exposure Assessment Methodology (TEAM) study: Personal exposures, indoor outdoor relationships, and breath levels of volatile organic compounds in New Jersey." Environ. Int. 12:369–387.

WALLACE, L., THOMAS, J., MAGE, D., and OTT, W. (1988). "Comparison of breath CO, CO exposure and Coburn model predictions in the U.S. EPA Washington-Denver (CO) study." Atmos. Environ. 22:2183–2193.

WALLACE, L., ÖZKAYNAK, H., SPENGLER, J., PELLIZZARI, E., and JENKINS, P. (1993). "Indoor, outdoor and personal air exposures to particles, elements and nicotine for 178 southern California residents." In: Indoor Air '93: Proceedings of the 6th International Conference on Indoor Air Quality and Climate (M.J. Jantunen, P. Kalliokoski, E. Kukkonen, K. Saarela, O. Seppänen, and H. Vuorelma, eds.). Helsinki, Finland, July 4–8, 1993. Vol. 3. Combustion Products, Risk Assessment, Policies." Gummerus, Jyväskylä, Finland, 1993. pp. 445–450.

WORKING GROUP "CLASSICAL" AIR POLLUTANTS AND HEALTH (1995). Update and Revision of the Air Quality Guidelines for Europe. WHO Regional Office for Europe, EUR/ICP/EHAZ 94 05/PB01.

WATTS, R. LEWTAS, J., STEVENS, R., HARTLAGE, T., PINTO, J., WILLIAMS, R., HATTAWAY, K., MISKOVA, I., BENES, I., KOTESOVEC, F., and SRAM, R. (1994). "Czech-United-States EPA health study: Assessment of personal and ambient air exposures to PAH and organic mutagens in thew Teplice district of Northern Bohemia." Intl. J. Environ. Anal. Chem. 56:271–287.

WORLD HEALTH ORGANIZATION (WHO) (1991). "An introductory guide to human exposure field studies—Survey methods and statistical sampling." WHO/PEP/HEAL/91.26, UNEP/GEMS/92.H4, 15 pp.

VAN WIJNEN, J.H., VERHOEFF, A.P., JANS, H.W.A., and VAN BRUGGEN, M. (1995). "The exposure of cyclists, car drivers and pedestrians to traffic-related air pollutants. Intl. Arch. Occup. Environ. Health. 67:187–193.

XUE, J., ÖZKAYNAK, H., WARE, J., SPENGLER, J., BILLICK, I., and SILVERS, A. (1993). "Alternative estimates of exposure to nitrogen dioxide and their implications to epidemiologic study design." In: Indoor Air '93: Proceedings of the 6th International Conference on Indoor Air Quality and Climate (M.J. Jantunen, P. Kalliokoski, E. Kukkonen, K. Saarela, O. Seppänen, and H. Vuorelma, eds.). Helsinki, Finland, July 4–8, 1993. Vol. 3. Combustion Products, Risk Assessment,

Mauscript Received: October 14, 1997
Manuscript Accepted: May 29, 1998

II

Kruize, H., Hänninen, O.O., Breugelmans, O., Lebret, E., Jantunen, M., 2003. **Description and demonstration of the *EXPOLIS* simulation model: Two examples of modeling population exposure to particulate matter**. *Journal of Exposure Analysis and Environmental Epidemiology* 13 (2): 87-99.

Journal of Exposure Analysis and Environmental Epidemiology (2003) 13, 87–99

www.nature.com/jea

Description and demonstration of the *EXPOLIS* simulation model: Two examples of modeling population exposure to particulate matter

HANNEKE KRUIZE,[a] OTTO HÄNNINEN,[b] OSCAR BREUGELMANS,[a] ERIK LEBRET[a] AND MATTI JANTUNEN[b]

[a]*National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands*
[b]*KTL, Department of Environmental Health, Kuopio, Finland*

As a part of the *EXPOLIS* study, a stochastic exposure-modeling framework was developed. The framework is useful to compare exposure distributions of different (sub-) populations or different scenarios, and to gain insight into population exposure distributions and exposure determinants. It was implemented in an MS-Excel workbook using @Risk add-on software. Basic concept of the framework is that time-weighted average exposure is a sum of partial exposures in the visited microenvironments. Partial exposure is determined by the concentration and the time spent in the microenvironment. In the absence of data, indoor concentrations are derived as a function of ambient concentrations, effective penetration rates and contribution of indoor sources. Framework input parameters are described by probability distributions. A lognormal distribution is assumed for the microenvironment concentrations and for the contribution of indoor sources, and a beta distribution for the time spent in a microenvironment and for the penetration factor. Mean and standard deviation values parameterize the distributions. In this paper, Latin Hypercube sampling is used for the input distributions. The outcome of the framework is an estimate of the population exposure distribution for the selected air pollutant. The framework is best suited for averaging times from 24 h upwards. Sensitivity analyses can be performed to determine the most influential factors of exposure. The application of the framework is illustrated in two examples. The *EXPOLIS* $PM_{2.5}$ example uses microenvironment measurement and time–activity data from the *EXPOLIS* study to model $PM_{2.5}$ population exposure distributions in four European cities. The results are compared to the observed personal exposure distributions from the same study. The Dutch PM_{10} example uses input data from several (Dutch) databases and from literature, and shows a more complex application of the framework for comparison of scenarios and subpopulations.
Journal of Exposure Analysis and Environmental Epidemiology (2003) **13,** 87–99. doi:10.1038/sj.jea.7500258

Keywords: *population exposure distributions, microenvironment approach, exposure model, air pollution, $PM_{2.5}$, PM_{10}.*

Introduction

Epidemiological research of the past 20 years has revealed significant mortality and morbidity effects in association to present European and North American levels of urban air pollution, especially fine particulate matter (PM). These studies are based on air pollution levels that have been measured at centrally located ambient air monitoring stations (Vedal, 1997; Spix et al., 1998; Dab et al., 2001; Pope et al., 2002). Health effects of air pollutants, however, are caused by the exposures people experience during their daily activities. People in Europe and North America spend most of their time indoors (Szalai, 1972; Schwab et al., 1990; Klepeis et al., 2001) where, in addition to pollution from outdoor sources, also indoor sources of air pollutants are present (Lioy, 1990). Indoor and personal pollution levels often correlate poorly with outdoor air levels (Dockery and Spengler, 1981; Letz et al., 1984; Sexton et al., 1984; Ott, 1985; Spengler et al., 1985; Ryan et al., 1986; Lioy, 1990, 1995; Law et al., 1997; Pellizzari et al., 1999; Kousa et al., 2002). Better understanding of the relationships between the personal exposures to various air pollutants and ambient air levels, and their relationships to other significant exposure determinants (such as indoor sources, sinks, and personal activities) are therefore needed before the epidemiological findings can be interpreted into efficient risk reduction policies (Ott, 1984; NRC, 1998).

Exposure can be defined as the contact of a target and a chemical, physical, or biological agent in an environmental carrier medium (Duan, 1982; Ott, 1985; Zartarian et al., 1997). It can be measured or modeled (Ryan, 1991), either directly (personal measurements) or indirectly (microenvironment approach) (Duan, 1982, 1991; Ott, 1984, 1985; Ott et al., 1988; Lioy, 1990; Ryan, 1991; Duan and Mage, 1997). Personal exposure measurements are expensive (Ott, 1984; Ryan, 1991), labor intensive and invasive (Letz et al., 1984; Sexton et al., 1984). Modeling requires a validated

1. Address all correspondence to: Hanneke Kruize, National Institute of Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, The Netherlands. Tel.: +31-30-2743599 or +31-30-2533398. Fax: +31 30-2744451. E-mail: Hanneke.Kruize@rivm.nl
Received 28 May 2002

model, and sufficient, representative, good-quality input data. Once these requirements are met, a model can be repeated for a large number of individuals or populations. Models can be used to assess past exposures, exposures of not sampled or undersampled groups in the population or to compare alternative future exposure scenarios (Letz et al., 1984; Lioy, 1990, 1995; Ryan, 1991). Only little demands have to be made on the study population in comparison with personal measurements (for which the study population, e.g., has to carry sampling equipment). These are all significant benefits when compared to measurements (Letz et al., 1984; Ryan et al., 1986).

Ryan (1991) describes three classes of human exposure models for air pollutants: statistical, physical, and physical–stochastic models. *Statistical* models can be used for descriptive analyses and testing of hypotheses on collected data. In the *physical* approach, the model is based on physical (and sometimes chemical) laws. The *a priori* defined physical model is transformed into a mathematical model (Ryan, 1991). An example of this deterministic type of models is the National Ambient Air Quality Standards (NAAQS) Exposure Model (NEM) (Johnson, 1995; McCurdy, 1995). *Physical–stochastic type* of models are based on physical equations like the pure physical models, but instead of relying on deterministic input data to fully describe the variability — or ignoring the variability — in input parameters, physical–stochastic models apply probabilistic techniques to propagate the variability through the model. These models describe parameters with frequency or probability distributions instead of single values (Ryan, 1991). Examples of this type of models are the Simulation of Human Air Pollution Exposures model (SHAPE) (Ott, 1984; Ott et al., 1988; Duan, 1991; Ryan, 1991), pNEM, the probabilistic version of the NEM model (McCurdy, 1995; Law et al., 1997), and the Air Pollution Exposure model (AirPEX) (Freijer et al., 1998). These models can be used to predict population exposures for both existing and past or scenario situations, and for subpopulations for whom no measurement data are available (Ryan, 1991), by simulating from the distributions of input parameters.

Full description of personal exposure to an air pollutant requires knowledge of the magnitude of pollutant concentration in the exposure environment, duration of exposure, and the time pattern of the exposure (Ryan, 1991). The microenvironment approach has been commonly used to model exposures (Fugas, 1975; Dockery and Spengler, 1981; Ott, 1984; Letz et al., 1984; Ryan et al., 1986; Freijer et al., 1998). In the microenvironment approach the exposure E is calculated as the sum of the partial exposures across the visited microenvironments Eq. (1) (e.g., Duan, 1982; Ryan et al., 1986):

$$E = \sum_{i}^{N} f_i C_i$$

where C_i is the concentration in microenvironment i, f_i the fractional time spent in microenvironment i, and N the number of microenvironments.

In literature, the exposure E is often defined as "total exposure", (e.g., Ryan, 1991). However, we prefer to use the term "time-weighted average exposure", because in our opinion it better expresses the fact that the exposure is the sum of weighted concentrations to which people are exposed in the microenvironments they visit. This equation can be used for any averaging time and any number of microenvironments, for any air pollutant. In case no measured data are available for indoor environments, the concentration can be derived as a function of outdoor concentration, the effective penetration factor, and the contribution of indoor sources (e.g., Dockery and Spengler, 1981; Ryan et al., 1986): (Eq (2))

$$C_i = C_a p_i + S_i$$

where C_a is the ambient concentration, p_i the effective penetration factor of the air pollutant in microenvironment i, and S_i the contribution of indoor sources in microenvironment i.

The effective penetration factor includes both first-order infiltration and first-order loss mechanisms (sinks) (Ryan et al., 1986). According to Ryan et al. (1986), p_i and S_i are dependent on many parameters, such as ventilation rates and family activity patterns. In Figure 1, this nested model is outlined for two types of outdoor environments (in this example, the urban and rural environment), with different indoor microenvironments nested within them.

As a part of the *EXPOLIS* study, a European multicenter study for measurement of air pollution exposures and microenvironment concentrations of working age urban populations (Jantunen et al., 1998, 1999), a population exposure simulation framework was developed to assess and predict exposure distributions of air pollutants of European urban populations. This simulation framework should be applicable in scenario studies, to assess the public health gain of environmental policy options in terms of population exposure. Another condition was that the framework should fit on information resulting from the *EXPOLIS* study. Furthermore, it had to be developed in such a way that all participating centers of the *EXPOLIS* study could use it without extensive developer's support. Also, the framework should be applicable to perform calculations for various (sub-) populations and air pollutants, and should produce population exposure distributions (Jantunen et al., 1998). Based on these conditions, the Dutch National Institute of Public Health and the Environment (RIVM) developed the modeling framework in collaboration with KTL (Finnish National Public Health Institute).

This paper describes the development and structure of the framework. Two examples demonstrate the application of the framework to simulate population exposure to particulate matter. The first one, the *EXPOLIS* $PM_{2.5}$ example, applies

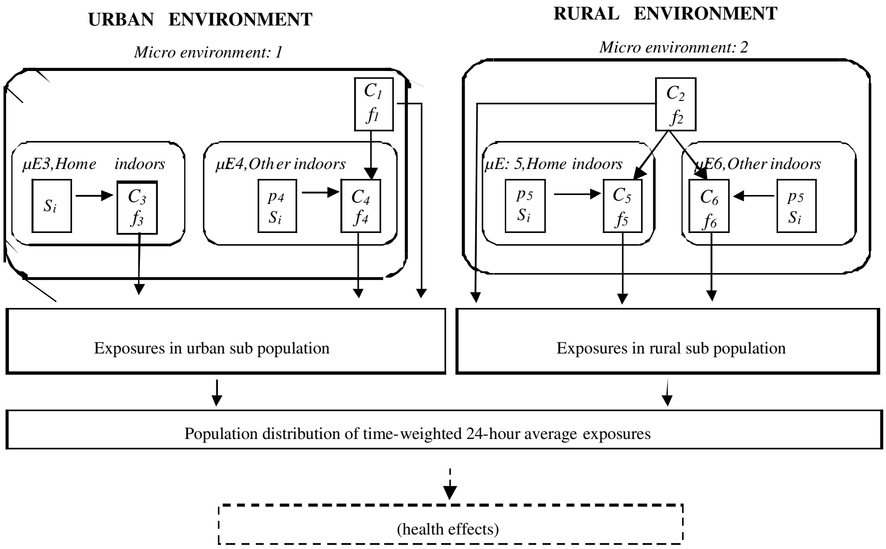

Figure 1. Outline of the nested structure of a microenvironmental exposure model. Concentration distributions for microenvironments 1 and 2 and for the microenvironment 3 are known. Concentrations for microenvironments 4–6 are modeled using penetrations and local sources.

the framework to four *EXPOLIS* cities to model the adult urban 48-h population exposure distributions to fine particulate matter ($PM_{2.5}$). The results are compared to the observed 48-h personal exposure distributions from the same study. The second one, the Dutch PM_{10} example, shows a more complex application, using the framework to model 24-h respirable particulate matter (PM_{10}) exposures of the whole Dutch population. In this example, the target population is divided into eight subpopulations based on age, work status and living in either a rural or an urban area. In the Dutch example, the results are calculated separately for the *current* scenario, including exposures to Environmental Tobacco Smoke (ETS), and for a hypothetical *non-ETS* scenario.

Methods

General Features of the Framework

The developed framework is based on Eqs. (1) and (2). It was implemented as a Microsoft Excel workbook. An Excel add-on software package @Risk (version 3.5, Palisade Corporation, 1994) is needed to supply the probabilistic functions for the stochastic functionality. The spreadsheet-based approach allows easy use of the framework by researchers who are not modelers or programmers by training. @Risk offers the user possibilities to choose their own simulation features, for example, selecting either sampling by the Latin Hypercube method or the Monte Carlo method. When Latin Hypercube sampling is used to create random realizations from the input distributions, the input probability distribution is stratified into equal intervals. Samples are taken randomly from each interval of the input distribution. Therefore, compared to the regular Monte Carlo sampling, fewer samples are needed to create the whole distribution. Owing to this way of sampling, also situations occurring with a lower probability are represented in the simulation output, for example, high concentrations sampled from the tail of a microenvironmental concentration distribution. Moreover, @Risk allows correlation among input variables, according to the researcher's specification. @Risk offers several options to present or analyze model outputs.

Required Input Data

Equations (1) and (2) show that three types of input data are required. First of all, relevant microenvironments need to be defined. The specification of microenvironments depends on the goal for which the framework is applied, data availability, and correlation between the microenvironments. The pollutant being studied is important for the selection of microenvironments, because the microenvironments, in which the source of the pollutant is present, vary between pollutants (Ott, 1985). Furthermore, a more detailed distinction with more microenvironments can produce a more accurate estimate, but it also requires more input data.

Secondly, concentration distributions need to be described for each microenvironment. In literature, concentration distributions and other distributions, which have a minimum level of zero and no upper limit, are often approached as a lognormal distribution (Ryan et al., 1986). Therefore, we assume all concentration distributions to be lognormal. For the microenvironments for which input data are available, Eq. (1) is used. In case the concentration distribution for an indoor microenvironment is not available, it is derived from the ambient concentrations, the effective penetration factors, and the contributions of indoor sources (Eq. (2)). For the distribution of the effective penetration factor, a beta distribution is assumed, limited between zero and one. This type of distribution allows many different shapes (Ryan et al., 1986). For the contribution of indoor sources a lognormal distribution is assumed. Furthermore, the percentage of indoor microenvironments with specified sources needs to be given.

Finally, data on time–activity patterns are needed, specified as the fraction of time spent in each microenvironment. People spend their time differently, depending on employment status, age (Letz et al., 1984), season, and day of week (Johnson, 1995), among other factors (Chapin, 1974). Therefore, it is important to define groups of people with similar time–activity patterns. For such subpopulations exposure distributions need to be simulated separately, and eventually merged together to get an exposure distribution for the whole population. We describe time fractions with a beta distribution, limited between zero and one, for the same reason as mentioned for the penetration factor. The simulation framework samples the time fractions from independent beta distributions. To scale the total fraction of time for each simulated individual to unity, each time fraction is divided by the sum of the fractions before calculating the partial exposures in microenvironments.

All distributions are entered as mean and standard deviation (SD) into the worksheet of the framework. The @Risk lognormal function is described by its mean and SD (note: not geometric mean and GSD) values. For the @Risk beta function, the mean and SD are transformed with Excel formulas to the needed function called parameters $\alpha 1$ and $\alpha 2$.

Since input variables might be correlated (for example, a person spending much time indoors will spend less time outdoors), a correlation matrix was implemented, in which the user can enter rank correlation coefficients. After having sampled from all relevant distributions using the Latin Hypercube sampling technique, the sampled values are combined, resulting in a partial exposure in one microenvironment for one individual. Summing all partial exposures and repeating this procedure according to the selected number of iterations generates the distribution of time-weighted average exposure levels for the target (sub-) population. From this population exposure distribution several exposure measures can be derived, such as the average exposure level, or exposure levels at different percentiles of the population exposure distribution. Also, sensitivity analyses can be performed to give an overview of the relative influence of the input parameters on the simulated population exposure distribution.

The following examples demonstrate the application of the framework.

The EXPOLIS $PM_{2.5}$ Example

The *EXPOLIS* $PM_{2.5}$ example demonstrates the use of the simulation framework in its simplest form. No subpopulations are defined, and no correlation structures between the model parameters are taken into account. ETS or any other indoor sources are not modeled separately, but are included in the observed total indoor microenvironment concentrations. One simulation was run for each city to estimate adult (age 25–55 years) urban population exposure distributions in Athens, Basel, Helsinki, and Prague. The modeled and measured 48-h population exposure distributions were compared to give a general impression of the validity of the framework. Table 1 summarizes the input data, which were

Table 1. Input values for the *EXPOLIS* $PM_{2.5}$ example: lognormal concentration distributions (arithmetic mean, SD) and beta distributions (arithmetic mean, SD) for time–activity for each of the three microenvironments of this model.

Microenvironment	Helsinki mean (SD)	Basel mean (SD)	Prague mean (SD)	Athens mean (SD)
		$PM_{2.5}$ concentrations ($\mu g/m^{-3}$)		
No of subjects	201	50	50	50
Home indoors	12.1 (15.1)	24.4 (24.7)	35.9 (30.0)	32.4 (20.9)
Work indoors	15.9 (34.8)	27.8 (38.5)	43.8 (44.6)	91.9 (81.3)
Outdoors[a]	9.3 (6.9)	21.4 (13.9)	26.9 (10.5)	36.6 (26.7)
		Time activity (fractions)		
No of diaries	434	322	83	100
Home indoors	0.58 (0.13)	0.56(0.14)	0.59(0.15)	0.64(0.18)
Work indoors	0.25 (0.18)	0.23(0.14)	0.23(0.14)	0.17(0.14)
Other places[b]	0.18 (0.10)	0.21(0.11)	0.18(0.12)	0.19(0.11)

[a] In Helsinki: fixed station 1 h, in other cities: *EXPOLIS* home outdoor 2-night concentration.
[b] "Outdoors" concentration is used for "Other places" in the simulation.

extracted from the *EXPOLIS* database Hänninen (et al., 2002). Three microenvironments were defined: "Home indoors", "Work indoors", and "Other places" (an aggregate microenvironment covering all other places visited). In the EXPOLIS study, data were collected from the fall of 1996 to the winter of 1997–1998. Measurements were carried out during two consecutive weekdays. Concentration distributions for $PM_{2.5}$ were available for the "Home indoors" (measurement period was $2 \times 16 = 32$ h) and "Work indoors" microenvironments (measurement period was $2 \times 8 = 16$ h). The measured indoor concentrations include both ambient $PM_{2.5}$ particles penetrated indoors, as well as particles from any indoor source. In Helsinki, 1-hour ambient concentrations measured at a traffic-oriented fixed monitoring station were randomly sampled for the "Other places" microenvironment, because the visits to the "Other places" microenvironments are typically short, and often occur in traffic. In Athens, Basel, and Prague, the approximately 32-h average concentrations measured outdoors at home were used, because no hourly fixed site $PM_{2.5}$ data were available for these cities. This treatment is likely to narrow the distribution of concentrations experienced in the "Other places" from their real, but unknown, values. No correction was applied to the standard deviations in spite of the different averaging times. Data on the fractions of time spent in the defined microenvironments were also available from the *EXPOLIS* database. The participants kept a 15 min resolution time–microenvironment–activity diary for 48 consecutive hours. Time–activity data were collected during weekdays, but not in weekends or holidays. Time spent in the microenvironment "Other places" was calculated by subtracting the time spent in the other microenvironments "Home indoors" and "Work indoors" from total time that the diary was kept. In all, 2000 iterations and a random number seed were selected for each of the four simulation runs.

The Dutch PM_{10} Example

We present the second example to show the use of the *EXPOLIS* framework for purposes other than the *EXPOLIS* study itself, in a larger and more complex set up, for comparisons of subpopulations and scenarios based on different policy options. This example is derived from work performed for the Dutch Health Inspectorate, in which rough estimates were made for the exposure of the Dutch population to fine particles. We estimated the population exposure distribution on the basis of directly available information from existing (Dutch) databases and from literature, gathered within a short time period. In this example, "fine particles" were defined to be PM_{10}, because Dutch air-quality guidelines are defined at PM_{10}. Consequently, more Dutch data were available for PM_{10} compared to $PM_{2.5}$, at least for the outdoor microenvironment. The input data are summarized in Tables 2–5. The subpopulations were formed on the basis of expected general similarity of time–activity patterns within groups and data availability. This resulted in the following subpopulations: children (0–12 years), the working/studying population (13–64 years), the nonworking and nonstudying population (13–64 years), and the elderly (≥ 65 years). In the following, we will refer to these groups as "Children", "Adults W", "Adults N", and "Elderly". The subpopulations, with their percentages of occurrence in the general Dutch population, are shown in Table 2. Two scenarios were simulated: the current Dutch situation including the presence of ETS in indoor environments (*current* scenario), in which we tried to give a rough estimate of the current exposure of the Dutch population to PM_{10}, and the hypothetical situation with no indoor smoking (*non-ETS* scenario). For the definition of microenvironments, we selected those for which we expected Dutch input data to be available, and for which the distinction would be meaningful in relation to PM. This resulted in four microenvironments "Outdoors", "Home indoors", "Other indoors", and "In transport".

Since measurements from fixed monitoring stations indicated that the ambient PM_{10} concentrations were higher in urban areas compared to rural areas (mean 39.7 $\mu g/m^3$, SD 17.4 $\mu g/m^3$ and mean 35.1 $\mu g/m^3$, SD 18.3 $\mu g/m^3$ respectively), we made a distinction between the urban and the rural part of the Netherlands (Kruize et al., 2000). Ambient concentrations were available from the National Air Quality Monitoring Network of the RIVM (Elzakker and Buijsman, 1999). We considered data on indoor PM_{10}

Table 2. Subpopulations, their percentages of occurrence in the Dutch population, and the number of iterations used in the Dutch PM_{10} example.

Subpopulation	Age (years)	Urban		Rural		Total	
		(%)	Iterations	(%)	Iterations	(%)	Iterations
Children	0–12	5.4	2170	10	4011	15.5	6181
Adults W[a]	13–64	18.2	7283	25.7	10,288	43.9	17,571
Adults N[b]	13–64	11.3	4516	15.9	6378	27.2	10,894
Elderly	65 +	5.9	2342	7.5	3012	13.4	5354
Total	—	40.8	16,311	59.2	23,689	100	40,000

[a]Working or studying adults.
[b]Adults not working or studying.

concentrations, available from Dutch studies (for example, Janssen, 1998; Fischer et al., 2000), not to be representative for Dutch homes in general, because measurements were performed in a limited number of Dutch homes, at a limited number of locations in The Netherlands. Therefore, the indoor concentration distribution for PM_{10} was derived from ambient concentrations using a penetration factor (Eq. (2)). In the absence of Dutch data on penetration factors, parameters of the probability distribution for the penetration factor were derived from calculations using a mass balance model, in which the input consisted of ambient concentrations, a fixed ventilation rate ($0.64\,h^{-1}$), and the half-life for PM_{10} (1.41 h) (Freijer and Bloemen, 2000). The resulting distribution for the effective penetration rate was parameterized with a mean of 0.6 and an SD of 0.04. These values are comparable with values presented in literature (Colome et al., 1992; Li, 1994). Owing to a lack of specific data for different types of indoor microenvironments, the distribution parameters of the microenvironment "Home indoors" were also used for the microenvironments "Other indoors" and "In transport".

The additional indoor concentrations caused by ETS were simulated in a separate simplified stochastical model. In this model (a modified version of the one described in Kruize et al., 2000), data on the additional indoor concentration of fine particles from one cigarette, the number of smoked cigarettes per person, and the number of smokers in a household, were combined to derive the input parameters of the lognormal distribution for the contribution of ETS. The additional indoor concentration per cigarette ($2.2\,\mu g/m^3$) was derived from the average emission per cigarette (12 mg; Koutrakis et al., 1992), the average volume of a Dutch house (assumed to be $250\,m^3$), a deposition rate of 12 per day, and a ventilation rate of 15.3 per day (Freijer and Bloemen, 2000). The number of smoked cigarettes per person was derived from a Dutch survey on smoking (Stivoro, 1999). For the adult and elderly subpopulations it was assumed that smoking would be present only in the "Home indoors" and "Other indoors" microenvironments. For the "Other indoors" microenvironment the same input parameters for the

Table 3. ETS concentration input values Dutch PM_{10} example: percentage of Dutch households with ETS, and the concentration distribution of additional PM_{10} in indoor environments caused by ETS (arithmetic mean, SD), by subpopulation.

Subpopulation	Age (years)	Percentage of households (%)	ETS-caused PM_{10} level ($\mu g/m^3$)	
			Mean	(SD)
Children	0–12	47.9	57.1	(51.7)
Adults W[a]	13–64	53.2	59.1	(55.0)
Adults N[b]	13–64	53.2	55.7	(50.6)
Elderly	65+	28.7	46.8	(41.2)

[a]Working or studying adults.
[b]Adults not working or studying.

Table 4. Input values for time–activity for the Dutch PM_{10} example: fractions of time spent daily in the microenvironments (arithmetic mean, SD), by subpopulation.

	Children 0–12 years		Adults W[a] 13–64 years		Adults N[b] 13–64 years		Elderly 65+ years	
	n=1101		n=2805		n=874		n=276	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Outdoors	0.13	(0.13)	0.13	(0.14)	0.14	(0.13)	0.15	(0.13)
Home indoors	0.72	(0.13)	0.62	(0.18)	0.76	(0.17)	0.78	(0.14)
Other indoors	0.11	(0.13)	0.19	(0.17)	0.06	(0.1)	0.04	(0.07)
In transport	0.04	(0.04)	0.05	(0.05)	0.04	(0.06)	0.03	(0.05)

Columns do not add up 1.00 due to rounding.
[a]working or studying adults.
[b]Adults not working or studying.

Table 5. Spearman rank correlation input values for time–activity fractions used in the Dutch PM_{10} example.

	Home indoors				Other indoors				Outdoors			
	Children	Adults W[a]	Adults N[b]	Elderly	Children	Adults W	Adults N	Elderly	Children	Adults W	Adults N	Elderly
Home indoors	1	1	1	1	—	—	—	—	—	—	—	—
Other indoors	−0.49	−0.56	−0.29	−0.11	1	1	1	1	—	—	—	—
Outdoors	−0.56	−0.24	−0.68	−0.72	−0.20	−0.49	−0.15	−0.22	1	1	1	1
In transport	−0.3	−0.42	−0.3	−0.29	0.33	0.28	0.43	0.5	−0.09	−0.05	−0.07	0.06

[a]Working or studying adults.
[b]Adults not working or studying.

contribution of ETS were applied as used for the "Home indoors" microenvironment, because no representative specific data could be found for different types of indoor microenvironments. For children, it was assumed that ETS exposure would only occur in "Home indoors", because we assumed no smokers to be present with children in "Other indoors" (for example, Dutch day nurseries) and "In transport" (Kruize et al., 2000). The number of indoors or household and the percentages of households with smoking were derived from a time–activity survey performed by the Dutch research institute "Intomart" in a sample of the Dutch population ($n = 5056$) (Freijer et al., 1998). The input data used to simulate the contribution of ETS in indoor microenvironments are summarized in Table 3.

The earlier mentioned Dutch time–activity survey performed by Intomart aimed at gathering data of different subpopulations in The Netherlands in such a way that they could be used to estimate exposures to air pollutants for these subpopulations. Therefore, we could use these data for the Dutch PM_{10} example. Data were collected for both week and weekend days, during three time periods: the summer period (July–September 1994), the winter period (November 1994–February 1995), and in episodes with predicted maximum temperatures above 25°C (July and September 1994). Intomart weighed time–activity data for age and gender in order to get representative time–activity data for The Netherlands. During 24 h people selected every 15 min, the location they visited at that moment, the activity they performed at that moment, and how strenuous the activity was, from a preformatted list. From these data statistics were calculated for the selected subpopulations and microenvironments as presented in this example. The time–activity input parameters are summarized in Table 4. Spearman rank correlation between distributions of time spent in the different microenvironments was calculated using data from the Dutch time–activity survey (Table 5).

For each scenario, separate simulations were performed for urban and rural inhabitants, and for the mentioned subpopulations. A weighted number of iterations were selected according to the occurrence of each subpopulation in the Dutch population, as derived from Dutch Census data (Table 2). For each subpopulation, we selected at least 2000 iterations, and we used a random seed. In total, 40,000 iterations were used for each scenario. Sensitivity analyses were performed using regression analyses, in order to determine the influence of the input parameters on the outcome (Kruize et al., 2000).

Results

The EXPOLIS $PM_{2.5}$ Example

The cumulative simulated 48 h-average population exposure distributions for the *EXPOLIS* $PM_{2.5}$ example are presented in Figure 2. Statistics of the simulated and the corresponding observed exposure distributions are summarized in Table 6.

The observed mean exposure levels were highest in Athens, followed by Prague, Basel, and Helsinki (37, 35, 31, and 16 $\mu g/m^3$, respectively). The corresponding simulated values rank to the same order, indicating that the relative exposure levels between cities can be estimated using these kind of models. The highest observed mean exposure level (Athens) was 2.3 times higher than the lowest one (Helsinki, 37/16). The corresponding ratio for simulated values is 3.3, clearly higher (43/13). Comparing the extreme cities the simulation models seem to exaggerate the difference.

In Basel and Helsinki, the simulation models underestimate the mean exposure levels, while in Athens and

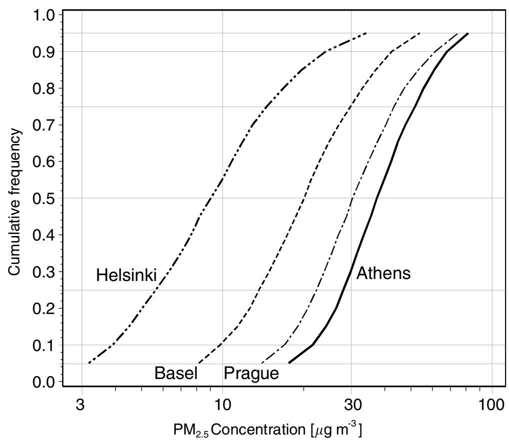

Figure 2. Simulation results for the *EXPOLIS* $PM_{2.5}$ example: exposure distributions of adults in four European cities (percentiles 5–95).

Table 6. Summary statistics of simulation results and corresponding observed exposures for the *EXPOLIS* $PM_{2.5}$ example.

	Helsinki		Basel		Prague		Athens	
	Sim.[a]	Obs.[b]	Sim.	Obs.	Sim.	Obs.	Sim.	Obs.
n	2000	193	2000	46	2000	47	2000	29
$PM_{2.5}$ exposures ($\mu g/m^3$)								
Mean	13	16	25	31	37	35	43	37
SD	30	19	20	43	30	26	30	25
25%	6	6	14	15	22	19	28	20
50%	9	10	20	20	30	25	37	29
75%	15	18	30	30	43	42	52	41
90%	24	33	43	50	62	57	68	70
95%	34	43	54	74	75	82	82	74

[a]Sim. = simulated in this work.
[b]Observed in *EXPOLIS* measurements.

Prague the simulated levels are higher than the observed ones. Differences between the observed and the simulated means range from $+2\,\mu g/m^3$ in Prague to $\pm 6\,\mu g/m^3$ in Athens and Basel. Relatively speaking, these maximum differences are +16 and −19%, respectively.

The simulated standard deviations do not rank to the same order as the corresponding observed values. Especially in Helsinki, the simulated standard deviation is too high and in Basel too low (+58 and −54%, respectively). Basel is the only city for which the standard deviation was underestimated.

The simulated main percentiles shown in Table 6 compare to the observed values similarly to the mean values in most cases. If the mean was underestimated, most of the percentiles are underestimated too; in fact, for Basel and Helsinki none of the percentiles was overestimated. For Athens and Prague most of the percentiles (except the 90th and 95th, respectively) were overestimated.

The Dutch PM_{10} Example

Figure 3 shows the cumulative simulated population exposure distributions for PM_{10} in the Dutch population for both the *current* scenario and the *non-ETS* scenario. Summary statistics of the distributions are presented in Table 7, for the whole population and for the subpopulations.

The average exposure level of the whole population appeared to be almost halved in the hypothetical case where people would not smoke in the indoor environments, indicating that roughly half of the present population exposures to PM_{10} are caused by passive exposure to tobacco smoke in indoor environments. Differences between urban and rural environments were analyzed, but appeared to be small (approximately $3\,\mu g/m^3$). Since all other input values were the same, this difference in modeled exposure levels is caused solely by the differences in ambient concentrations.

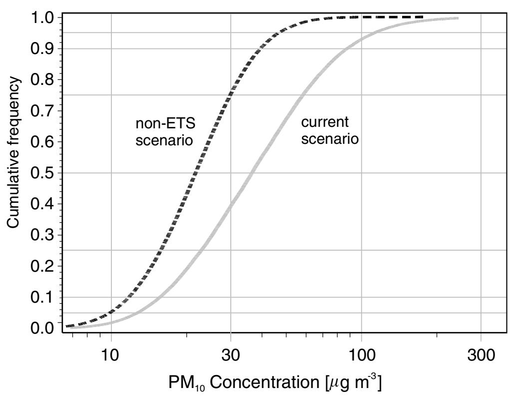

Figure 3. Simulation results for the Dutch PM_{10} example: exposure distributions of the Dutch population for the *current* and the hypothetical *non-ETS* scenarios.

In the *current* scenario, the differences between the subpopulations (Table 7) were mainly because of the differences in ETS exposure, but were also caused by differences in time–activity patterns. Elderly people clearly appeared to experience the lowest exposure levels, with a largest difference in means compared to the adult population ($36\,\mu g/m^3$ *versus* 49 and $50\,\mu g/m^3$, respectively). The sensitivity analyses confirmed that the contribution of ETS and time spent indoors were the most influential factors in this scenario.

In spite of the differences in time–activity patterns, variations in the exposure distributions of the subpopulations were very small (maximum difference of $1\,\mu g/m^3$) in the *non-ETS* scenario. From the sensitivity analyses, the ambient concentrations appeared to have the largest influence on the population exposure distribution.

In this example, largest differences were found between the two presented scenarios, again emphasizing the influence of ETS on population exposure to PM. Compared to these differences, the differences between urban and rural dwellers, and between the subpopulations were small.

Discussion

In this paper, the *EXPOLIS* simulation framework is presented as a tool to provide insight into population exposure to air pollutants, without costly and invasive personal measurements. The examples presented in this paper illustrate the framework structure and usability. PM exposures are modeled in both examples, but the framework can be applied for other air pollutants as well. The *EXPOLIS* $PM_{2.5}$ example demonstrated the use of the simulation framework in its simplest form, and the Dutch PM_{10} example showed that, although the *EXPOLIS* framework was developed as a part of the *EXPOLIS* study, it can be used for purposes beyond this project as well, in a larger and more complex set up, as long as input parameters can be derived from literature or existing databases. The presented results of the Dutch PM_{10} example should be interpreted carefully. For the work ordered by the Dutch Health Inspectorate, input data were gathered within a short time period, and should therefore preferably be directly available Dutch data. Since not all the required input data were directly available, several assumptions were made, and proxy's were used. Also, not all directly available data appeared to be representative for the general situation in The Netherlands. For example, the time–activity data used in the Dutch example were not gathered during the whole year, and might therefore not give a representative idea of yearly average time–activity patterns. Consequently, the simulation results of the current scenario only give a rough estimation of the PM_{10} population exposure in The Netherlands, and the effect of a virtual "no indoor smoking" policy. Once more,

Table 7. Simulation result statistics for the subpopulations and the whole population in the Dutch PM_{10} example.

	Children		Adults W[a]		Adults N[b]		Elderly		All	
	Current ($\mu g/m^3$)	Non-ETS ($\mu g/m^3$)	Current ($\mu g/m^3$)	Non-ETS ($\mu g/m^3$)	Current ($\mu g/m^3$)	Non-ETS ($\mu g/m^3$)	Current ($\mu g/m^3$)	Non-ETS ($\mu g/m^3$)	Current ($\mu g/m^3$)	Non−ETS ($\mu g/m^3$)
	n = 6181		n = 17,571		n = 10,894		n = 5354		n = 40,000	
Mean	44	24	50	24	49	24	36	24	47	24
SD	36	12	39	12	39	12	28	12	38	12
25%	20	15	25	16	23	16	19	16	23	16
50%	34	21	40	22	38	22	28	22	37	22
75%	55	30	62	30	61	30	43	30	58	30
90%	84	39	93	40	92	40	67	40	88	40
95%	110	46	119	48	116	47	87	47	114	47

[a]Working or studying adults.
[b]Adults not working or studying.

we would like to emphasize that this example presented mainly to show the usability of the *EXPOLIS* framework for comparisons of subpopulations and scenarios, more than showing the most accurate results.

Usability of the Framework

The presented examples show that the framework can be used well for comparison between several existing or nonexisting situations or populations. First, we presented a comparison of population exposures to $PM_{2.5}$ in four European cities (Athens, Basel, Helsinki, and Prague), included in the *EXPOLIS* study, which demonstrates how models can be built to estimate population exposures in different cities. From the study of Rotko et al. (2000b) it appears that response rates were below US standards. However, because both the input for the simulations and the personal measurement data were derived from the same *EXPOLIS* database, and the low response rates do not stand in the way the comparison between simulation and measured data as presented here. Furthermore, the fact that the nonresponse was high, emphasizes the need of modeling next to measuring. The simulated means compared rather well to observed ones; absolute maximum differences were $+6\,\mu g/m^3$ in Athens and $-6\,\mu g/m^3$ in Basel, and both of these values are within $\pm 20\%$ of the observed levels. The main reason for these differences is that the ETS exposures are not fully captured by the microenvironment measurements used as inputs for the *EXPOLIS* framework, as some ETS exposure occurs also in other places (and rooms) than where the microenvironmental measurements were taken. Furthermore, the simulated variances of three out of four cities were overestimated probably because of the fact that the lognormal concentration distributions used in this work were not truncated. One or two extreme concentration values generated by the Latin Hypercube sampling increase the simulated standard deviation significantly. In the case of Helsinki, the simulated maximum exposure was higher ($1128\,\mu g/m^3$) than in any of the other cities and more than an order of magnitude higher than any of the hundreds of measured concentrations, even though all microenvironment concentration means were the lowest in Helsinki. This artefact caused by the sampling technique does not affect any of the percentiles below 99th and, owing to the large number of samples, affects the mean only slightly. The artefact, however, does affect the standard deviation, and thus the truncation of lognormal distributions should be considered, as suggested by Hänninen et al. (2003) and Hänninen and Jantunen, 2003), when the standard deviations (or other measures of variance) are reported.

Secondly, we performed a scenario analysis for The Netherlands, considering the current population exposure to PM_{10} (including ETS) with a *non-ETS* scenario, serving as an example of determining the effect of a potential policy option, in this case a "no indoor smoking" policy. The average exposure level of the whole population appeared to be almost halved in case people would not smoke in the indoor environments. Also, the sensitivity analyses showed ETS to be an important determinant of exposure, together with time spent indoors (where ETS exposure took place). Other exposure studies also indicate that tobacco smoking is the most or one of the most important contributors of personal exposure to PM (Dockery and Spengler, 1981; Letz et al., 1984; Sexton et al., 1984; Spengler et al., 1985; Koutrakis et al., 1992; Rotko et al., 2000a; Koistinen et al., 2001). It is important to keep in mind that both the measured and modeled exposures in relation to tobacco smoke, both for smokers and nonsmokers, only include the impact of passive smoking and inhalation of environmental tobacco smoke (ETS). We realize that the exposure of a smoker from active smoking is much greater, but it was not assessed in this study.

The same ETS occurrences and concentration parameters modeled for a standard Dutch home as explained in the methods were used for the "Home indoors" and "Other indoors'' microenvironments for adults and elderly people, which can be considered an important approximation done in the PM_{10} model. For example, Hänninen and Jantunen (2002) report $PM_{2.5}$ ETS concentrations analyzed from *EXPOLIS* Helsinki data that are almost double in workplaces compared to homes. It is likely that also in the Netherlands the ETS concentrations in workplaces and, for example, restaurants are different from those in homes. This question cannot be answered without representative measurements. The current ETS model used in the Dutch PM_{10} example can be considered to be the best possible estimate, using directly available data.

A third comparison made in this paper was on subpopulations of the Dutch population. Elderly people appeared to experience lower exposures compared to the other subpopulations according to the *current* scenario. In the presented example, these differences between the subpopulations are caused for a greater part by the differences in the estimated exposures to ETS; a small fraction of the differences between the subpopulations is attributable to the differences in the ratio of times spent indoors and outdoors. In general, this type of comparison between subpopulations can be used to determine, for example, what part of the population is at risk, or to model exposure for specific sensitive groups, such as the elderly or children.

Another subdivision presented in the Dutch example was based on a (spatial) distinction between urban and rural dwellers. The differences in exposures between the Dutch urban and rural dwellers in the example are caused solely by the difference in ambient levels, being approximately $0.6 \times (40–35)\ \mu g\,m^3 = 3\ \mu g/m^3$, because all other input parameters were assumed to be equal between these two subpopulations.

With the simulated population exposure it is possible to estimate the health effects of the exposure as well, if reliable and comparable exposure–response relationships are available for PM from different sources and in different microenvironments. The elemental composition of indoor and outdoor PM can be rather different (Letz et al., 1984; Spengler et al., 1985; Koutrakis et al., 1992), and data are only emerging on the differences in the risk levels of PM from different sources (Laden et al., 2000; Pope et al., 2002). The only indoor source, for which broad-based risk assessments are available, is ETS (Hackshaw et al., 1997; Law et al., 1997), which is also the most important indoor source for PM. In order to be able to calculate the risks from multiple indoor and outdoor sources, the outcome measure of the modeling result should be the same as the exposure measure used in the exposure–response relationship.

The outcomes of the *EXPOLIS* framework can be tested against air-quality standards considering both indoor and outdoor environments. Again, for the PM-oriented examples presented in this paper it was not possible, because the standard for PM is based on only ambient concentrations.

Another aspect on the usability of the *EXPOLIS* framework concerns the averaging time for which the model can be used. As the time–activity is modeled as fractions of time, the framework can be applied to calculation of exposures with wide range of averaging times, but is best suitable for averaging times from 24 h upwards.

A last remark should be made on the way total fraction of time is calculated in the presented version of the model. Since fractions of time spent in the defined microenvironments are sampled from independent distributions for each defined microenvironment, the total sampled time might end up below or above one. In the presented version of the EXPOLIS framework, total time spent in the visited microenvironments is scaled to unity, by dividing each time fraction by the sum of all of the fractions of time of the visited microenvironments, before calculating the partial exposures in microenvironments, as applied in the Dutch PM_{10} example. In the *EXPOLIS* $PM_{2.5}$ example scaling was not necessary, because time spent in the microenvironment ''Other places'' was calculated by subtracting the time spent in the other microenvironments ''Home indoors'' and ''Work indoors'' from total time that the diary was kept a day, automatically resulting in a total sampled time of one, being another possibility to deal with this problem. We are aware that there are probably more (accurate) solutions (e.g., binary trees of probability distributions for occurrence of time fractions). Further research is needed to find out what is the effect of the way this problem is treated, and what solution would be best.

Data Availability

Users of the *EXPOLIS* framework have many opportunities to adapt the framework to their own needs within the limitations of data availability. These limitations of data availability became also apparent in the examples presented in this paper.

For example, in both examples presented in this paper, some data referred to only a short period of time for each study participant. In the *EXPOLIS* study, the participants kept a diary during two weekdays. The Dutch participants filled out the diary during only one day. This short period that the diary was kept does not necessarily represent the normal (average) time–activity pattern of a person. Therefore, the presented population exposure distributions do not allow analyses on an individual level (Klepeis, 1999). However, if time–activity data and concentration data would be available on a longer consecutive period of time, it would be possible to derive distributions of personal exposure levels of specific individuals.

Apart from this, data availability was not a big problem in the *EXPOLIS* $PM_{2.5}$ example, because the framework was built around the design of the *EXPOLIS* study. However, the Dutch PM_{10} example illustrates that data availability may become a problem when using the framework for other purposes. Although the current *EXPOLIS* framework has simple input requirements, representative PM_{10} concentration data for indoor environments and indoor sources (even on ETS) were not directly available for the Netherlands. Therefore, we had to decide whether to use available foreign data or Dutch data based on a limited number of measurement locations, assuming that these data are representative for the Netherlands, or to estimate input data making assumptions and using proxies. In this case, we chose the latter solution. For example, to be able to calculate the effect of passive smoking, we estimated the contribution of ETS to indoor concentrations with a separate stochastical model. It is well possible that another researcher would decide to use the available foreign or Dutch data of Janssen (1998) or Fischer et al. (2000) instead. This example makes it clear that the researcher's decision can influence the modeling results. In the absence of validation, however, it is questionable what solution would approach the real exposure situation most accurately.

As this problem must be recognizable for many researchers in the field of exposure modeling, it would be very helpful if more databases on (indoor) concentration data and exposure-relevant time–activity data were published and made available (Klepeis, 1999). In the USA, several useful initiatives have been taken already so far (Sexton et al., 1994; Johnson, 1995). For example, the CHAD is a large database in which several time–activity databases from the US and Canada are included, such as the NHAPS database (Klepeis, 1999; Klepeis et al., 2001; McCurdy et al., 2000). However, it is questionable as to what extent data from these studies can be used for other (European) countries, because time–activity patterns might be rather different in different countries, caused by the cultural differences in how people spend their time for example. Therefore, more large, international, multicenter exposure studies would offer a rich source of (comparable) data, often for more countries at the same time. The *EXPOLIS* study is an example of such a study, from which the database (including outdoor, indoor, and personal concentration data and time–activity data among others) will become accessible for the international research community. When more data on (indoor) concentrations of air pollutants, indoor sources, and time–activity patterns would become available in the future, one still needs to be careful in using them as input for the model, and one should evaluate the representativeness of the input data case by case.

Conclusions

The *EXPOLIS* simulation framework, described in this paper, is a helpful tool for researchers to support policy makers and policy evaluating processes by evaluating air pollution exposures in different scenarios, population groups, and locations. It is also useful for helping researchers to understand the factors affecting exposure levels.

The *EXPOLIS* example showed that the model predicted mean population exposure levels in four European cities with better than 20% accuracy. The presented version of the simulation framework, not applying truncation to lognormal concentration distributions, however, overestimated variances in three out of four cases.

The Dutch population PM_{10} exposure example demonstrated the use of the framework in modeling exposure levels of a large complex population in alternative scenarios, for different subpopulations. The results support the findings from field surveys that exposure to tobacco smoke approximately doubles the population exposures to particulate matter. Limited data availability asked for creative ways to derive input parameters for the EXPOLIS framework, and emphasizes the need for general accessibility of databases on (indoor) concentration data and exposure-relevant time–activity patterns to support researchers in the field of exposure modeling.

Acknowledgments

EXPOLIS has been supported by European Union 4th Framework Research, Technology and Development (RTD) program, contracts ENV4-CT96-0202 (DG12-DTEE) and ERB IC20-CT96-0061 (Prague); Academy of Finland contract No. 36586 (Helsinki); Swiss Ministry for Education and Science Contract BBW No. 95.0894 (Basel); French National Environment Agency (ADEME), Union Routière de France Communauté de Communes (Grenoble); and other national research funds, and intramural funding from the participating institutes. The Dutch PM_{10} example was derived from calculations performed as a part of a work commissioned by the Dutch Health Inspectorate.

We thank Dr. J.I. Freijer for his helpful comments on the draft versions of this manuscript.

References

Chapin Jr FS. *Human Activity Patterns in the City — Things People Do in Time and Space*. John Wiley & Sons, Inc., New York, USA, 1974.

Colome SD., Kado NY., Jacques P., and Kleinman M. Indoor–outdoor air pollution relations: particulate matter less than 10 μm in aerodynamic diameter (PM_{10}) in homes of asthmatics. *Atmos Environ* 1992: 6A (6): 2173–2178.

Dab W., Ségala C., Dor F., Festy B., Lamelose P., Moullec le Y., Tertre le A., Médina S., Quénel P., Wallaert B., and Zmirou D. Air pollution and health: correlation or causality? The case of the relationship between exposure to particles and cardiopulmonary mortality. *J Air Waste Manage Assoc* 2001: 51: 220–235.

Dockery DW., and Spengler JD. Indoor–outdoor relationships of respirable sulfates and particles. *Atmos Environ* 1981: 15: 335–343.

Duan N. Models for human exposure to air pollution. *Environ Int* 1982: 8: 305–309.

Duan N. Stochastic microenvironmental models for air pollution exposure. *J Expos Anal Environ Epidemiol* 1991: 1 (2): 235–257.

Duan N., and Mage DT. Combination of direct and indirect approaches for exposure assessment. *J Expos Anal Environ Epidemiol* 1997: 7 (4): 439–470.

Elzakker BG., and Buijsman E. Measurement activities in 1999 by the National Air Quality Monitoring Network (in Dutch) (Report No. 723101 032). National Institute of Public Health and the Environment, Bilthoven, The Netherlands, 1999.

Fischer PH., Hoek G., Reeuwijk van H., Briggs DJ., Lebret E., Wijnen van JH., Kingham S., and Elliott PE. Traffic related differences in outdoor and indoor concentrations of particles and volatile organic compounds in Amsterdam. *Atmos Environ* 2000: 34: 3713–3722.

Freijer JI., Bloemen HJTh., Loos de S., Marra M., Rombout PJA., Steentjes GM., and Veen van MP. Modelling exposure of the Dutch population to air pollution. *J Hazard Mater*. 1998: 61: 107–114.

Freijer JI., and Bloemen HJT. Modelling relationships between indoor and outdoor air quality. *J Air and Waste Manage Assoc* 2000: 50: 292–300.

Fugas M. Assessment of total exposure to an air pollutant. Proceedings of the International Conference on Environmental Sensing and Assessment. Paper No. 38-5, Vol. 2, IEEE #75-CH 1004-1, ICESA, 1975, Las Vegas.

Hänninen OO., and Jantunen MJ. Simulating personal $PM_{2.5}$ Exposure Distributions in Helsinki using ambient measurements as inputs (submitted).

Hänninen OO., and Kaarakainen E., and Alm S. *EXPOLIS* Databases. Project document manuscript. 2002. KTL, Kuopio, Finland.

Hänninen OO., Kruize H., Lebret E., and Jantunen, and MJ. *EXPOLIS* simulation model: $PM_{2.5}$ application and comparison to measurements in Helsinki. *J Expos Anal Environ Epidemiol* 2003: 13: 74–85.

Hackshaw AK., Law MR., and Wald NJ. The accumulated evidence on lung cancer and environmental tobacco smoke. *BMJ* 1997: 315: 980–988.

Jantunen MJ., Katsouyanni K., Knöppel H., Künzli N., Lebret E., Maroni M., Saarela K., Srám R., and Zmirou D., et al. Air pollution exposure in European cities: the *EXPOLIS* study. Final EU Report Contract nos. ENV4-CT96-0202 and ERB IC20-CT96-0061, 1998.

Jantunen MJ., Hänninen O., Katsouyanni K., Knöppel H., Künzli N., Lebret E., Maroni M., Saarela K., Srám R., and Zmirou D. Air pollution exposure in European cities: the "*EXPOLIS*" study. *J Expos Anal Environ Epidemiol* 1999: 8: 495–518.

Janssen NAH. Personal exposure to airborne particles — validity of outdoor concentrations as a measure of exposure in time series studies. PhD dissertation, Agricultural University of Wageningen, Wageningen, 1998.

Johnson TR. Recent advances in the estimation of population exposure to mobile source pollutants. *J Expos Anal Environ Epidemiol* 1995: 5 (4): 551–571.

Klepeis NE. An introduction to the indirect exposure assessment approach: modeling human exposure using microenvironmental measurements and the recent National Human Activity Pattern Survey. *Environ Health Perspect* 1999: 107 (2): 365–374.

Klepeis NE., Nelson WC., Ott WR., Robinson JP., Tsang AM., Switzer P., Behar JV., Hern SC., and Engelmann WH. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J Expos Anal Environ Epidemiol* 2001: 11 (3): 231–252.

Koistinen KJ., Hänninen O., Rotko T., Edwards RD., Moschandreas D., and Jantunen MJ. Behavioral and environmental determinants of personal exposures to $PM_{2.5}$ in *EXPOLIS*-Helsinki, Finland. *Atmos Environ* 2001: 35 (14): 2473–2481.

Kousa A., Oglesby L., Koistinen K., Künzli N., and Jantunen M. Exposure chain of urban air $PM_{2.5}$ – association between ambient fixed site, residential outdoor, indoor, workplace and personal exposures in four European cities in the *EXPOLIS*-study. *Atmos Environ* 2002: 36: 3031–3039.

Koutrakis P., Briggs SLK., and Leaderer BP. Source apportionment of indoor aerosols in Suffolk and Onondaga Counties, NewYork. *Environ Sci Technol* 1992: 26 (3): 521–527.

Kruize H., Freijer JI., Franssen EAM., Fischer PH., Lebret E., and Bloemen HJTh. Verdeling van de blootstelling aan fijn stof in de Nederlandse bevolking (in Dutch) (Report no. 263610 005). RIVM, Bilthoven, The Netherlands, 2000.

Laden F., Neas LM., Dockery DW., and Schwartz J. Association of fine particulate matter from different sources with daily mortality in six U.S. cities. *Environ Health Perspect* 2000: 108 (10): 941–947.

Law MR., Morris JK., and Wald NJ. Environmental tobacco smoke exposure and ischaemic heart disease: an evaluation of the evidence. *BMJ* 1997a: 315: 973–980.

Law PL., Lioy PJ., Zelenka MP., Huber AH., and McCurdy TR. Evaluation of a probabilistic exposure model applied to carbon monoxide (pNEM/CO) using Denver personal exposure monitoring data. *J Air Waste Manage Assoc* 1997b: 47: 491–500.

Letz R., Ryan PB., and Spengler JD. Estimating personal exposures to respirable particles. *Environ Monit Assess* 1984: 4: 351–359.

Li Ç-S. Relationships of indoor/outdoor inhalable and respirable particles in domestic environments. *Sci Total Environ* 1994: 151: 205–211.

Lioy PJ. Assessing total human exposure to contaminants — a multidisciplinary approach. *Environ Sci Technol* 1990: 7: 938–945.

Lioy PJ. Measurement methods for human exposure analysis. *Environ Health Perspect* 1995: 103 (3): 35–44.

McCurdy T. Estimating human exposure to selected motor vehicle pollutants using the NEM series of models: lessons to be learned. *J Expos Anal Environ Epidemiol* 1995: 5 (4): 533–550.

McCurdy T., Glen G., Smith L., and Lakkadi Y. The National Exposure Research Laboratory's Consolidated Human Activity Database. *J Expos Anal Environ Epidemiol* 2000: 10 (6): 566–578.

NRC (National Research Council) Research Priorities for Airborne Particulate Matter. I: Immediate Priorities and Long Term Research Portfolio. National Academy Press, Washington, DC, 1998, 127pp.

Ott WR. Exposure estimates based on computer generated activity patterns. *J Toxicol-clin Toxicol* 1984: 21 (1 and 2): 97–128.

Ott WR. Total human exposure — an emerging science focuses on humans as receptors of environmental pollution. *Environ Sci Technol* 1985: 19 (10): 880–886.

Ott W., Thomas J., Mage D., and Wallace L. Validation of the simulation of human activity and pollutant exposure (SHAPE) model using paired days from the Denver, Colorado carbon monoxide field study. *Atmos Environ* 1988: 22: 2101–2113.

Palisade Corporation. @Risk — Risk Analysis and Simulation Add-in for Microsoft Excel or Lotus 1-2-3. Palisade Corporation, Newfield, USA, 1994.

Pellizzari ED., Clayton CA., Rodes CE., Mason RE., Piper LL., Fort B., Pfeifer G., and Lynam D. Particulate matter and manganese exposures in Toronto, Canada. *Atmos Environ* 1999: 33 (5): 721–734.

Pope CA., Burnett RT., Thun MJ., Calle EE., Krewski D., Ito K., and Thurston GD. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J Am Med Assoc* 2002: 287: 1132–1141.

Rotko T., Oglesby L., Künzli N., and Jantunen MJ. Population sampling in European air pollution exposure study, EXPOLIS:

comparisons between the cities and representativeness of the samples. *J Expos Anal Environ Epidemiol* 2000a: 10 (4): 355–364.

Rotko T., Koistinen K., Hänninen O., and Jantunen M. Sociodemographic descriptors of personal exposure to fine particles ($PM_{2.5}$) in *EXPOLIS*-Helsinki. *J Expos Anal Environ Epidemiol* 2000b: 10 (4): 385–393.

Ryan PB., Spengler JD., and Letz R. Estimating personal exposures to NO_2. *Environ Int* 1986: 12: 395–400.

Ryan PB. An overview of human exposure modeling. *J Expos Anal Environ Epidemiol* 1991: 1 (4): 453–474.

Schwab M., Colome SD., Spengler JD., Ryan PB., and Billick IH. Activity Patterns applied to pollutant exposure assessment: data from a personal monitoring study in Los Angeles. *Toxicol Ind Health* 1990: 6 (6): 517–532.

Sexton K., Spengler JD., and Treitman RD. Personal exposure to respirable particles: a case study in Waterbury, Vermont. *Atmos Environ* 1984: 18 (4): 1385–1398.

Sexton K., Wagener DK., Selevan SG., Miller TO., and Lybager JA. An inventory of human exposure-related data bases. *J Expos Anal Environ Epidemiol* 1994: 4 (1): 95–109.

Spengler JD., Treitman RD., Tosteson TD., Mage DT., and Soczek ML. Personal exposures to respirable particulates and implications for air pollution epidemiology. *Environ Sci Technol* 1985: 19: 700–707.

Spix C., Anderson HR., Schwartz J., Vigotti MA., LeTertre A., Vonk JM., Touloumi G., Balducci F., Piekarski T., Bacharova L., Tobias A., Ponka A., and Katsouyanni K. Short-term effects of air pollution on hospital admissions of respiratory diseases in Europe: a quantitative summary of APHEA study results. *Arch Environ Health* 1998: 53 (1): 54–64.

Stivoro (Stichting Volksgezondheid en Roken). Jaarverslag 1998 (in Dutch). The Hague, The Netherlands, 1999.

Szalai A. (Ed.). *The Use of Time: Daily Activities of Urban and Suburban Populations in Twelve Countries*. Mouton, The Hague, Netherlands, 1972.

Vedal S. Ambient particles and health: lines that divide. *J Air Waste Manage Assoc* 1997: 47:551–581.

Zartarian VG., Ott WR., and Duan N. A quantitative definition of exposure and related concepts. *J Expos Anal Environ Epidemiol* 1997: 7 (4): 411–437.

III

Hänninen, O.O., Kruize, H., Lebret, E., Jantunen, M., 2003. ***EXPOLIS* Simulation Model: $PM_{2.5}$ Application and Comparison with Measurements in Helsinki**. *Journal of Exposure Analysis and Environmental Epidemiology* 13 (1): 74-85.

Journal of Exposure Analysis and Environmental Epidemiology (2003) 13, 74–85

www.nature.com/jea

EXPOLIS simulation model: $PM_{2.5}$ application and comparison with measurements in Helsinki

OTTO HÄNNINEN,[a] HANNEKE KRUIZE,[b] ERIK LEBRET[b] AND MATTI JANTUNEN[a]

[a] *KTL, Kuopio, Finland*
[b] *RIVM, Bilthoven, The Netherlands*

$PM_{2.5}$ exposure distributions of adult Helsinki citizens were simulated using a probabilistic simulation framework. Simulation results were compared to corresponding personal exposure distributions measured in the *EXPOLIS* study in Helsinki. The simpler models 1 and 2 (with two and three microenvironments, respectively) predict the general outline of the exposure distributions reasonably well. Compared to the observed exposure distribution, the mean is underestimated by less than 3 μg m^{-3} (20%) and the standard deviation by 23–35%. In the improved simulation models (3 and 4), the environmental tobacco smoke (ETS)-exposed subjects are excluded, the time–activity models of working and nonworking subpopulations are modeled separately, and the correlations of input concentration and time fraction variables have been accounted for. The output of these models was very close to the observed distributions; the differences in the means were less than 0.1 μg m^{-3} and the differences in standard deviation less than 1%. We conclude that when the required input data are available or can be reliably estimated, the target population $PM_{2.5}$ exposure distributions can be predicted accurately enough for most practical purposes using this kind of a microenvironment model.
Journal of Exposure Analysis and Environmental Epidemiology (2003) **13,** 74–85 doi:10.1038/sj.jea.7500260

Keywords: *exposure, model, particles, population, simulation, validation.*

Introduction

Modeling is recognized as a tool for assessing population exposures to air pollution. Exposure models allow estimation of pollutant exposure for groups of people and time periods (e.g., future) for which personal monitoring has not been conducted. Models can be used to combine information from different sources to produce estimates for population exposures that would be very expensive or impossible to measure (e.g., Letz et al., 1984).

Models should be tested against observed data to see how well the model performs in reality. Two related objectives of testing strategies usually are (1) to quantify how closely predictions match observed parameters and (2) to identify model component deficiencies that might be responsible for poor predictions (Parrish et al., 1992).

Law and Kelton (1991) identified comparison of predicted values to observed ones only as one part of validation. In their opinion (p. 299), validation is concerned with determining whether the conceptual simulation model construct is an accurate representation of the system under study. The conceptual model behind the factual implementation must match reality — in other words, phenomena truly affecting the exposures should drive the model. Koistinen et al. (2000) analyzed behavioral determinants of personal $PM_{2.5}$ exposures using *EXPOLIS* Helsinki data. Exposure to environmental tobacco smoke (ETS) was found to be the strongest single determinant of personal $PM_{2.5}$ exposures. Other yet weaker determinants were the concentrations at home and at work place. Cooking was not found to be important at population level, but this is partly due to the fact that gas stoves are rare in Helsinki. Rotko et al. (2000a) analyzed socio-economic factors connected to different $PM_{2.5}$ exposure levels. In this analysis, too, the tobacco smoke exposure was the strongest single factor, while the working status and type of occupation had weaker yet statistically significant relationships with the exposure levels.

Letz et al. (1984) compared respirable particle (RSP) exposure modeling results to personal monitoring data from Kingston-Harriman, TN, USA, for validation. The microenvironment model approach presented is also used as basis for the *EXPOLIS* simulation model (Kruize et al., 2002). Letz et al. (1984) used time-weighted averages in the analytical model, and the variances were estimated using Gauss' law of error propagation.

They used five microenvironments: outdoors, home when awake, home when asleep, other indoors, and in travel. Results were shown separately for the ETS-exposed

1. Address all correspondence to: Dr. Otto Hänninen, KTL, PO Box 95, FIN-70701, Finland. Tel.: +358-17-201-211, +358-17-201-171. Fax: +358-17-201-184. E-mail: otto.hanninen@ktl.fi
Received 1 October 2001.

Table 1. Summary of the presented simulation models 1–4.

Model number	Target population	Subgroups	Model microenvironments	Correlations	*EXPOLIS* subjects	Simulation iterations
1	*EXPOLIS* population		HI+W	no	194	2000
2	*EXPOLIS* population		HI+W+O	no	194	2000
3	Non-ETS-exposed	working/not	HI+W	yes	126	3770+570
4	Non-ETS-exposed	working/not	HI+W+O	yes	126	3770+570

Microenvironment codes: HI=home indoors, W=workplace, O=all other places.
In all models, the fitted lognormal concentration distributions were truncated at 99.9%.

and the nonexposed. The predicted means were 6–8% and standard deviations 8–14% lower than the observed values. Closer values for both mean and standard deviation were obtained for the ETS-exposed target population. Other exposure distribution characteristics, like the shapes of the simulated or observed distributions, were not reported.

In a later work, the same authors applied Monte Carlo simulation technique to the same microenvironment model for NO_2 exposures (Ryan et al., 1986), but no model validation was presented.

Ott et al. (1988) developed the SHAPE model for assessing population exposure distributions to carbon monoxide (CO). They used the Denver 1982–1983 personal monitoring data for validation. The model approach was much more detailed than in the Letz et al. (1984) study; the observed data were practically continuous and 22 microenvironments were identified from the time–activity data. Ott et al. (1988) reported the observed and simulated personal 1- and 8-h maximum exposure distributions on log–probability charts. The mean, standard deviation, and maximum values were also listed numerically for comparison. For the 8-h maximum exposures, all the three presented simulation model versions estimated the mean well (observed 4.9 ppm, composite fixed station model 4.8 ppm) but underestimated the standard deviations (observed 4.2 ppm, composite fixed station model 2.4 ppm: a 43% underestimation). For the 1-h maximum exposures, the simulated mean (10.6 ppm) was close to that observed (10.2 ppm), but the standard deviations were, again, underestimated by over 30% (observed 8.9 ppm, modeled 6.0 ppm). Ott et al. listed the finite nature of histogram distributions used in the sampling, autocorrelation of microenvironment concentrations, and serial dependencies of personal activities as possible causes for the underestimation.

Behar et al. (1990) modified the SHAPE model to simulate benzene exposures. The modified model is called Benzene Exposure Assessment Model (BEAM). Microenvironment concentrations were taken from 12-h measurements performed in the BEAM studies and the corresponding daytime and overnight exposure measurements were used for validation. The predicted and observed cumulative distributions were shown for daytime, overnight, and 22-h average exposures together with numerical values of the mean, standard deviation, and maximum. The means matched well; the predicted overnight mean was the same, while the predicted daytime and 22-h means were 16% and 8% higher than the corresponding observed values. The standard deviations in all reported simulations were — again — considerably underestimated (by 39–45%). Behar et al. explained the differences by the absence of extremes in the distributions used in sampling process.

Law et al. (1997) evaluated the probabilistic NAASQ Exposure Model applied to CO (pNEM/CO) using the same Denver 1982–1983 data as Ott et al. (1988) earlier. The simulated results, maximum 1- and 8-h running average daily exposures, were the same as in Ott et al.'s paper, but number of microenvironments was smaller (13 compared to 22). Homes with and without gas stoves were simulated separately. The target population was divided into 84 cohorts according to home and work districts, demographic groups, and cooking fuel used at home.

The simulation results were compared to observed exposure distributions by plotting the cumulative 1- and 8-h maximum exposure distributions on log–probability charts and by tabulating eight percentile values for comparison. The median values were slightly (3–4%)

Table 2. Summary of time activity distributions (fractions of time) used in the simulations (f = fraction of time, HI = home indoors, WI = work indoors, O = other).

Variable	Whole *EXPOLIS* target population				
	n	Mean	SD	Min	Max
f(HI)	434	0.576	0.126	0.16	1.00
f(WI)	434	0.248	0.130	0.00	0.51
f(O)	434	0.176	0.103	0.00	0.70
Working subpopulation					
f(HI)	377	0.544	0.091	0.16	0.86
f(WI)	377	0.285	0.093	0.00	0.51
f(O)	377	0.171	0.099	0.02	0.70
Nonworking subpopulation					
f(HI)	57	0.789	0.123	0.42	1.00
f(WI)	57	0.000	0.000	0.00	0.00
f(O)	57	0.211	0.123	0.00	0.58

Table 3. Summary of *EXPOLIS* exposure data used in comparison and concentration data used as simulation inputs.

Variable	Whole *EXPOLIS* population					Non-ETS-exposed				
	n	Mean	SD	Min	Max	*n*	Mean	SD	Min	Max
Personal, 48 h[a]	194	15.4	18.8	1.7	177.4	126	9.8	6.4	1.7	37.4
Home indoors[b]	192	12.2	15.1	1.8	121.6	126	8.9	5.7	1.9	26.7
Work indoors[b]	151	15.9	34.9	0.5	280.2	98	9.7	10.0	1.4	82.6
Ambient, 1 h[b]	6436	9.6	6.8	0.1	128.4	6436	9.6	6.8	0.1	128.4

[a]Distribution used in comparison.
[b]Distribution used as simulation input.

underestimated for the 1-h exposures and overestimated for the 8-h exposures (7% for nongas stove homes, 16% for gas stove homes). All models also overestimate the 5th percentile level by 40% or more and underestimate the 95th percentile level by 24–67%. Thus, the variation was once more underestimated in the simulated results, although direct variation measures were not reported.

Law et al. (1997) list four possible reasons for the discrepancy between simulated and observed values: (i) only two (gas stove and smoking) of all known indoor sources (wood stove, kerosene heaters, water heaters, fire places, and garages) of CO were included in the model; (ii) the population time–activity autocorrelations were not modeled; (iii) the time–activity database (Washington, DC) was from different area than the validation data (Denver); and (iv) the model may under predict high exposures due to the constant values used in mass balance model and other empirical pNEM/CO model parameters.

Law et al. (1997) also calculated modified Kolmogorov–Smirnov (K–S) statistics to test the differences between the observed and simulated distributions. All simulated distributions except the 1-h exposures of subjects with non-gas stove homes were rejected in the K–S statistical tests using 5% risk level (P=0.05). Thus, according to the K–S statistics, the simulated exposure distributions are not similar to the measured ones.

In summary, all the reviewed simulation model validations predicted the mean or median values with fair to good accuracy, but all underestimated the exposure variability. According to the authors, there is a need to improve the model performance especially in the high-end exposure levels.

Kruize et al. (2002) described the simulation framework developed within the *EXPOLIS* project. As described in Jantunen et al. (1998), the general objectives of the probabilistic exposure simulation development in *EXPOLIS* are to assess the population exposure simulations of selected subpopulations, urban areas, and future scenarios. The aims of the current paper are (1) to use the model to simulate Helsinki adult population exposures to $PM_{2.5}$; (2) to compare the simulations to observed exposure distributions; and (3) to evaluate model components to identify possible model development needs.

Materials and methods

The structure of the *EXPOLIS* exposure simulation framework used in this work is presented in detail by Kruize et al. (2002). The model is based on average concentration experienced over visits to multiple environments (Fugas, 1975). The microenvironment formulation of the approach is shown in Eq. (1) (e.g., Duan 1982; Letz et al., 1984; Ryan et al., 1986):

$$E = \sum_i f_i C_i \quad (1)$$

where E is the time-weighted average exposure of an individual and f_i and C_i are the time fraction spent in and concentration in microenvironment i. This equation is applicable for any averaging time and any number of microenvironments and can, in principle, be used for any air

Table 4. Pairwise rank correlation coefficients between input time fraction and concentration variables in models 3 and 4.

Time–activity			
	Home	Work	Other
Working subjects			
Home	1		
Work	−0.37	1	
Other	−0.52	−0.50	1
Nonworking subjects			
Home	1		
Work	N/A	N/A	
Other	−1.00	N/A	1
Concentrations			
	Home	Work	Ambient
Non-ETS-exposed subjects			
Home indoors	1		
Work indoors	0.40	1	
Ambient	0.83	0.57	1

pollutant. In a model run, E is calculated for a large number of simulated individuals, based on a random drawing of input variables f_i and C_i from specified probability distributions. The modeling framework described by Kruize et al. (2002) allows also for a nested approach, where the Eq. (1)-based microenvironment model is supplemented with (i) modeling of indoor microenvironment concentrations using ambient concentrations and probabilistic penetration factor following beta distribution, and (ii) modeling of indoor emissions using probabilistic lognormal emission factors. This nested approach is not used in this work.

The *EXPOLIS* Helsinki $PM_{2.5}$ exposure and microenvironmental concentration measurement data are used to test the applicability of microenvironment-based simulations to $PM_{2.5}$ exposures. The *EXPOLIS* study is described in detail in Jantunen et al. (1998) and the $PM_{2.5}$ sampling methods and data quality in Koistinen et al. (1999). The

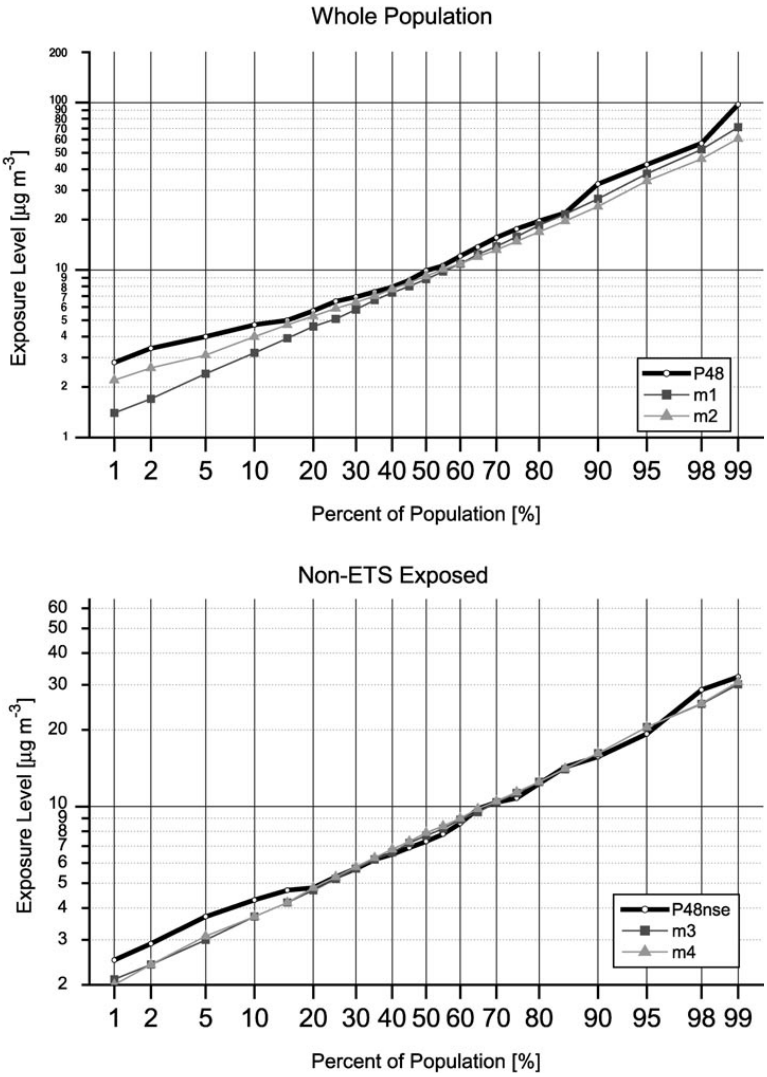

Figure 1. Comparison of simulated and observed exposure distributions (P_{48} = observed 48-h exposure to $PM_{2.5}$; NSE = non-ETS-exposed subjects only).

measurements were carried out in 1996–1997. The population sampling procedures and sample comparison to the whole metropolitan area population are presented by Rotko et al. (2000b). The randomly drawn *EXPOLIS* population sample in Helsinki consisted of Finnish-speaking 25- to 55-year-old Helsinki Metropolitan area residents.

$PM_{2.5}$ concentrations at home indoors, home outdoors, and at work place, and personal exposures during 48 h were measured for 201 subjects. The personal monitor briefcase was carried by the subject and kept in the vicinity for 48 h. Because the measurement did not catch true exposure to active smoking, the exposure levels of active smokers are processed here as exposures to ETS. The residence and work microenvironment measurements were programmed to occur during the actual hours that the participants were expected to spend there. A 15-min resolution time–microenvironment–activity diary (TMAD) with 11 microenvironments and three activities were kept during the measurements and all subjects filled a detailed microenvironment, behavioral, and socio-demographic questionnaire. A larger population sample participated in a questionnaire-only study by filling the diaries and questionnaires. The measurement and questionnaire data were stored into the *EXPOLIS* Access Database (EADB; Microsoft, Seattle, WA, version 7.0) developed for this purpose (Hänninen et al., 2002).

Four simulation runs, listed in Table 1, were performed using the Risk add-on software (Palisade, Newfield, NY, version 4.0) with Excel (Microsoft version 8.0). All simulations were run using Latin hypercube sampling. Models 1 and 2 targeted the whole *EXPOLIS* population while models 3 and 4 excluded the ETS-exposed subjects from the target population and used separate time–activity models for working and nonworking subpopulations. Models 1 and 3 were built on two microenvironments, "Home indoors" (HI) and "Work indoors" (WI), and used the exposure equation $E = f_{HI}C_{HI} + f_{WI}C_{WI}$. Models 2 and 4 added the "Other" (O) microenvironment by lumping all other nine diary microenvironments together. The exposure equation for these models was $E = f_{HI}C_{HI} + f_{WI}C_{WI} + f_{O}C_{O}$.

The time–activity parameters used here are calculated from all of the diaries, including both the exposure measurement sample and the questionnaire-only sample (total $n=434$). The time fractions were calculated for the two or three microenvironments and beta distributions were fitted on these data. Goodness of the fits were evaluated by plotting observed histograms overlaid with the corresponding fitted beta density function and by calculating K–S, Anderson–Darling, and Chi-square test statistics (using Risk 4.0 software). The time–activity simulation inputs as fractions of time spent in each microenvironment are listed in Table 2.

The concentration input parameters were obtained from the *EXPOLIS* "Home indoors" and "Work indoors" microenvironment measurements. Measurement repeatability for the *EXPOLIS* concentrations was 3% (relative standard deviation) (Koistinen et al., 1999). Hourly measured ambient $PM_{2.5}$ data were used as the concentration distribution for the "Other" microenvironment. The Helsinki Metropolitan Area Council carried out the ambient measurements and provided the data. The simulation inputs (means, standard deviations) and the numbers of observations of each type are listed in Table 3, including observed exposure values. The lognormality of the concentration distributions was tested using Shapiro–Wilk's test. Two-parameter lognormal distributions were fitted to the concentration data. Goodness of the fits were evaluated visually using function overlays on a histogram and statistically by t-test, Wilcoxon rank-sum test, and K–S test statistics for each fit (STATA 5.0 statistical software; STATA, College Station, TX).

All lognormal concentration distributions were truncated at 99.9th percentile in the simulations to prevent unrealistically high concentration values. Spearman's rank correlation (r) matrixes for the time fraction and concentration variables were used in models 3 and 4. The rank correlation inputs are shown in Table 4.

Results

Comparison of Model Outputs to Observed Exposure Distributions

Simulated exposure distributions are compared graphically to the observed ones in Figure 1. Models 1 and 2, targeting the whole *EXPOLIS* population (Finnish-speaking 25- to 55-year-old persons), are shown in the top chart. In models 3 and 4 in the second chart, the ETS-exposed subjects have been excluded, and the models have been enhanced with separate time–activity models for working and nonworking subpopulations and by taking the rank correlations within fraction of time and concentration variables into account.

Table 5. Observed and simulated population exposure distribution values [$\mu g\ m^{-3}$].

	Whole *EXPOLIS* population			Non-ETS-exposed subpopulation		
	Observed	Model 1	Model 2	Observed	Model 3	Model 4
Mean	15.4	13.2	12.6	9.2	9.1	9.2
SD	18.8	14.4	12.3	5.8	5.8	5.8
Percentiles						
25%	6.5	5.1	5.9	5.3	5.2	5.3
50%	9.9	8.8	9.2	7.3	7.7	7.9
75%	17.6	15.8	14.8	10.8	11.3	11.4
90%	32.6	26.6	23.9	15.7	16.2	16.2
95%	42.6	37.5	34.0	19.3	20.5	20.5

Each chart shows two simulations: one with two and the other with three microenvironments. Visually evaluated, all the simulated distributions are roughly similar to the observed ones.

Numerical comparisons of the mean, standard deviation, and percentiles of the observed and simulated distributions are shown in Table 5. Models 1 and 2 slightly underestimate the mean and all percentiles. The three-microenvironment model (model 2) is closer to observed values in low percentiles while the two-microenvironment model (model 1) is performing slightly better in the high end of the distribution. Both models clearly underestimate the standard deviation. Models 3 and 4 estimate the means and standard deviations very close and almost identical to the observed values. All modeled percentiles are closer to the observed values than in models 1 and 2, especially in the high end of the distributions. The lowest percentiles are still underestimated, but the absolute differences are less than 1 $\mu g\,m^{-3}$.

Agreement of the simulated distributions with the observed exposures was statistically tested using the t-test, Wilcoxon rank-sum test, and K–S test statistics. The K–S P values were 0.01, 0.11, 0.53, and 0.69 for models 1–4, respectively. The t-test and Wilcoxon rank-sum tests produced similar results.

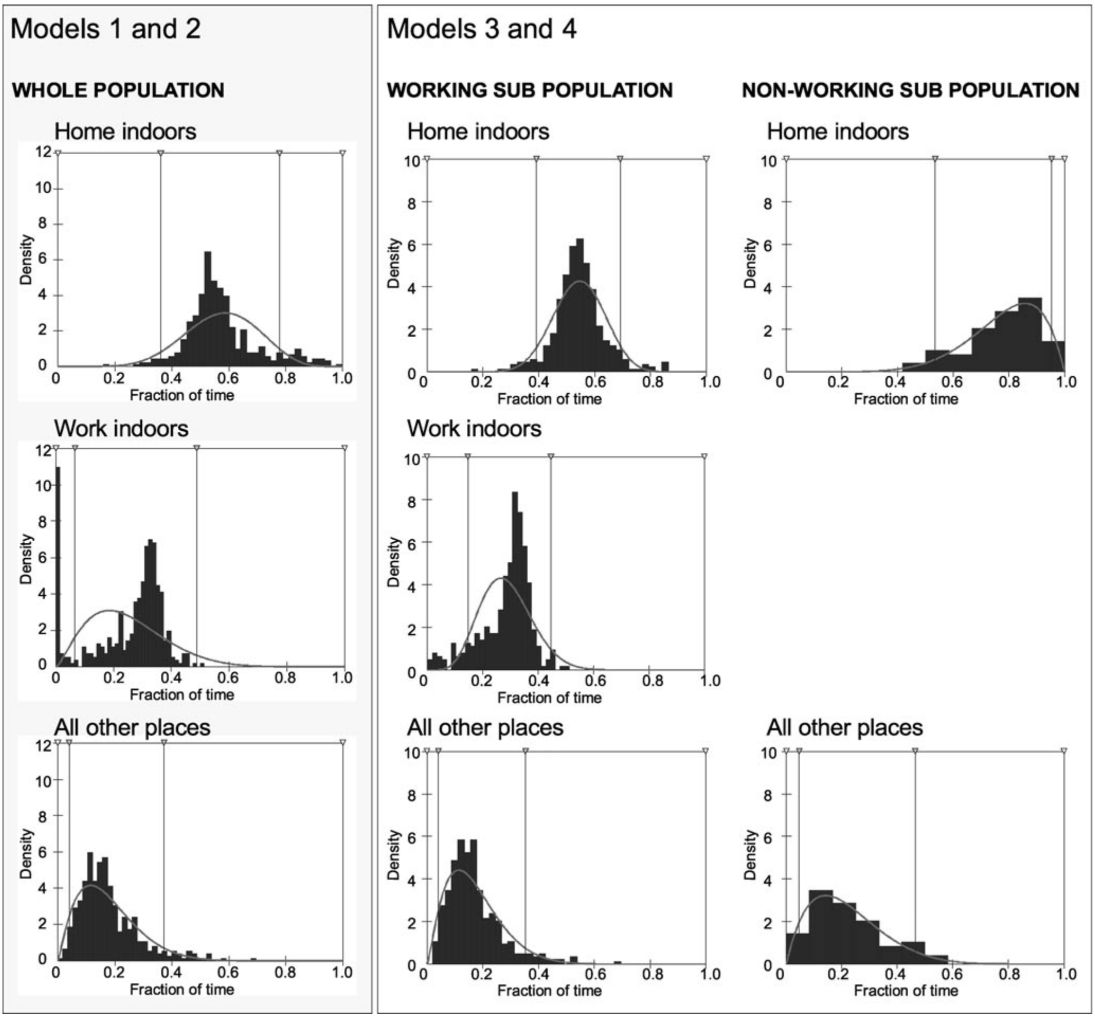

Figure 2. Fitted beta functions (lines) compared to observed time activity distributions (bars).

Evaluation of Model Components

Fitted Time–Activity Functions Observed fractions of time histograms are shown with corresponding beta density functions in Figure 2. Different subpopulations are shown in the columns and microenvironments in the rows. The fitted beta function for home indoors had a similar range as the observed data, with central tendency close to the observed. The observed data, however, show a more pronounced peak for the whole and the working populations. The K–S statistics (P values) testing the goodness of the fits are 0.000, 0.002, and >0.25 for the whole, the working, and the nonworking populations, respectively.

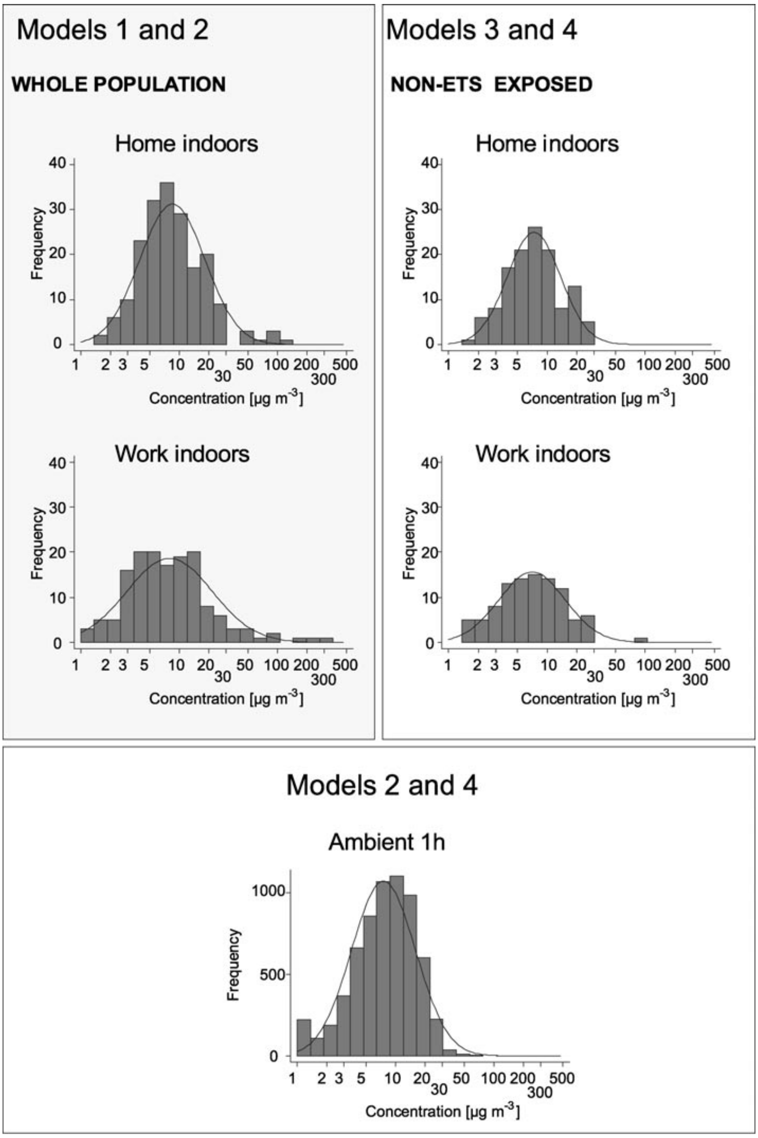

Figure 3. Histograms and fitted lognormal distributions for the concentration inputs. Ambient concentration, used for the "Other" microenvironment in models 2 and 4, is not affected by ETS exposure status of the subjects.

By definition, the nonworking subpopulation, with fraction of time spent at work being zero, can be seen in the leftmost bar of the histogram. This high bar forces the beta fit leftwards, causing the mode to be misplaced compared to the observed data. The beta function properties do not reflect the working day length well. Even the longest working individuals in the observed data do not spend more than approximately 50% of their time at work. The normal workday length, 8 h, is the mode of the observed data. Thus, the observed histogram is skewed to the left, while the fitted beta function is skewed to the right. The histogram for the working subpopulation is identical to the whole population, except that the zero time bar has been removed. Now the fit is clearly better, but it still overpredicts short and long workdays with a mode closer to 6 than 8 h. This indicates that the beta fit, so applied, is not very good for the work microenvironment. The K–S statistics (P values) are 0.000 for both the whole and the working populations, indicating that statistically, the fitted and observed distributions are different.

For the sum of time fractions spent in all other environments (than in "Home indoors" and "Work indoors"), the beta fit is quite good for all three populations. It can be expected that beta fit works well for most minor microenvironments, where the distribution mode is close to zero and the distribution is skewed to the right. The K–S statistics (P values) are 0.01, 0.01, and >0.25 for the whole, the working, and the nonworking populations, respectively. This indicates that the fits for the first two populations are still statistically poor but that the last one is good.

Fitted Concentration Functions Concentration histograms measured in the Helsinki *EXPOLIS* study are shown in Figure 3. The fitted lognormal density functions are plotted on logarithmic x-axis scale. The concentrations in the "Home indoors" and "Work indoors" of the whole Helsinki study population (models 1 and 2) are shown in the left column of figures. Observed data for the non-ETS-exposed subpopulation (models 3 and 4) are shown in the right column. The single chart on the third row of figures shows the distribution of 1-h ambient concentration used as the concentration distribution for the "Other" microenvironment in models 2 and 4.

The overall visual appearances of the fits are good. Shapiro–Wilk's test results for lognormality, however, were poor. The P values were <0.00 for all distributions for the whole population and ambient 1-h data. The P values for the non-ETS-exposed subpopulation were 0.20 ("Home indoors") and 0.64 ("Work indoors"), indicating statistically acceptable fits. This indicates that the most important cause for the poor fit of the distributions for the whole population is smoking; the ETS-exposed indoor environments appear in the concentration distribution as weak but statistically evident multimodality.

Detailed inspection of the simulated concentrations (data not shown) reveals that in some cases, the open-ended nature of lognormal distribution (from zero to infinity) does not describe the range of realistic concentrations. In models 1 and 2, the highest simulated concentrations without truncation exceeded 5000 μg m^{-3}, 20 times the observed maximum. The problem was reduced in models 3 and 4, but the simulated maximum concentrations were still clearly higher than the observed maxima in these models, too. Latin hypercube sampling, used in the simulations, highlights this problem by ensuring that one sample is taken from the extreme of each distribution.

Intercorrelation of the Simulation Inputs Pairwise rank correlation coefficients within the observed fraction of time and concentration variables are shown in Table 4. The fractions of time correlations are negative. In the three-microenvironment model for the nonworking subjects (model 4), the correlation between the fraction of time spent in "Home indoors" and in the "Other" microenvironments is -1.00, by definition.

The rank correlation between the "Home indoors" concentration and the "Work indoors" concentration of the same person is lower (r=0.4) than the correlations of both of these indoor microenvironments with simultaneous ambient concentrations. The rank correlation of "Home indoors" for the non-ETS-exposed subjects with the ambient concentration was 0.83.

Discussion

Comparing Simulation Results to Literature
In the validations of simulated exposures in the reviewed literature, the mean or the median values usually matched the observed ones quite well. Our own results for estimating the means are similar. Models 1 and 2 underestimate the means by 15% and 18%, respectively, but models 3 and 4 come very close to the observed values, the relative difference being less than 1%.

The variance was usually underestimated in the cited studies. In the current work, models 1 and 2 behave similarly. The standard deviations were underestimated by 23% (model 1) and 35% (model 2). Adding the correlation matrix to model 2 improved the standard deviation estimate only slightly (underestimation by 27%, data not shown), and did not affect the main problem of underestimating all the levels. Our models 3 and 4 outperformed the earlier models by predicting standard deviations within 1% of the observed values. This is at least partly explained by (i) incorporation of concentration intercorrelations in these models, and (ii) the use of factual microenvironment concentration distributions measured in times when people were truly present.

The simulated standard deviations were quite sensitive to the simulated maximum concentrations. Without truncation at 99.9th percentile, the standard deviation of model 1 was four times of the observed (over estimation 400%, data not shown). Truncation at 99th percentile led to 10% underestimation in the standard deviation (data not shown). The effect was much smaller in models 3 and 4.

All the reviewed models in the literature have been built on five or more microenvironments, while the current models have only two or three. This indicates that at least for a pollutant like $PM_{2.5}$ — having very small outdoor spatial variations and with major indoor source ETS excluded — a very simple microenvironment model can work remarkably well. For the whole population, the three-microenvironment model performed slightly better; but for the non-ETS-exposed, there was practically no difference between the two- and three-microenvironment models.

Evaluation of the Goodness of Fit

Classical statistical tests were used in combination with graphical comparison of the model outputs to observed distributions to evaluate the parametric input distribution fits. As discussed by, e.g., Firestone et al. (1997), the statistical power of the tests increases with the number of data points. In a simulation with a large number of iterations, the possibility to find small but statistically significant differences between distributions increases. The overall shape of a fitted distribution can be similar to the observed one and the values fit the same ranges, but statistically, the distributions are still different. On the other hand, the input distribution can be found to be lognormal using a statistical test, but using the fit as input in a Latin hypercube sampling model may produce unrealistically high concentration inputs. Both of these cases are demonstrated in the results.

The statistical tests reduce the goodness-of-fit evaluation to a single P value, but the interpretation of such a number may be difficult. Graphically, the fit might look acceptable, having approximately the same range of values with roughly the same shape, while the statistical test indicates a poor fit. The tests assess the probability that the observed values have been sampled from the fitted distribution. They do not evaluate whether the difference would invalidate the fitted distribution. Thus, the result of a statistical test should neither be the only reason to accept nor reject a particular fit or output.

Firestone et al. (1997) emphasize the use of graphical comparisons in assessing the goodness of fits. Different formats of graphical distribution comparisons have different benefits. The log–probability plots used in this work and in the literature (e.g., Ott et al., 1988; Behar et al., 1990; Law et al., 1997) show the relative goodness of the fits clearly. On the other hand, the logarithmic y-axis scale partly masks the absolute differences at the high end. Our work follows Law et al. (1997) and lists the main percentiles numerically for direct (linear) comparison (Table 5).

In the Helsinki metropolitan area, there are approximately 440,000 people in the age range 25–55 years. The corresponding *EXPOLIS* measurement sample was only 201 persons (0.05 %). Each sampled subject represents 2189 persons in the target population. The highest indoor concentration observed was close to 300 $\mu g\ m^{-3}$ (Table 3). Many other studies have reported higher indoor particle levels in smoker's homes, e.g., in the Netherlands, weekly average indoor RSP levels in the winter period were 400–500 $\mu g\ m^{-3}$ (Lebret, personal communication). Thus, the observed maximum in Helsinki may be an underestimation of the true extreme. The simulated concentration level of 5000 $\mu g\ m^{-3}$ would, however, be unrealistic due to the visual, olfactory, and irritative properties of tobacco smoke. The reliability of the measured extreme exposures to represent the true extreme exposures within this population is equally uncertain as the reliability of the modeled extremes. Above 95th percentile, they both have upwards-increasing uncertainty, which should be kept in mind when comparing the simulated and observed levels beyond these percentiles.

The uncertainty and the variability were not modeled separately in this work. The concentration and exposure measurements were carried out with a 3% relative precision. Repeatability of the time–activity measurement by participant-applied time–microenvironment–activity diaries has not been assessed by us or others. One can only theorize that for a 6- to 8-h working day, the diary error might be 30 min, which is less than 10%. For the fraction of time spent at home, the relative error is probably smaller. For microenvironments where the time spent is shorter, the relative error would increase. These measurement errors slightly increase the observed variance in both the observed exposure distributions and in the simulation outputs. In both cases, most of the variance seems to be real.

Modeling Input Concentrations

Environmental pollutant concentrations have often been found to follow lognormal distribution (e.g., Ryan et al., 1986; Ott, 1990; Ott et al., 1988). The microenvironment concentrations, used as inputs for "Home indoors" and "Work indoors" in models 1 and 2, visually appeared close to lognormal. The overall population concentration distributions were in fact multimodal due to the combination of ETS-exposed and nonexposed subjects, and Shapiro–Wilk's statistical test rejected the lognormality at 1% confidence level. This may be one explanation for the poorer results for models 1 and 2. In models 3 and 4, the ETS-exposed subjects were excluded, and the microenvironment concentrations were lognormal.

Ott (1990) used computer simulations to show how concentrations after random dilution process followed roughly lognormal distributions. At the highest percentiles, however, the concentrations fell below the lognormal ones. In Ott's results, this starts to appear above 98th percentile and becomes even stronger above 99th percentile. Ott simulated the levels up to the 99.9th percentile. This feature and its consequences were well demonstrated in the current work. Using Latin hypercube sampling with 2000 or more iterations, the lognormal concentration inputs produced concentrations in excess of 5000 μg m^{-3}. The highest measured indoor concentrations in *EXPOLIS* Helsinki data were below 300 μg m^{-3} even with ETS exposures included and below 90 μg m^{-3} when ETS exposures were excluded. While it is conceivable that such extreme concentrations could sometimes occur in Helsinki, we are not implying that our model could estimate them any more than randomized *EXPOLIS* measurements could capture them. Thus, the concentration model was enhanced with truncation. Concentrations exceeding the 99.9th percentile level were truncated to that level. The selection of the truncation percentile affects the simulated standard deviations (a lot in extreme cases), and might have a minor effect on the mean values, too. It has no effect on the exposure percentiles below the 99th one.

Besides the selected lognormal distributions to model indoor concentrations, other distribution shapes could be considered. Lognormal distribution is, however, the most common distribution used to model concentration in the literature. Ott's (1990) work has shown that there, in fact, is a physical explanation to the fact that many observed concentration distributions appear to be approximately lognormal. Thus, in the current work, the lognormal distribution was used as a default distribution shape for the concentrations. The results indicate that despite of the observed deviancies from the lognormality, this assumption works fine in the current models.

Time–Activity Inputs

The current work uses modeled time–activity inputs. As suggested by Ryan et al. (1986), fitted beta distributions were used to describe the time fractions spent in each microenvironment. Beta distribution is flexible, is limited by definition to the range [0,1] as a time fraction parameter must be, and allows for symmetry or skewness to the left or to the right. Time–activity diaries of true persons have also been used as time–activity inputs for simulation modeling (e.g., Ott et al., 1988; Behar et al., 1990; Law et al., 1997; Freijer et al., 1998). The SHAPE model was originally developed to use modeled time activities, but the versions used in the CO (Ott et al., 1988) and benzene validations (Behar et al., 1990) were modified to use actual diary data as input instead. Ott et al. (1988) point out that also the US Science Advisory Board panel members suggested the use of actual diaries instead of time–activity models to avoid errors from overlooked auto- and intercorrelation structures of time–activity variables.

By definition, the fractions of times spent in different microenvironments must, in general, be inversely related. The time spent in any one microenvironment reduces the time available for all others. In the current *EXPOLIS* simulation framework, the time–activity model has been built so that the individual fractions of time are sampled from the fitted beta functions of each microenvironment, and are then divided by the total of the sampled fractions. This effectively scales the used time to unity — an important and necessary property of any time fraction model. But this approach also has the problem that after the division, the sampled time activities do not follow anymore the original fitted beta function. Using negative correlations slightly helps in this problem, the total sampled time being in average closer to unity before the division.

There is no natural reason for fractions of times spent in microenvironments to follow beta distribution. Ryan et al. (1986) selected beta distribution as the preferred function because it is bounded by zero from below and one from above — matching the definition of fraction of time variable. The goodness-of-fit plots and tests for β fits showed that the fits are not ideal even in the best cases. One clear reason for this is the difference in time patterns between subgroups of people, e.g., the working and nonworking. We did not explore further subdivision into subgroups as a means of improving the fits. If the time–activity model would need to be improved, then maybe the use of experimental density functions could be considered. But in simulations with more than few microenvironments, this opens the question as to whether the kind of simple correlation matrix approach used here is sufficient.

In spite of the theoretical problems in the beta fit-based time–activity model discussed above, the simulation outputs compare well to the observed exposures. This suggests that at least for a pollutant like $PM_{2.5}$, the beta fit model is good enough.

ETS Exposure

It is important to realize that neither the measured nor modeled exposures include active smoke inhalation by smokers. Because only ETS was included in the sampling, both passive and active smoker measurements reflect the exposure of passive smoking, ETS, only. The true exposures of active smokers are much higher.

The simulations shown in this paper clearly confirm the general finding in many studies (e.g., for Helsinki, Koistinen et al., 2000, Rotko et al., 2000a) that ETS exposure greatly affects total population exposures to $PM_{2.5}$. The simulations including ETS-exposed subjects underestimated the exposure levels. This is probably caused by the fact, that the fixed point home indoor and workplace measurements catch only part of the ETS exposure. In Helsinki, smoking in the workplaces is very often restricted to special rooms, and even at home, many people smoke only outdoors. Adding the concentration correlations to the ETS-included models improves the standard deviation estimates, but does not solve the underestimation of the levels. Thus, models 1–4, not including the microenvironments important for ETS exposures (models with only "Home indoors" and "Work indoors") or lumping these microenvironments together with many other types of microenvironments (models with the "Other" microenvironment), are not suitable for modeling ETS exposures. ETS exposures should be specifically handled in a model, or should just be excluded from the simulation. The nested design of the modeling framework described by Kruize et al. (2002) allows also for modeling of ETS and other indoor source exposures. Exclusion of ETS, selected in this work, is justified when the focus is in pollution with ambient origin.

Were there other specific indoor sources than smoking, similar analysis of input representativeness should be carried out to assess the model applicability.

Limitations of the presented model include the need for microenvironment concentration distributions for home and work place indoor air, optionally with the correlations with each other and the ambient air. The concentration submodel needs to be developed further to allow the use ambient air quality measurements to model the microenvironment $PM_{2.5}$ input concentrations.

Uncertainty in the Models

When a model is developed to assess exposure levels in a hypothetical scenario, it is important to assess also the uncertainties in the model outputs, caused by uncertainties in model structure, exposure scenario, and model parameters. Model uncertainty includes uncertainties in the selection of the distributions (different parametric vs. empirical), methods of fitting the parameters, definition of the microenvironments and modeled activities, selection of averaging times and number of iterations, and generation of the random numbers, and so forth (Morgan and Henrion, 1990; Cullen and Frey, 1999).

The microenvironment model, shown in Eq. (1) earlier, is in fact the definition of exposure as time-weighted average over the microenvironments visited. Thus, in this basic equation of the model, there is no uncertainty. The simplifications used in the selection of microenvironments and the selection of parametric distributions, however, introduce uncertainties to the model structure.

Model uncertainty has been examined to some extent by evaluation of the input distributions and by comparing the two- and three-microenvironment models with and without ETS exposure. As an example, missing such microenvironments as bars and smoking lounges in the model is obviously one reason for the poor results in simulation of the exposures of the smokers and ETS-exposed. Full analysis of the model uncertainty would significantly broaden the focus and volume of this article. Because the comparison of the modeled and measured exposures for the non-ETS-exposed subjects shows that there is little remaining uncertainty to be explained, this analysis will not be pursued here.

Assessment of scenario uncertainty is crucially important when a model is applied into a new setting (scenario is changed). In the present comparison study, the modeled and measured scenarios are identical and, thus, scenario uncertainty has been removed by the study setup. Using the same population sample for both model inputs and comparison data also removes the biggest source of parameter uncertainty, namely population sampling. Thus, only measurement errors are causing parameter uncertainty in the presented models. According to the quality assurance results published elsewhere (Koistinen et al., 1999), the effect of measurement errors seems to be negligible and was considered to be out of the scope of the current work.

Conclusions

The probabilistic two- to three-microenvironment simulation model predicts the population $PM_{2.5}$ exposures fairly well. When ETS exposures were excluded and correlations between the input variables were taken into account, the match was very good over the whole distribution for this subgroup.

With ETS-exposed subjects included and ignoring the input correlations, the simulation outputs underestimated the exposures at the mean and all percentile levels. To solve this problem, ETS exposure should be specifically modeled, e.g., using a nested model. Ignoring the input correlation matrix leads to underestimation of the exposure variance.

The beta distribution model used for time–activity performed acceptably, in spite of the deviations in the fits compared to the observed distributions. Lognormal fits for microenvironment concentrations had to be truncated at 99.9th percentile to prevent overestimation of exposure variation.

Acknowledgments

This work has been supported by EU Contract N ENV4-CT96-0202 (DG12-DTEE); Academy of Finland Contracts 36586, 40835, and 42610; and intramural funding by KTL, the National Public Health Institute of Finland, and other institutions in other centers.

References

Behar J.V., Thomas J., and Pandian M.D. Development of the Benzene Exposure Assessment Model (BEAM). In: *Total Exposure Assessment Methodology. Proceedings of the EPA/A&WMA Specialty Conference, Las Vegas, NV, November 1989.* Air and Waste Management Association, Pittsburgh, USA, 1990, pp. 436–450.

Cullen A.C., and Frey H.C. *Probabilistic Techniques in Exposure Assessment.* Plenum, New York, 1999, 335 pp.

Duan N. Models for human exposure to air pollution. *Environ Int* 1982: 8: 305–309.

Firestone M., Fenner-Crisp P., Barry T., Bennet D., Chang S., Callahan M., Burke A.M., Michaud J., Olsen M., Cirone P., Barnes D., Wood W., and Knot S. *Guiding Principles for Monte Carlo Analysis. EPA Technical Panel Report* 1997; EPA/630/R-97/001; 35 pp.

Freijer J.I., Bloemen H.J.T.H., de Loos S., Marra M., Rombout P.J.A., Steentjes G.M., and van Veen M.P. Modelling exposure of the Dutch population to air pollution. *J Hazard Mater* 1998: 61: 107–114.

Fugas M. Assessment of total exposure to an air pollutant. *Proceedings of the International Conference on Environmental Sensing and Assessment* 1975; Vol. 2, Paper No. 38-5, IEEE #75-CH 1004-1, ICESA.

Hänninen O., Alm S., Kaarakainen E., and Jantunen M. The EXPOLIS Databases. Publications of the National Public Health Institute, BB/2002, KTL, Kuopio. ISBN: 951-740-321-6 (printed), ISBN: 951-740-322-4 (pdf), ISSN: 0359-3576; 408 pp.

Jantunen M.J., Hänninen O., Katsouyanni K., Knöppel H., Kuenzli N., Lebret E., Maroni M., Saarela K., Sram R., and Zmirou D. Air pollution exposure in European cities: the "Expolis" study. *J Expos Anal Environ Epidemiol* 1998: 8(4): 495–518.

Koistinen K.J., Kousa A., Tenhola V., Hänninen O., Jantunen M.J., Oglesby L., Kuenzli N., and Georgoulis L. Fine particle ($PM_{2.5}$) measurement methodology, quality assurance procedures, and pilot results of the EXPOLIS study. *J Air Waste Manage Assoc* 1999: 49: 1212–1220.

Koistinen K.J., Hänninen O., Rotko T., Edwards R.D., Moschandreas D., and Jantunen M.J. Behavioral and environmental determinants of personal exposures to $PM_{2.5}$ in EXPOLIS-Helsinki. *Atmos Environ* 2001: 35: 2473–2481.

Kruize H., Hänninen O., Breugelmans O., Lebret E., and Jantunen M. Description and demonstration of EXPOLIS simulation model: two examples of modeling population exposure to particulate matter. *JEA* 2002 (in print).

Law A.M., and Kelton W.D. Simulation modeling and analysis. McGraw-Hill international editions. In: *Industrial Engineering Series* 1991, McGraw-Hill, Singapore. ISBN 0-07-100803-9; 759 pp.

Law P.L., Lioy P.J., Zelenka M.P., Huber A.H., and McCurdy T.R. Evaluation of a probabilistic exposure model applied to carbon monoxide (pNEM/CO) using Denver personal exposure monitoring data. *J Air Waste Manage Assoc* 1997: 47: 491–500.

Letz R., Ryan B.P., and Spengler J.D. Estimated distributions of personal exposure to respirable particles. *Environ Monit Assess* 1984: 4: 351–359.

Morgan M.G., and Henrion M. *Uncertainty. A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, New York, 1990, ISBN 0-521-42744-4; 332 pp.

Ott W.R. A physical explanation of the lognormality of pollutant concentrations. *J Air Waste Manage Assoc* 1990: 40: 1378–1383.

Ott W., Thomas J., Mage D., and Wallace L. Validation of the Simulation of Human Activity and Pollutant Exposure (SHAPE) model using paired days from the Denver, CO, carbon monoxide field study. *Atmos Environ* 1988: 22: 2101–2113.

Parrish R.S., Smith C.N., and Fong F.K. Tests of the pesticide root zone model and the aggregate model for transport and transformation of aldicarb, metolachlor and bromide. *J Environ Qual* 1992: 21: 685–697.

Rotko T., Koistinen K., Hänninen O., and Jantunen M. Sociodemographic descriptors of personal exposure to fine particles ($PM_{2.5}$) in EXPOLIS-Helsinki. *J Expos Anal Environ Epidemiol* 2000a: 10(4): 385–393.

Rotko T., Oglesby L., Künzli N., and Jantunen M. Population sampling in European air pollution exposure study, EXPOLIS: comparisons between the cities and representativity of the samples. *J Expos Anal Environ Epidemiol* 2000b: 10(4): 355–364.

Ryan P.B., Spengler J.D., and Letz R. Estimating personal exposures to NO_2. *Environ Int* 1986: 12: 395–400.

IV

Hänninen, O.O., Lebret, E., Ilacqua, V., Katsouyanni, K., Künzli, N., Srám, R., Jantunen, M.J., 2004. **Infiltration of ambient $PM_{2.5}$ and levels of indoor generated non-ETS $PM_{2.5}$ in residences of four European cities**. *Atmospheric Environment*, 38 (37): 6411-6423.

ELSEVIER

AE International – Europe

Atmospheric Environment 38 (2004) 6411–6423

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

Infiltration of ambient $PM_{2.5}$ and levels of indoor generated non-ETS $PM_{2.5}$ in residences of four European cities

O.O. Hänninen[a,*], E. Lebret[b], V. I1acqua[a], K. Katsouyanni[c], N. Künzli[d,f], R.J. Srám[e], M. Jantunen[a]

[a]KTL, Department of Environmental Health, Kuopio, Finland
[b]RIVM, Bilthoven, The Netherlands
[c]University of Athens, Medical School, Athens, Greece
[d]University of Basle, Institute of Social and Preventive Medicine, Basel, Switzerland
[e]Institute of Experimental Medicine, AS CR, and Health Institute of Central Bohemia, Prague, Czech Republic
[f]University of Southern California, Department of Preventive Medicine, Los Angeles, USA

Received 28 April 2004; received in revised form 23 June 2004; accepted 12 July 2004

Abstract

Ambient fine particle ($PM_{2.5}$) concentrations are associated with premature mortality and other health effects. Urban populations spend a majority of their time in indoor environments, and thus exposures are modified by building envelopes. Ambient particles have been found to penetrate indoors very efficiently (penetration efficiency $P \approx 1.0$), where they are slowly removed by deposition, adsorption, and other mechanisms. Other particles are generated indoors, even in buildings with no obvious sources like combustion devices, cooking, use of aerosol products, etc.. The health effects of indoor generated particles are currently not well understood, and require information on concentrations and exposure levels.

The current work apportions residential $PM_{2.5}$ concentrations measured in the *EXPOLIS* study to ambient and non-ambient fractions. The results show that the mean infiltration efficiency of $PM_{2.5}$ particles is similar in all four cities included in the analysis, ranging from 0.59 in Helsinki to 0.70 in Athens, with Basle and Prague in between. Mean residential indoor concentrations of ambient particles range from 7 (Helsinki) to 21 $\mu g\,m^{-3}$ (Athens). Based on $PM_{2.5}$ decay rates estimated in the US, estimates of air exchange rates and indoor source strengths were calculated. The mean air exchange rate was highest in Athens and lowest in Prague. Indoor source strengths were similar in Athens, Basle and Prague, but lower in Helsinki. Some suggestions of possible determinants of indoor generated non-ETS $PM_{2.5}$ were acquired using regression analysis. Building materials and other building and family characteristics were associated with the indoor generated particle levels. A significant fraction of the indoor concentrations remained unexplained.

Keywords: Exposure; Particle; *EXPOLIS*, Sulphur; Indoor air quality; Indoor sources

1. Introduction

Ambient fine particulate matter ($PM_{2.5}$) concentrations have been associated with excess mortality and morbidity at current urban levels (e.g. Pope et al., 2002).

*Corresponding author. KTL, P.O.B. 95, FIN-70701 KUOPIO, FINLAND. Tel.: +358-17-201 211 (department), +358-17-201 171 (personal); fax: +358-17-201 184 (laboratory)
E-mail address: otto.hanninen@ktl.fi (O.O. Hänninen).

1352-2310/$ - see front matter
doi:10.1016/j.atmosenv.2004.07.015

Nomenclature

ETS: environmental tobacco smoke
EXPOLIS: air pollution exposure of adult urban populations in Europe –study
KTL: Finnish National Public Health Institute
PM: particulate matter
$PM_{2.5}$: fine particles, i.e. PM with aerodynamic diameter smaller than 2.5 μm
PM_{10}: PM with aerodynamic diameter smaller than 10 μm
RSP: respirable suspended particles ($\approx PM_{3.5}$)
SD: standard deviation

Because urban populations typically spend large fractions of time in indoor environments, it is important to understand how exposures to outdoor particles are affected by building envelopes (Wallace, 1996). According to population-based studies conducted during the last couple of decades in the US, summarized shortly below, it seems that exposures are reduced (i) substantially by decay processes of particles in indoor environments, being effective especially in (ii) reduced air exchange rates, and (iii) to a lesser extent or not at all by limiting the penetration of ambient particles indoors. (Wallace, 1996).

On the other hand, the total personal exposure to particulate matter may be significantly increased by particles generated by indoor sources like smoking, cooking and other activities (Wallace, 1996; Özkaynak et al., 1996). The relative contribution of these particles to exposures increases when air exchange rates are lower. It is still not clear whether non-ambient particles have health effects similar to those that have been shown for ambient particles. For efficient control of health effects, however, exposures to particles from different sources, having potentially different health effects or exposure-response relationships, should be estimated separately (Wilson et al., 2000;Wallace, 1996).

The following background section gives a short description of the techniques used earlier to estimate the infiltration of ambient particles indoors and the concentrations of indoor generated particles. The objective of the current work is to: (i) modify the analysis technique suggested by Ott et al. (2000) and Wilson et al. (2000) to use $PM_{2.5}$ bound sulphur as a marker substance, (ii) apportion the observed residential indoor $PM_{2.5}$ concentrations to fractions of ambient and indoor origin, and (iii) investigate potential determinants of indoor generated particle concentrations, including building and family characteristics.

2. Background

Five large population based studies have been conducted on the relationships of indoor and outdoor particles: the Harvard six cities—study, the New York State ERDA—study and the EPA particle TEAM (PTEAM)—study in US, the Ethyl Corporation study in Toronto, Canada, and the European *EXPOLIS*-study. The first three of these and dozens of smaller studies are reviewed in detail by Wallace (1996). Dockery and Spengler (1981) used indoor and outdoor measurements of respirable particles (RSP, $\approx PM_{3.5}$) from six US cities to study the indoor–outdoor relationships of particles in 68 residences over a one-year measurement period. The same cities were studied later in the famous Harvard six cities epidemiological study (Dockery et al., 1993). Dockery and Spengler elaborated on the mass-balance equation, assuming uniform mixing within the building and steady state conditions, i.e. that the penetration efficiency P, the air exchange rate a and the decay rate k stay constant over the sampling period Δt. Accordingly, the average indoor concentra tion $\overline{C_i}$ ($\mu g\,m^{-3}$) can be expressed as:

$$\overline{C_i} = \frac{Pa}{a+k}\overline{C_a} + \frac{\overline{Q}}{V(a+k)} - \frac{\Delta C_i}{\Delta t(a+k)}, \quad (1)$$

where C_i is the indoor concentration ($\mu g\,m^{-3}$), C_a the ambient (outdoor) concentration ($\mu g\,m^{-3}$), P the penetration efficiency (dimensionless), a the air exchange rate (h^{-1}), k the decay rate indoors (h^{-1}), Q the source strength ($\mu g\,h^{-1}$) (symbol used by Dockery and Spengler was S) and V the interior volume of the building (m^3).

The third term on the right side of the equation is reorganized here to use Δ- variables: $\Delta C = C_i(t_1) - C_i(t_0)$, indoor concentration change during the sampling period ($\mu g\,m^{-3}$), $\Delta t = t_1 = t_0$, sampling period (h).

The third term in the equation, representing the lag of indoor concentrations in reaching equilibrium, becomes relatively small and can be ignored for 24-h or longer sampling periods. Dockery and Spengler estimated the remaining two model parameters using a regression technique, giving slope values of 0.70 for RSP and 0.75 for sulphates and a constant term of 15.0 $\mu g\,m^{-3}$ for RSP. Based on theoretical considerations and earlier results from Dockery's Ph.D. theses, they state that for fine particles the decay term k is negligible (<0.5 h^{-1} for

RSP, $<0.05\,h^{-1}$ for PM_1) in comparison to typical air exchange rates ($1.5\,h^{-1}$) and that thus the indoor–outdoor regression slope can be interpreted as the particle mass penetration efficiency P.

Koutrakis et al. (1992) analyzed a similar population based data set from Suffolk and Onondaga counties, New York, for source apportionment of indoor $PM_{2.5}$ and 16 elements. The final data set consisted of indoor measurements of 178 residences and 57 corresponding outdoor measurements. The outdoor measurements were used for several residences measured at the same time in different locations. Koutrakis et al. simplified the mass-balance equation by assuming that air exchange rate ($0.51\pm0.28\,h^{-1}$), residence volume ($341\pm184\,m^3$) and particle deposition velocity ($1.46\,m\,h^{-1}$, calculated from other values reported) presented little variability in the study and were thus considered as constants. Based on these assumptions Koutrakis et al. were able to solve the penetration efficiency and source flux terms using the regression estimate for the outdoor–indoor slope. The regression slope, however, was not statistically significant for $PM_{2.5}$ mass and 8 of the 16 elements (including sulphur). Thus, the authors used the average of the statistically significant slopes (0.49) instead, yielding a value of 0.84 for the $PM_{2.5}$ penetration efficiency. Using regression techniques, the authors also derived $PM_{2.5}$ source strengths for smoking (12.6 mg/cigarette) and other (unidentified) sources ($1.1\,mg\,h^{-1}$). Estimated source strengths for wood burning and kerosene heating were statistically non-significant.

Lewis analyzed VOC and $PM_{2.5}$ sources using measurement data from 10 residences in Boise, Idaho (1991). The observed air exchange rates varied between 0.2 and $0.8\,h^{-1}$ (mean $0.5\,h^{-1}$). None of the residences had obvious major indoor sources. Lewis separated particles from ambient and indoor sources using the methods presented by Dockery and Spengler (1981), based on simultaneous indoor and outdoor concentrations and one or more species with negligible indoor sources. Lewis identified sulphur, lead, zinc, and soil-corrected potassium as species with no notable indoor sources. Calculated infiltration factors for fine particle species averaged 0.5 and varied in a reasonable way with air change rates. The $PM_{2.5}$ mass concentrations attributable to indoor sources varied between -5 and $22\,\mu g\,m^{-3}$ (mean $3\,\mu g\,m^{-3}$).

In the PTEAM study personal exposures to PM_{10} and residential indoor and outdoor levels of $PM_{2.5}$ and PM_{10} were measured using a probability sample of the non-smoking population of Riverside, CA, over 10 years old. Each of the 178 subjects was monitored for two consecutive 12-h periods in September–November, 1990. Using these data, Clayton et al. (1993) identified house work (including vacuuming, dusting, carpet cleaning, cooking, using a clothes dryer), spraying (using paints, cleaners and other consumer products in spray-form), and tobacco smoke as indoor sources affecting personal exposures to PM_{10}. Özkaynak et al. (1996) performed detailed analyses of the mass-balance equation terms infiltration, penetration, decay and indoor source strength for PM_{10} and $PM_{2.5}$ particles. They modified the mass-balance model developed by Koutrakis et al. (1992) and derived population averages for the equation terms using nonlinear least squares methods. The original solution estimated the penetration efficiency to be slightly above unity, which is impossible by definition, and thus the model was constrained by setting $P=1$ (Wallace, 1996; Wilson et al., 2000). Unknown indoor sources were found to account for a substantial fraction (25%) of indoor concentrations.

Several smaller-scale studies have used continuous monitoring techniques to investigate the indoor behaviour of particles. Thatcher and Layton (1995) studied the penetration, deposition and resuspension properties of particles as a function of particle size in a single residence during the summer months of 1993 in Livermore, California. Air exchange rates varied between 0.14 and $0.3\,h^{-1}$. The calculated penetration efficiencies were close to unity (0.9–1.4, except one higher outlying value). The particle deposition loss rates ranged from $0.0\,h^{-1}$ for particle sizes between 0.3–0.5 µm to $4.1\,h^{-1}$ for particles larger than 25 µm. Deposition loss rates calculated for 2–3 µm particles were $0.55–0.75\,h^{-1}$. The concentration of resuspended particles below 1 µm was minimal, but concentrations increased by over $20\,\mu g\,m^{-3}$ for the 1–5 µm particle size range due to vigorous house cleaning activities.

Abt et al. (2000) and Long et al. (2001) studied the behaviour of particles of different sizes in 4 and 9 residences, respectively, in the Boston area, Massachusetts. They used scanning mobility particle sizer (SMPS) and aerodynamic particle sizer (APS) monitors (TSI, Inc.) together with a tapered element oscillating microbalance (TEOM) for particle monitoring, and the mass-balance equation for deriving penetration, decay and source terms for different particle sizes. Due to the numerous days of monitoring they were able to report variability of these parameters in the buildings over a period of approximately 1-week. Air exchange rates reported by Abt et al. (2000) based on analysis of decay rates varied between 0.16 and $0.66\,h^{-1}$. The decay rates varied from just below 1 to slightly over $3\,h^{-1}$ for particle sizes from 0.3 to 7 µm. The decay rate was the lowest for particle sizes around 0.5 µm. Long et al. (2001) found penetration efficiencies to be between 0.7 and 1.0 for particle sizes up to 2.5 µm and 0.3 for particle sizes 2.5–10 µm. The decay rates varied between 0.1 and $0.5\,h^{-1}$, being lowest for particle sizes between 0.1 and 0.5 µm and 0.72 ± 0.34 (SE) for 2–3 µm particles. Abt et al. saw the effects of indoor sources (cooking, cleaning, indoor work and washing) mostly on particles larger than 2 μm. The effect of cooking on particles <2 μm was small and not statistically significant. Long et al. (2000), on the other hand, reported significant peaks due to various cooking events especially in the particle size range 0.1–0.5 μm. The peaks were typically short.

Ott et al. (2000) developed a statistical method called random component superposition (RCSP) model for separating personal exposures to particulate matter into ambient and non-ambient components and applied it on PM_{10} data from Phillipsburg, NJ (THEES study), Riverside, CA (PTEAM study) and Toronto, Canada (Ethyl Corporation study). Ott et al. used the regression slope of exposure against ambient concentration from population level data to estimate the fraction of personal exposures attributable to particles from ambient sources. The observed personal exposure level, minus this fraction, was then interpreted as non-ambient exposure level, consisting of exposures to particles from indoor sources as well as a personal cloud. Ott et al. also demonstrated how the suggested analysis can be applied to modelling of residential indoor PM_{10} concentrations in Phillipsburg instead of exposure data (Fig. 2. in Ott et al., 2000), and derived a slope of 0.53 and intercept of 18 μg m^{-3}. The slopes of exposure versus ambient PM_{10} concentration in the three cities were 0.54, 0.55 and 0.61, and corresponding mean non-ambient PM_{10} exposure levels 53, 59 and 52 μg m^{-3}, respectively.

Wilson et al. (2000) elaborate the same equations and estimation methods as Ott et al. (2000) using symbols that are familiar from the mass-balance equations presented by Dockery and Spengler (1981) and Koutrakis et al. (1992). The following summary of the technique for separating ambient and indoor generated PM levels is essentially the same model as developed by Ott et al. (2000), but presented using a notation adapted from Wilson et al. (2000). Apportionment is based on Eq. (2):

$$C_{ig} = C_i - C_{ai}, \tag{2}$$

where C_i is the total indoor concentration, C_{ai} the concentration of ambient PM that has infiltrated indoors and C_{ig} is the concentration of indoor generated particles. Based on the mass-balance equation presented by Dockery and Spengler (1981) these concentration fractions can be expressed as

$$C_{ai} = \frac{Pa}{a+k} C_a, \tag{3a}$$

$$C_{ig} = \frac{Q}{V(a+k)}, \tag{3b}$$

where the other symbols are same as in Eq. (1). The indoor–outdoor ratio of the ambient PM, i.e. the infiltration fac tor, can be calculated as

$$F_{INF} = \frac{C_{ai}}{C_a} = \frac{Pa}{a+k}. \tag{4}$$

Combining Eqs. (2) and (4) the indoor concentration can be expressed as

$$C_i = F_{INF} C_a + C_{ig} \tag{5}$$

and the parameters F_{INF} and C_{ig} can solved from the regression of indoor concentration against the ambient concentration as demonstrated by Ott et al. (2000). The slope of the regression (β_1) estimates the F_{INF} and the intercept (β_0) the average concentration of indoor-generated PM (C_{ig}).

3. Material and methods

The multi-centre *EXPOLIS*—study funded by the European Union was conducted in seven European cities during 1996–2000. In each city a population sample was drawn and personal exposures and micro-environment concentrations of $PM_{2.5}$ (including both gravimetric mass and elemental composition), carbon monoxide, nitrogen dioxide and 30 volatile organic compounds were monitored for 48 h over a one-year period. The current work uses the $PM_{2.5}$ measurements conducted at each subject's residence in Athens, Basle, Helsinki, and Prague. Home indoor and outdoor air was sampled for two consecutive working days from evening to morning, when the study subject was expected to be at home. Thus, the sampling time for each residential $PM_{2.5}$ sample was typically 30–32 h. Indoor and outdoor pumps were programmed to run simultaneously, but in a few cases different sampling times indicated exceptions to the normal procedure. Besides being an indication that the sampling was not completed as planned, different sampling times may lead to errors in the following calculations. To limit these errors, residences where the indoor and outdoor sampling times differed by more than 3 h were excluded from the current analysis. The data is summarized in Table 1. Population sampling, response rates and sample quality are described in detail in Rotko et al. (2000) and the overall study design in Jantunen et al. (1998). The $PM_{2.5}$ measurement techniques and general quality assurance results are described in Koistinen et al. (1999) and Hänninen et al. (2002a). The elemental composition of the filters was analysed using energy dispersive X-ray fluorescence (ED-XRF) by the University of Basle (Mathys et al., 2001). Data were accessed from the combined international *EXPOLIS* access database (CIDB, version September, 2002), described in Hänninen et al. (2002b).

Sulphur is suitable marker for ambient PM, as there are typically no indoor sources. Sulphur is emitted as

Table 1
Summary of the $PM_{2.5}$ and sulphur concentration and residence volume data

	$PM_{2.5}$ ($\mu g\,m^{-3}$)								
	Residence outdoors			Indoors, all			Indoors, non-ETS		
	Mean±SD	(min, max)	n	Mean±SD	(min, max)	n	Mean±SD	(min, max)	n
Athens	37±27	[9, 140]	47	31±17	[12, 75]	35	23±11	[12, 52]	21
Basle	19±12	[5, 59]	47	26±26	[6, 140]	40	17±8	[6, 39]	29
Helsinki	10±7	[2, 45]	170	13±16	[2, 122]	170	9±6	[2, 27]	135
Prague	27±10	[10, 48]	20	36±30	[10, 124]	47	25±16	[10, 96]	32

	$PM_{2.5}$ –bound sulphur ($ng\,m^{-3}$)						Residence volumes (m^3)		
	Residence outdoors			Indoors, all					
	Mean±SD	(min, max)	n	Mean±SD	(min, max)	n	Mean±SD	(min, max)	n
Athens	7564±5095	[2047, 29175]	37	5291±2020	[2739, 9016]	28	290±83	[140, 508]	50
Basle	3291±1594	[951, 7822]	41	2619±1625	[728, 7215]	30	280±169	[95, 910]	50
Helsinki	2151±1502	[175, 6434]	98	1586±1287	[215, 5930]	84	205±83	[63, 524]	189
Prague	3971±1536	[1275, 7802]	20	3073±1278	[1268, 5492]	16	233±80	[79, 360]	49

gaseous sulphur dioxide and oxidized to sulphate in the atmosphere. Sulphate condenses on existing particles and forms new ones. As a result, a large fraction of the ambient particles contain traces of sulphur, but the largest fraction of particle bound sulphur is in the submicron particle size range. Because PM mass on the other hand is concentrated in the larger particles, which have on the average higher settling velocities and lower diffusion coefficients, it can be expected that decay rates of $PM_{2.5}$ particles differ from that of PM-bound sulphur, and thus corresponding infiltration rates also differ. In the current work sulphur I/O ratios are adjusted for $PM_{2.5}$ infiltration by using the ratio of $PM_{2.5}$ and sulphur indoor–outdoor regression slopes in each city. The indoor levels of ambient $PM_{2.5}$ and concentrations of indoor-generated $PM_{2.5}$ in residences are then estimated from the EXPOLIS data from four European cities.

The overall population values of infiltration factors for $PM_{2.5}$ and $PM_{2.5}$ bound sulphur were calculated as slopes (β_1) of the corresponding indoor–outdoor concentration regression. The assumption of no indoor sources was supported by the small and statistically non-significant intercepts (β_0) of the indoor–outdoor regressions, varying on both sides of zero (Fig. 1). The $PM_{2.5}$ regression was calculated using only non-ETS exposed residences to avoid influence of unevenly distributed high ETS concentrations (the slopes were slightly affected by the ETS residences in Basle and Prague while in Athens and Helsinki the effect was small as can be seen from the right-side charts in Fig. 1). Because the slopes for $PM_{2.5}$ were lower than for sulphur for all cities (Table 2), we concluded that, as suggested earlier by e.g. Dockery and Spengler (1981), the decay rate of $PM_{2.5}$ particles is slightly higher than that of PM bound sulphur due to the differences in particle sizes and other properties, and thus the infiltration efficiency is also different, too.

To estimate the $PM_{2.5}$ infiltration factors for individual residences we first calculated the sulphur infiltration factor using Eq. (6) for each residence

$$C_i^S = F_{INF}^S C_a^S + C_{ig}^S,\ C_{ig}^S = 0,$$
$$\Rightarrow F_{INF}^S = \frac{C_i^S}{C_a^S}, \tag{6}$$

where F_{INF}^S = infiltration factor for sulphur for a single residence, C_i^S, C_a^S, C_{ig}^S are the indoor, outdoor and indoor generated concentrations of sulphur, respectively.

Then, the observed difference in $PM_{2.5}$ and sulphur infiltration factors was corrected using the ratio of the corresponding regression coefficients according to Eq. (7):

$$F_{INF}^{PM_{2.5}} = \frac{\beta_1^{PM_{2.5}}}{\beta_1^S} F_{INF}^S, \tag{7}$$

where F_{INF}^S is the infiltration factor for sulphur for a single residence, $\beta_1^{PM_{2.5}}, \beta_1^S$ are the $PM_{2.5}$ and sulphur indoor–outdoor regression slopes for each city (Table 2).

Indoor concentrations of ambient $PM_{2.5}$ were then calculated using infiltration factors and residential outdoor concentrations according to Eq. (3a). Concentrations of indoor generated $PM_{2.5}$ were calculated as

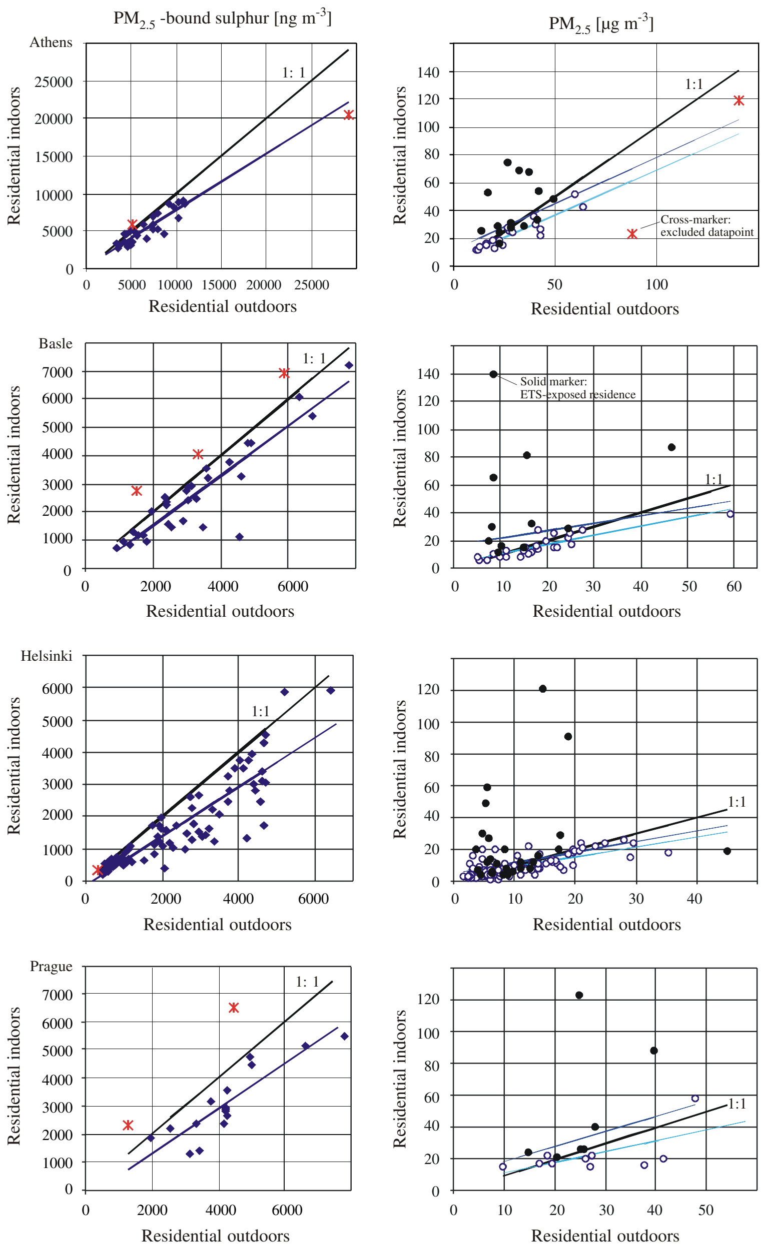

Fig. 1. Sulphur (left column) and $PM_{2.5}$ indoor–outdoor relationships. Besides the 1:1 line, regression lines are shown; for $PM_{2.5}$ two regression lines, upper one including the ETS exposed residences (solid markers). Data points discarded in jack knife sensitivity analysis of the result variables are shown with x-markers.

the difference of the total measured indoor concentrations and estimated contributions of ambient $PM_{2.5}$ according to Eq. (2). To estimate air exchange rates and indoor generated particle concentrations we solved these from the Eqs. (3b) and (4):

$$F_{\mathrm{INF}} = \frac{Pa}{a+k} \Rightarrow a = \frac{kF_{\mathrm{INF}}}{P - F_{\mathrm{INF}}}, \tag{8a}$$

Table 2
Regression analysis of the sulphur and $PM_{2.5}$ indoor–outdoor relationships

	Sulphur					$PM_{2.5}$					β_1 ratio
	β_1	SE	r^2	p	n	β_1	SE	r^2	p	n	$PM_{2.5}$/S
Athens	0.75	0.072	0.81	<0.000	28	0.64	0.070	0.83	<0.000	19	0.86
Basle	0.88	0.078	0.82	<0.000	30	0.69	0.090	0.71	<0.000	26	0.79
Helsinki	0.76	0.041	0.80	<0.000	84	0.64	0.057	0.53	<0.000	113	0.84
Prague	0.79	0.114	0.78	<0.000	16	0.67	0.293	0.40	0.050	10	0.85

$$C_{ig} = \frac{Q}{V(a+k)} \Rightarrow Q = C_{ig} V(a+k). \quad (8b)$$

Based on results from the US studies summarized by Wallace (1996), we then assumed that the penetration of $PM_{2.5}$ particles is approximately 1.0 and that the decay factor for $PM_{2.5}$ is $0.39\,h^{-1}$ (Wallace, 1996; Özkaynak et al., 1996) and calculated estimates for values of a and Q for individual homes.

Univariate single and stepwise-multiple regression analyses were run using the *EXPOLIS* questionnaire data, describing the residences, occupant characteristics and exposure related activities during the sampling period, against the calculated concentrations of indoor generated particles (C_{ig}). Only independent variables with a potential theoretical connection to indoor air quality, including generation or decay of particles in the indoor environment, were included in the analysis. Only households not exposed to tobacco smoke were considered. Regression models were calculated using Stata software, version 5.0 (Stata Corporation, College Station, TX).

4. Results

Indoor versus outdoor regression analyses produced slopes varying between 0.75 (Athens) and 0.88 (Basle) for sulphur (all regressions statistically highly significant, $p<0.001$) and between 0.64 (Athens and Helsinki) and 0.69 (Basle) for $PM_{2.5}$ (others highly significant, Prague $p=0.050$). The slopes of the regression with data points are shown in Fig. 1 and corresponding statistics are shown in Table 2. The magnitude of the sulphur intercepts were 5–10% of the average outdoor sulphur concentration (all statistically non-significant, $p>0.2$), confirming that there were no significant population level indoor sources for sulphur. From the graphs in Fig. 1 it can be seen, however, that there were seven residences with I/O ratios above 1; these cases might be influenced by sulphur indoor sources. The single case in Helsinki occurs on a very low sulphur level and might be caused by measurement error. These and one outlier case in Athens were excluded from the analysis, as otherwise they produced infiltration rates above one and resulted in highly negative indoor sources. Regression slopes for ETS-included and the ETS-excluded datasets differed in Athens and in Prague due to unevenly distributed high indoor concentrations in ETS-exposed residences (indicated with solid markers in Fig. 1). Non-ETS regression coefficients were used in the following analysis.

Mean infiltration factors for $PM_{2.5}$ varied between 0.59 (Helsinki) and 0.70 (Athens, Table 3). The values for Basle, Helsinki, and Prague were very close to each other, while in Athens the infiltration rate was somewhat larger. This difference can be seen in all percentiles of the infiltration levels as depicted in the first chart in Fig. 2. This is reasonable for the warmer climate in Athens, which favours less tightly sealed buildings and thus higher air exchange rates. The estimated mean air exchange rate for Athens was $1.3\,h^{-1}$, while in the three other cities it ranged between 0.7 and $0.9\,h^{-1}$. The air exchange rate percentiles in Athens were much higher than in the other cities (Fig. 2). Variability of the air exchange rates was clearly smaller in Basle and Prague compared to Athens and Helsinki (Table 3). At least in Prague this was probably caused by the fact that all residences were sampled from the downtown area.

Indoor concentrations of ambient PM (C_{ai}) and of indoor generated non-ETS $PM_{2.5}$ (C_{ig}) are summarized in Table 4 and in Fig. 2. The indoor concentrations of ambient PM are clearly highest in Athens ($21 \pm 13\,\mu g\,m^{-3}$) and lowest in Helsinki ($7 \pm 5\,\mu g\,m^{-3}$). Concentrations caused by indoor generated $PM_{2.5}$ were highest in Basle and Prague, where air exchange rates were the lowest. Source strengths for indoor generated non-ETS $PM_{2.5}$ were comparable in the other three cities (1.4–$1.5\,mg\,h^{-1}$), but were clearly lower in Helsinki ($0.6\,mg\,h^{-1}$).

The lowest percentiles of indoor generated concentrations (C_{ig}) and source strengths (Q) are negative as can be seen in the bottom charts in Fig. 2. The sample size of non-ETS exposed residences in Prague was small ($n=9$) and as a result the standard error of the slope of $PM_{2.5}$ indoor–outdoor regression was high. Negative source strengths, however, can be expected in this kind of analysis and also were reported by Lewis et al. (1991).

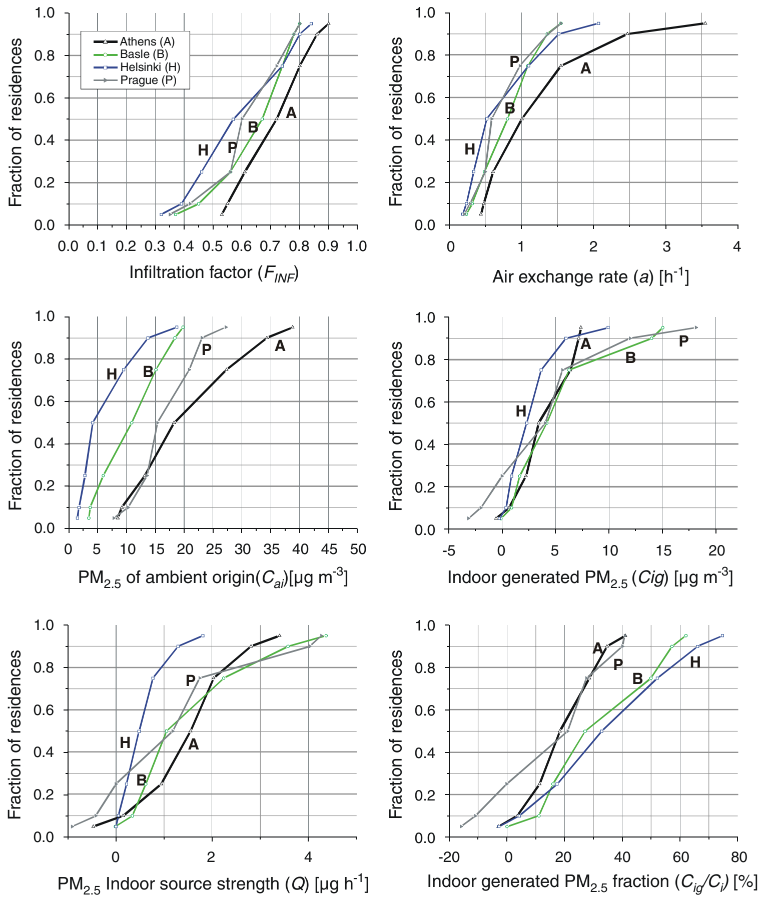

Fig. 2. Distributions of the estimated variables: $PM_{2.5}$ infiltration factors, concentrations of ambient $PM_{2.5}$ and non-ETS indoor generated $PM_{2.5}$, air exchange rates (air changes per hour) and non-ETS indoor $PM_{2.5}$ generation rate.

Infiltration factors are lower than one, because of decay of particles indoors. When there are no indoor sources, and for some reason the decay is faster than average in a certain building (for reasons like static electricity in materials, use of air cleaners etc., or differences in particle size distribution), in this kind of analysis the decay processes appear as negative indoor sources.

As pointed out by Özkaynak et al. (1996) and Wallace (1996), a large fraction of indoor $PM_{2.5}$ concentrations observed in previous population based studies was unexplained. Because the current analysis estimated infiltration ratios for individual residences, and thus allowed for estimation of indoor generated particles in individual residences, a series of regression analyses was run in an attempt to identify determinants of indoor generated particle levels. Regression analyses were run for the combined dataset from the four cities using indoor source concentrations (C_{ig}) of non-ETS exposed residences as the dependent variable ($n=125$, mean $\pm$ SD $3.6\pm4.0\,\mu g\,m^{-3}$). The dependent variable included one outlying high value ($24.3\,\mu g\,m^{-3}$). This value, when included in the models, nearly doubled the coefficient of determination, and increased the statistical significance of variables associated with this special case (e.g. education level and marital status). A gas appliance was used for 36 h in this residence and was probably connected to the high level, but to draw conclusions on a single data point is unjustified and thus this single data point was excluded from reported models.

Regression analyses were able to identify few variables that were associated with indoor source concentrations with statistical significance (Table 5). In simple regression these variables were (i) wooden building material (three variables for floor, panels, and both combined), (ii) city, (iii) building age (in decades), (iv) floor of the residence (0 = ground floor) and (v) use of

Table 3
Infiltration factors for $PM_{2.5}$ and sulphur particles and estimated air exchange rates

	F_{INF} (Sulphur)		F_{INF} ($PM_{2.5}$)			Ventilation rate	
	Mean [1]	SD [1]	Mean [1]	SD [1]	n	Mean (h^{-1})	SD (h^{-1})
Athens	0.82	0.14	0.70	0.12	28	1.3	1.1
Basle	0.80	0.19	0.63	0.15	30	0.83	0.46
Helsinki	0.70	0.20	0.59	0.17	84	0.81	0.85
Prague	0.72	0.16	0.61	0.14	16	0.75	0.43

Table 4
$PM_{2.5}$ concentrations of ambient and non-ETS indoor origin, and corresponding $PM_{2.5}$ indoor source strengths

	$PM_{2.5}$ of ambient origin			Non-ETS indoor sources				
	Mean ($\mu g\,m^{-3}$)	SD ($\mu g\,m^{-3}$)	n	Mean ($\mu g\,m^{-3}$)	SD ($\mu g\,m^{-3}$)	Mean ($mg\,h^{-1}$)	SD ($mg\,h^{-1}$)	n
Athens	21.3	12.7	28	4.0	*2.8*	1.4	1.3	16
Basle	11.4	7.0	30	5.3	4.9	1.5	1.4	23
Helsinki	6.5	5.3	84	2.9	3.1	0.6	0.6	71
Prague	16.9	6.7	16	5.0	8.2	1.4	1.9	9

stove other than electric. The slope for the stove variable was negative and the data was available only from Helsinki, where there were four residences with non-electric stoves. Variables and regression coefficients are listed in Table 5).

Five variables were entered to a forward-selection stepwise regression model with significance limit for entering variables set to $p = 0.2$. Only the first variable remained significant at 0.95 level in the final model: (i) PVC floors, (ii) air exchange rate, (iii) building age (decades), (iv) attached garage, and (v) wood panels. In the corresponding backward model eleven variables stayed at the same p-value limit set for removal of variables. Four of these were the same variables as in the forward model, but their ranking (in the order of decreasing statistical significance in the final model) was changed. The wood panel variable was dropped and variables for number of family members, floor area per person, chipboard walls, residence size (m^2), pets, traffic density near home and gender of the subject were included in the model. Six variables remained statistically significant in the final model (variables that were entered in the forward selection model too are indicated with an asterisk): (i) air exchange rate*, (ii) PVC floors*, (iii) number of family members, (iv) floor area per person, (v) building age in decades*, and (vi) attached garage*.

Other tested variables included time spent cooking, vacuuming, soft furnishing materials (including carpets and curtains), use of burning devices (both as hours of use and as binary variable, separately for devices fuelled with wood, oil, coal or gas, and all of them combined), hours that windows were kept open, marital status, years of education, and time of year (both as month and as season).

The coefficient of determination remained below 10% for all variables in simple regressions. The corresponding multiple regression coefficients of determination were 0.22 for the forward and 0.34 for the backward selection models. Variables consistently associated with indoor generated particle levels included air exchange rates and descriptors of building materials (wooden floors and panels, PVC floors and chipboard walls). Increased air exchange rate decreased indoor generated levels as expected. The role of indoor materials is not as clear; they might be descriptors of building, family and activity characteristics not included in exposure questionnaires, but they might also be causally connected to e.g. resuspension and decay rates of particles. It seems that the origin of the non-ETS indoor generated particles is heterogeneous and that it is not possible to completely quantify the various sources from population based non-continuous $PM_{2.5}$ measurements. The achieved coefficient of determination in the backward selection model gives, however, indication on variables that are worth studying in more detail.

5. Discussion

Mean indoor levels of ambient $PM_{2.5}$ are two to five times higher than corresponding levels caused by indoor

Table 5
Univariate single and stepwise multiple regressions for estimating the indoor generated $PM_{2.5}$ concentration using residence related questionnaire data as independent variables. The stepwise regression significance limit was set to $p=0.2$

		Single regression				Stepwise regression 1 Enter: $r^2 = 0.22$, $p = 0.006$			Stepwise regression 2 Remove: $r^2 = 0.34$, $p = 0.006$		
Variable	Description	β_1	r^2	p	n	Rank	β_1	p	Rank	β_1	p
Wood panels	No = 0, yes = 1	−2.2	8%	**0.00**	94/30	5	−1.44	0.15			
Wood material	No wood panels/floor = 0; either = 1;both = 2	−1.31	8%	**0.00**	39/57/28						
City	4 dummies: A,B,H,P; 0 = no, 1 = yes	n/a	7%	**0.03**	16/22/78/8						
Building age	Decades, 1–3, 4 or more	0.61	4%	**0.03**	16/25/24/59	3	0.657	0.12	5	0.96	**0.02**
Wood floor	No = 0, yes = 1	−1.32	3%	**0.04**	41/83						
Building floor	Integer, 0 (ground)-9	0.32	3%	**0.05**	1-45/floor						
Non-electric stove	No = 0, yes = 1	−3.38	3%	**0.05**	120/4						
PVC floor	No = 0, yes = 1	1.05	2%	0.12	86/38	1	2.09	**0.02**	2	2.13	**0.02**
Wall-to-wall carpet	No = 0, yes = 1	1.12	2%	0.14	98/26						
Attached garage	No = 0, yes = 1	1.77	2%	0.16	116/8	4	3.26	0.13	6	4.33	**0.05**
Material status	No = 0, yes = 1	−0.85	1%	0.19	43/81						
Chipboard walls	No = 0, yes = 1	−0.83	1%	0.25	94/30				7	−1.46	0.11
Highrise building	No = 0, yes = 1	0.73	1%	0.26	44/80						
Air exchange rate	h^{-1}	−0.39	< 1%	0.27	(cont.)[a]	2	−0.795	0.12	1	−1.24	**0.02**
Family size	Persons (1–6)	−0.26	< 1%	0.29	(too many)[b]				3	−1.79	**0.02**
Plaster walls	No = 0, yes = 1	−0.63	< 1%	0.32	69/55						
Wall paper	No = 0, yes = 1	−0.60	< 1%	0.34	54/70						
Windows kept open	h	−0.015	< 1%	0.35	(cont.)[a]						
Cooking time	min	0.0017	< 1%	0.38	(cont.)[a]						
Use of burning devices	No = 0, yes = 1	−1.48	< 1%	0.40	120/4						
Use of gas appliances	h	−0.22	< 1%	0.43	(cont.)[a]						
Soft furnishing	No = 0, yes = 1	−0.60	< 1%	0.47	20/104						
Age of the subject	Years (range 29–59)	−0.025	< 1%	0.48	(cont.)[a]						
Annoyance experi nced	1(none)-8(max.)	0.099	< 1%	0.54	(too many)[b]						
Other than w2w-carpet	No = 0, yes = 1	−0.47	< 1%	0.55	24/100						
Vacuming	No = 0, yes = 1	0.34	< 1%	0.59	77/47						
Residence floor area	m^2	−0.004	< 1%	0.60	(cont.)[a]				8	0.036	0.12
Years of education	Integer, range 6–33	−0.045	< 1%	0.62	(cont.)[a]						
Gender	Female = 0; male = 1	−0.307	< 1%	0.63	72/52				11	1.14	0.19
Linoleum floor	No = 0, yes = 1	0.38	< 1%	0.68	108/16						
Traffice density	Estimated, integer 0–5	0.077	< 1%	0.68	(too many)[b]				10	−0.34	0.18
Pets	No = 0, yes = 1	−0.36	< 1%	0.70	86/38				9	−1.33	0.16
Floor area/person	m^2	−0.006	< 1%	0.71	(cont.)[a]				4	−0.115	**0.02**
Month	Integer 1–12	−0.025	< 1%	0.79	(too many)[b]						
Winter	Dec–Feb = 1;other = 0	−0.17	< 1%	0.80	90/34						
Summer	June–Aug = 1;other = 0	−0.17	< 1%	0.80	88/36						
Time using burning dev.	h	−0.017	< 1%	0.81	(cont.)[a]						
Time burning wood	h	0.066	< 1%	0.88	(cont.)[a]						
Residence volume	m^3	−0.0001	< 1%	0.96	(cont.)[a]						
Downtown location	No = 0, yes = 1	0.059	< 1%	0.96	96/28						
Curtains	No = 0, yes = 1	−0.04	< 1%	0.97	11/113						

[a]Continuous variable
[b]Too many classes to be listed. Variables with p-value $\leqslant 0.05$ in one or more models are shown in bold.

generated PM in Helsinki and Athens, respectively (Table 4). Residential indoor levels caused by ETS, excluded from the current analysis, were even higher. Thus, assuming equal toxicities for the sake of comparison, the relative importance of these PM exposures can be ranked into the order of: (i) ETS, (ii) ambient, and (iii) indoor generated non-ETS PM. Those PM fractions for which we have the strongest evidence of being a health problem, ETS and ambient $PM_{2.5}$, clearly dominate exposures—particularly in areas with high ambient PM. As the impacts of public health education continue to result in reduced exposures to tobacco smoke, and a combination of development of low pollution vehicles and successful emission control measures in industry reduce ambient $PM_{2.5}$ levels, more attention will be focused on indoor air pollution. There is a need, therefore, to understand the relative toxicities and health effects of PM from indoor sources, to determine whether policies to reduce exposure to these sources may be warranted.

Besides being air pollution sources by themselves, buildings can also be considered as a complementary means to reduce the exposure levels to ambient particles by adjusting the air exchange rates and using efficient filtration in mechanical ventilation systems. In particular, use of efficient filters in mechanical ventilation systems remains as an interesting option to further reduce exposures to ambient particles.

The current paper demonstrates estimation of the variability of infiltration factors in a random population sample. The standard error (SE) of the regression slope allows for assessment of the uncertainty in the mean infiltration level, but does not allow for assessment of parameter variability. The use of a marker element (sulphur in the current work) for calculation of infiltration factors for individual residences, allows for assessment of the variability of the infiltration efficiencies, but might be biased due to different decay rates of particles of different sizes. The current work combines the advantages of the both techniques by correcting the bias using a correction factor calculated as the ratio of the regression coefficients for $PM_{2.5}$ and the marker substance.

Current and previous attempts to identify and quantify non-ETS indoor sources of $PM_{2.5}$ have not been very successful. Short-term studies in single buildings using continuous monitoring techniques have been able to easily detect effects of cooking, cleaning and other indoor activities (Thatcher and Layton, 1995;Abt et al., 2000;e.g. Long et al., 2001). When these emission events are mixed with each other, diluted to a longer sampling time and mixed with measurement errors and variability of concentrations caused by outdoor sources and variations in the air exchange rates and infiltration ratios, it becomes extremely difficult to detect them with statistical significance in datasets of current sample sizes. The time dilution alone reduces the effect of a 30-min cooking event of $20\,\mu g\,m^{-3}$ to $0.4\,\mu g\,m^{-3}$ in the corresponding 24-h average. Questionnaire variables used to model these sources also include large variability. For example a cooking event might indicate boiling of water in one time activity diary entry, while in another it might refer to sautéing or frying, which are much more plausible sources of indoor particles. The dilution ratio for this kind of variability in the questionnaire data is difficult to estimate, but it does not seem too far-fetched to say that a large fraction of the population cooking time in Europe is not a significant source of indoor particles. Thus, even sources that are strikingly evident in the continuous short-term measurements end up being only vaguely visible, if at all, in the kind of population data used in the current work. On the other hand, people are present near the stove while cooking, and the effect may be larger on personal exposure than on indoor concentrations measured further away from the stove. To improve population exposure studies, exposure questionnaires must be improved to quantify the particle generating activities more accurately, including separation from non-source events.

Studies based on small, non-representative samples of buildings are very useful for deepening the understanding of the processes affecting I/O relationships and indoor concentrations of pollutants. Such studies are needed to supply relevant information that serves as input to models for estimation of exposure of general populations. Population based values of exposure variables, including description of variability, are needed for efficient application of models for exposure and risk analyses as pointed out by e.g. Ryan et al. (1986), Freijer et al. (2000) and Kruize et al. (2003). Further studies, based on development of exposure questionnaire variables through short-term indoor source studies using continuous monitoring combined with current findings on potential determinants of indoor sources, will allow for more detailed understanding of the indoor generated $PM_{2.5}$ levels.

6. Conclusions

The current paper demonstrates a method to estimate the variability of $PM_{2.5}$ infiltration ratios in a sample of buildings, and thus to more accurately estimate the variability of the levels of indoor generated particles. Indoor concentrations of ambient $PM_{2.5}$ varied substantially between cities; mean levels in Athens ($21\,\mu g\,m^{-3}$) were more than three times higher than in Helsinki ($6.5\,\mu g\,m^{-3}$). The building envelope is a significant modifier of personal exposures to ambient particles; for example in Athens residential indoor concentrations of ambient PM were only 56% of corresponding outdoor levels. Levels of indoor generated particles were similar in four different European cities (3–5 $\mu g\,m^{-3}$), being highest in Basle and Prague, where the air exchange rates were the lowest. Indoor source strengths were similar in Athens, Basle and Prague ($\approx 1.5\,mg^{-1}$) but only $0.6\,mg\,h^{-1}$ in Helsinki. These estimates are being used further in *EXPOLIS* exposure modelling.

Besides air exchange rate, building materials, building age, an attached garage, and family characteristics were associated with levels of indoor generated particles. Wooden floors and/or panels were associated with decreased levels of indoor generated particles, while PVC or plastic floors were associated with increased levels. The predictive power of regression models remained below 35% for reasons discussed. Refinement of exposure related questionnaires will be required in population exposure studies for more accurate prediction of levels of indoor generated particles, and thus improve our ability to estimate exposures.

Acknowledgements

The authors want to express their gratitude to Professor Wayne Ott, Department of Civil and Environmental Engineering and Department of Statistics, Stanford University, California, who provided significant insight to this work on statistical modelling of particulate exposure and indoor concentration distributions.

The *EXPOLIS* Centers are KTL (coordinating center) in Helsinki (Finland), University of Athens (Greece), University of Basle (Switzerland), University Joseph Fourier in Grenoble (France), University of Milan (Italy), Inst. Exp. Medicine AS CR and Health Inst. of Central Bohemia (Czech), VTT in Helsinki (Finland), RIVM in Bilthoven (The Netherlands) and Imperial College (United Kingdom). This work has been supported by EU contracts ENV4-CT96-0202 (*EXPOLIS*, DG12-DTEE), ERB IC20-CT96-0061 (Prague) and EVK4-CT-2002-00097 (FUMAPEX), Academy of Finland contracts N:o 36586, 40835 and 42610 (Helsinki), Swiss Ministry for Education and Science contract BBW N:o 95.0894 (Basel), French National Environment Agency (ADEME), Union Routière de France and Grenoble Communauté de Communes (Grenoble), DEFRA contract EPG1/5/106 (Oxford), other national research funds, and intramural funding from participating institutes.

References

Abt, E., Suh, H.H., Catalano, P., Koutrakis, P., 2000. Relative contribution of outdoor and indoor particle sources to indoor concentrations. Environmental Science & Technology 34 (17), 3579–3587.

Clayton, C.A., Perrit, R.L., Pellizzari, E.D., Thomas, K.W., Whitmore, R.W., Wallace, L.A., Özkaynak, H., Spengler, J.D., 1993. Particle total exposure assessment methodology (PTEAM) study: distributions of aerosol and elemental concentrations in personal, indoor and outdoor air samples in an southern California com. Journal of Exposure Analysis and Environmental Epidemiology 3, 227–250.

Dockery, D.W., Pope, I.C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, J.B.G., Speizer, F.E., 1993. An association between air pollution and mortality in six US cities. The New England Journal of Medicine 329, 1753–1759.

Dockery, D.W., Spengler, J.D., 1981. Indoor–outdoor relationships of respirable sulfates and particles. Atmospheric Environment 15, 335–343.

Freijer, J.I., Bloemen, H.J.T., 2000. Modeling Relationships between indoor and outdoor air quality. Journal of the Air & Waste Management Association 50, 292–300.

Hänninen, O., Koistinen, K.J., Kousa, A., Keski-Karhu, J., Jantunen, M.J., 2002a. Quantitative analysis of environmental factors in differential weighing of blank Teflon filters. Journal of the Air & Waste Management Association 52 (2), 134–139.

Hänninen, O.O., Alm, S., Kaarakainen, E., Jantunen, M., 2002b. The EXPOLIS Databases B13/2002. Publications of the National Public Health Institute, Kuopio KTL.

Jantunen, M.J., Hänninen, O., Katsouyanni, K., Knoppel, H., Kuenzli, N., Lebret, E., Maroni, M., Saarela, K., Sram, R., Zmirou, D., 1998. Air pollution exposure in European cities: the "EXPOLIS" study. Journal of Exposure Analysis and Environmental Epidemiology 8 (4), 495–518.

Koistinen, K.J., Kousa, A., Tenhola, V., Hänninen, O., Jantunen, M.J., Oglesby, L., Kuenzli, N., Georgoulis, L., 1999. Fine particle (PM2.5) measurement methodology, quality assurance procedures, and pilot results of the EXPOLIS study. Journal of the Air & Waste Management Association 49 (10), 1212–1220.

Koutrakis, P., Briggs, S.L.K., Leaderer, B.P., 1992. Source apportionment of indoor aerosols in Suffolk and Onondaga counties, New-York. Environmental Science & Technology 26 (3), 521–527.

Kruize, H., Hänninen, O., Breugelmans, O., Lebret, E., Jantunen, M., 2003. Description and demonstration of the EXPOLIS simulation model: two examples of modeling population exposure to particulate matter. Journal of Exposure Analysis and Environmental Epidemiology 13 (2), 87–99.

Lewis, C.W., 1991. Sources of air pollutants indoors: VOC and fine particulate species. Journal of Exposure Analysis and Environmental Epidemiology 1 (1), 31–44.

Long, C.M., Suh, H.H., Catalano, P.J., Koutrakis, P., 2001. Using time- and size-resolved particulate data to quantify indoor penetration and deposition behavior. EST 35, 2089–2099.

Long, C.M., Suh, H.H., Koutrakis, P., 2000. Characterization of indoor particle sources using continuous mass and size monitors. Journal of the Air & Waste Management Association 50, 1236–1250.

Mathys, P., Stern, W.B., Oglesby, L., Braun-Fahrländer, C., Ackermann-Liebrich, U., Jantunen, M.J., Künzli, N., 2001.

Elemental analysis of airborne particulate matter by ED-XRF within the European EXPOLIS study. ICP Information Newsletter 27 (3), 29–34.

Ott, W., Wallace, L., Mage, D., 2000. Predicting particulate (PM10) personal exposure distributions using a random component superposition statistical model. Journal of the Air & Waste Management Association 50, 1390–1406.

Özkaynak, H., Xue, J., Spengler, J., Wallace, L., Pellizzari, E., Jenkins, P., 1996. Personal exposure to airborne particles and metals: results from the particle TEAM study in Riverside, California. Journal of Exposure Analysis and Environmental Epidemiology 6, 57–78.

Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Journal of the American Medical Association 287 (9), 1132–1141.

Rotko, T., Oglesby, L., Kunzli, N., Jantunen, M.J., 2000. Population sampling in European air pollution exposure study, EXPOLIS: comparisons between the cities and representativeness of the samples. Journal of Exposure Analysis and Environmental Epidemiology 10 (4), 355–364.

Ryan, P.B., Spengler, J.D., Letz, R., 1986. Estimating personal exposures to NO2. Environment International 12, 395–400.

Thatcher, T.L., Layton, D.W., 1995. Deposition, resuspension, and penetration of particles within a residence. Atmospheric Environment 29, 1487–1497.

Wallace, L., 1996. Indoor particles: a review. Journal of the Air & Waste Management Association 46, 98–126.

Wilson, W.E., Mage, D.T., Grant, L.D., 2000. Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: why and how. Journal of the Air & Waste Management Association 50 (7), 1167–1183.

V

Hänninen, O.O., Lebret, E., Tuomisto, J.T., and Jantunen, M.J., 2005. **Characterization of Model Error in the Simulation of $PM_{2.5}$ Exposure Distributions of the Working Age Population in Helsinki, Finland**. *JAWMA*. 55: 446-457.

TECHNICAL PAPER

ISSN 1047-3289 *J. Air & Waste Manage. Assoc.* **55**:446–457

Characterization of Model Error in a Simulation of Fine Particulate Matter Exposure Distributions of the Working Age Population in Helsinki, Finland

Otto O. Hänninen, Jouni T. Tuomisto, and Matti J. Jantunen
Finnish National Public Health Institute (KTL), Centre for Environmental Health Risk Analysis, Kuopio, Finland

Erik Lebret
National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

ABSTRACT

Exposure models are needed for comparison of scenarios resulting from alternative policy options. The reliability of models used for such purposes should be quantified by comparing model outputs in a real situation with the corresponding observed exposures. Measurement errors affect the observations, but if the distribution of these errors for single observations is known, the bias caused for the population statistics can be corrected. The current paper does this and calculates model errors for a probabilistic simulation of 48-hr fine particulate matter ($PM_{2.5}$) exposures. Direct and nested microenvironment-based models are compared. The direct model requires knowledge on the distribution of the indoor concentrations, whereas the nested model calculates indoor concentrations from ambient levels, using infiltration factors and indoor sources. The model error in the mean exposure level was $<0.5\ \mu g\ m^{-3}$ for both models. Relative errors in the estimated population mean were $+1\%$ and -5% for the direct and nested models, respectively. Relative errors in the estimated SD were -9% and -23%, respectively. The magnitude of these errors and the errors calculated for population percentiles indicate that the model errors would not drive general conclusions derived from these models, supporting the use of the models as a tool for evaluation of potential exposure reductions in alternative policy scenarios.

IMPLICATIONS

Exposure models are needed for optimization of policies for controlling exposures to toxic substances. Costly and limiting controls must be based on scientifically sound estimates of the intended exposure reductions. In the current work, various error terms affecting exposure estimates are quantified and compared with each other for observed population exposure data, including quantitative validation of a microenvironment-based simulation model. The results show that the model errors are comparable with measurement error bias below the 90th percentile and above that with the sampling error. The presented general population model underestimates the highest exposures, indicating a need for special attention.

INTRODUCTION

Much interest has been paid to validating environmental models, particularly in the late 1980s and early 1990s, because of their use in public policy debates regarding climate change and other environmental impacts. These models contain large uncertainties and raise controversy because of the large costs of reducing the suggested impacts. The debate triggered critical reviews of the concept of model "validity". In their rather philosophical examination, Oreskes et al.[1] pointed out that validation or verification of any physical model is impossible in the sense that the truthfulness of a model can never be proven conclusively. A model is always developed in a given setting that implicitly defines its limits of applicability. Instead of conclusive validation, a model can be tested against real data. In case of test failure, the model is invalidated; if the results are favorable, the test can be interpreted as confirming, but not validating, the model.[2] Although Oreskes et al.[1] are absolutely correct, their conclusion is useful only if it is understood as a cry for more specific model-assessment descriptors than "validation" or "verification". The term validation is widely used in the scientific literature, and to make it meaningful, it should be understood as a demonstration that a model is capable of making accurate predictions in a given real setting.[3]

Model accuracy and reliability are continuous quantities, and model assessment is shifting toward a more quantitative analysis of model reliability.[4] A commonly used technique to assess model reliability is uncertainty analysis. Uncertainties in model outputs can be estimated by propagating known or estimated uncertainties through the model, and they are described by confidence limits that are assumed to contain the true value with a given probability.[5]

Our approach, however, is different. Our aim is to calculate the difference between the model outputs and the corresponding true values representing the model errors. These are not statistical measures of the probable differences between the model output and the true value but quantitative measures of the model accuracy in the given situation. The complementary nature of uncertainty analysis and the current approach is depicted in Figure 1. Uncertainty analysis can be conducted even when the true values are unknown, whereas the model errors can be calculated only in the opposite case. In the current work, analysis of model errors is applied to a probabilistic model used to estimate population variability of exposures to fine particulate matter ($PM_{2.5}$).

Our comparison of model outputs with observed values includes all errors created along the model and the measurement chains, as depicted in Figure 2. To be able to estimate the model errors, the experiment must be designed so that the other error terms can be excluded or quantified. In the current work, the errors caused by statistical sampling and representativity were excluded from the study design by use of input values and observed output values measured from the same population sample, and the effects of measurement errors are corrected for.

Probabilistic modeling has been useful in modeling of exposure distributions (i.e., variability).[6-11] In this context, probabilistic models are based on equations describing the causal physical relationships between the model variables, but instead of using inputs describing specific individuals, probabilistic model inputs are entered as probability distributions. Physical equations and values sampled from the input distributions are then used to calculate the outputs; thus, the simulated cases do not represent real individuals. It is therefore only meaningful to compare population statistics, such as mean, SD, and percentiles, with the corresponding observed values (Figure 2). In a deterministic model, in which specific

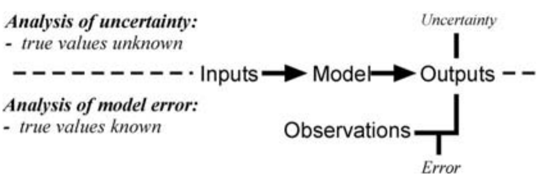

Figure 1. Uncertainty analysis propagates (known) uncertainties through the model to estimate confidence limits for the outputs. Analysis of model error quantifies the errors instead of the confidence limits.

individuals are modeled, the model error can in principle also be calculated for individuals. Probabilistic technique is also often used for the analysis of uncertainty, but in the current work, it is used only to model variability.

The probabilistic exposure model used was developed as part of the *EXPOLIS* study in collaboration by RIVM and KTL. The goals set for the model development included capability to assess population exposure distributions for (1) selected subpopulations and (2) urban areas in (3) different policy scenarios.[12] The developed microenvironment-based model can be run in either direct or nested mode. In the direct mode, lognormal microenvironment concentration distributions are assumed and the parameters entered as inputs. In the nested mode, the indoor concentration of ambient origin is modeled by use of ambient concentration distribution and infiltration factor distribution, as depicted in Figure 3. In both modes, additional indoor sources can be defined for a defined fraction of microenvironments.[13] Kruize et al.[13] applied the model to direct mode simulation of the annual distribution of 48-hr $PM_{2.5}$ exposures in Athens, Basle, Helsinki, and Prague and to nested mode simulation of the daily PM_{10} exposures of the general Dutch population, including all age groups and both rural and urban areas.

The model has also been applied to more detailed evaluation of the direct mode simulation of $PM_{2.5}$ exposures in Helsinki.[14] The distribution assumptions for the log normality of concentration distributions and the β distribution for the time fractions were tested; the concentrations followed lognormal distributions quite well, but the fits for the time fractions were more problematic. Statistically significant deviations in the time–activity distributions, however, did not lead to significant errors in the outputs. The current work continues by adding traffic as a fourth microenvironment to the model (including analysis of $PM_{2.5}$ concentrations experienced while in traffic by use of measurements in traffic), compares the direct and nested approaches (including the analysis of the infiltration factors and indoor sources in residences and workplaces), and takes the quantitative comparison of model outputs to observations one step further by accounting for the measurement errors.

The objectives of the current work are (1) quantification of the model error in simulation of population statistics of cross-sectional 48-hr $PM_{2.5}$ exposures (including correction of the measurement error bias in the observed exposure distribution required for this purpose); (2) comparison of the model errors with the other errors affecting population exposure assessment: measurement error, measurement error bias, and sampling error; and (3) comparison of the model errors in direct and nested simulation.

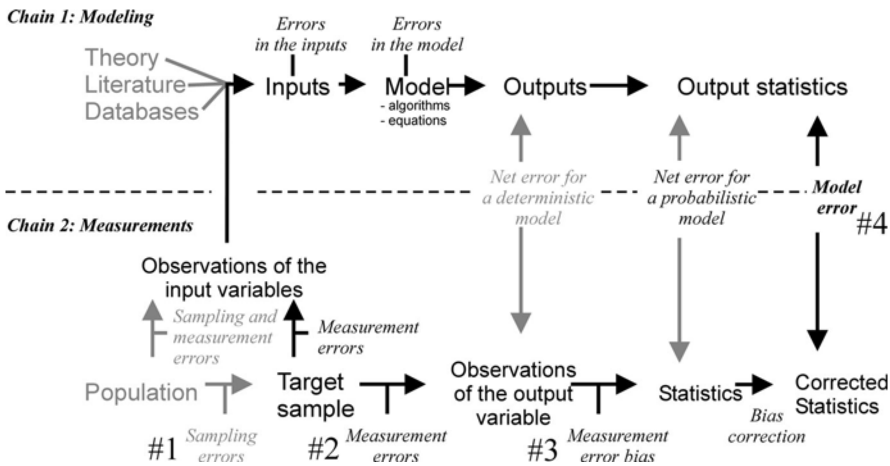

Figure 2. Comparison of modeling results to observations includes all errors along the two chains. The terms quantified in Figure 6 are indicated by #1–#4.

MATERIALS AND METHODS

Model Description

The current work defines exposure as the time-weighted average pollutant concentration experienced by the subject during a given averaging period. Exposures are modeled by a microenvironment approach in which the temporal and spatial variations of the concentration field experienced by the subject are reduced to the average concentration in each microenvironment during the subject's presence. Below is the mathematical representation of our definition of exposure:

$$\bar{E} = \frac{1}{t_{\text{avg}}}\sum_{i=1}^{u}(t_i \times C_i) = \sum_{i=1}^{n} f_i \times C_i \qquad (1)$$

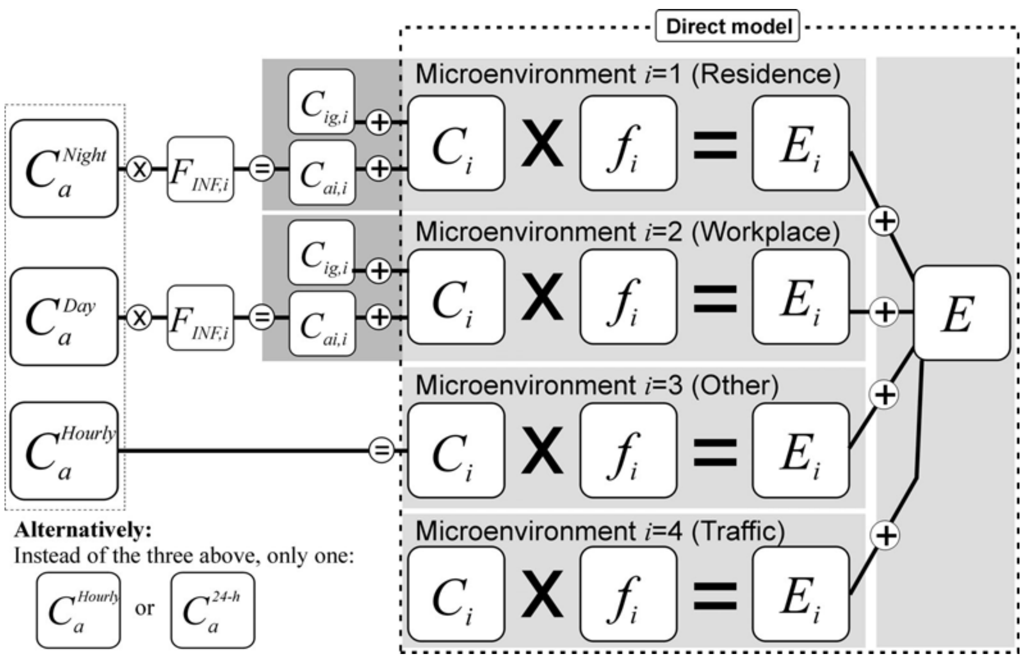

Figure 3. Diagram of the nested-model structure as an extension of the direct model. Symbols are defined in the text (see eqs 1 and 2).

where $\bar{E}$ is the average exposure level for the averaging time, t_{avg} ($=\Sigma t_i$); t_i is the time spent in the microenvironment i; C_i is the average concentration therein; and f_i is the fraction of time spent in each microenvironment, i.e., the ratio t_i/t_{avg}.

In the nested approach, the microenvironment concentrations in indoor environments are calculated from the corresponding ambient levels and indoor sources according to the following equation:[10,15-18]

$$C_i = F_{INF} C_a + C_{ig} \quad (2)$$

where C_i is the indoor concentration ($\mu g\ m^{-3}$); F_{INF} is the infiltration factor (dimensionless) for ambient particles; C_a is the ambient (outdoor) concentration ($\mu g\ m^{-3}$); and C_{ig} is the concentration of indoor generated particles ($\mu g\ m^{-3}$).

Major differences compared with the previous published work[14] include splitting of the earlier "other" microenvironment into "traffic" and "other" microenvironments, increasing the number of microenvironments from three to four, and use of the nested approach. The models were run with the *EXPOLIS* simulation framework[13] with Microsoft Excel version 8 (Seattle, WA) and @Risk add-on software version 4 (Palisade, Newfield, NY). Simulation settings included Latin hypercube sampling, 2000 iterations, and a fixed pseudorandom number seed.

Model Inputs

The simulation model inputs for concentrations, time activity, infiltration factors, indoor sources, and intervariable correlations were created by use of the *EXPOLIS*-Helsinki time–activity diaries and microenvironment measurements. The inputs used in the simulation models are listed in Table 1. The model outputs were compared with 48-hr personal exposure measurements. The *EXPOLIS* study design,[12] the population sampling,[19] and the $PM_{2.5}$ measurement techniques[20-22] have been published in more detail elsewhere. The *EXPOLIS* data were accessed from the *EXPOLIS* database dated September 2002.[23]

Subpopulations and Time Activities. The *EXPOLIS* population was classified into working (86.2%) and nonworking (13.8%) subpopulations according to the questionnaire

Table 1. Simulation inputs for the frequency distribution variables.

Input Category/Variable	Form	Parameters: Mean	Parameters: SD	Obs[a] (n)	Used in Models: Direct	Used in Models: Nested
Time–activity (% of time)						
Working subpopulation (86.2%)						
Residence indoors	β	57	8	374	x	x
Workplace	β	28	9	374	x	x
Traffic	β	8	6	374	x	x
Other combined	β	6	7	374	x	x
Nonworking subpopulation (13.8%)						
Residence indoors	β	85	13	60	x	x
Traffic	β	9	13	60	x	x
Other combined	β	7	7	60	x	x
Concentrations ($\mu g\ m^{-3}$)						
Residence indoors	Lognormal	8.9	5.7	126	x	
Workplace indoors	Lognormal	9.7	10.0	98	x	
Ambient night (17-07)	Lognormal	9.5	5.9	297		x
Ambient day (07-17)	Lognormal	9.1	6.1	297		x
Ambient 1-hr	Lognormal	9.6	6.8	7036	x	x[b]
Ambient 24-hr	Lognormal	9.3	5.4	298		x[b]
Traffic	Lognormal	17.2	13.9	37	x	x
Infiltration factors (fractions)						
Residences	β	0.64	0.20	98		x
Workplaces	β	0.47	0.24	94		x
Other	β	1.00	0.00	N/A[c]		x
Indoor sources ($\mu g\ m^{-3}$)						
General/residences	Lognormal	2.48	3.18	78		x
General/workplaces	Lognormal	4.18	4.98	41		x

[a]Number of observations used in parameter estimation; [b]Used in alternative nested models; [c]N/A, not applicable.

data. The time activities of these subpopulations were created separately by combining the original 11 diary categories into four microenvironments in the models: (1) residence; (2) workplace indoors (for the working sub population); (3) traffic; and (4) all other places. The diary entries, measured as minutes spent in each microenvironment, were transformed to corresponding fractions of time, and two-parameter β distributions were fitted to the fraction of time distributions by use of the common statistical technique of matching moments.[5] Goodness of similar fits was evaluated earlier.[14]

Concentrations. The Helsinki Metropolitan Area Council (YTV) supplied continuous $PM_{2.5}$ data measured during the *EXPOLIS* field phase in the Vallila monitoring station ~3.5 km northeast of the Helsinki downtown area. The data were measured by a β-radiation absorption technique with an Eberline FH 62 I-R analyzer. Comparison of fixed monitoring station data and the *EXPOLIS* residential outdoor measurements showed good correlation.[24]

Hourly $PM_{2.5}$ data were used to create the ambient inputs listed in Table 1. A lognormal probability distribution was fitted to each of these observed distributions by the method of matching moments.[5] In the simulations, the fitted lognormal distributions were truncated at the 99.9th percentile to prevent unrealistic concentration values created by the open-ended nature of the lognormal distribution. In reality, the upper concentration limit in the environment is set by the concentration in emission gases (e.g., exhaust gas) and mixing conditions. The upper limit depends on the location and the averaging time. The truncation percentile was selected based on literature and earlier results.[14,25]

During the *EXPOLIS* study in 1996–1997, $PM_{2.5}$ concentrations were measured in 37 vehicles. Four measurements in cars and taxis, and 20 in busses and trams were combined to describe concentrations in street traffic. Thirteen measurements in trains and metros were used to describe concentrations in rail traffic. The ambient 1-hr concentration distribution was used for walking and biking. Distribution of the average concentrations experienced while in traffic was simulated by use of these measurements and the *EXPOLIS* time–activity data. The resulting lognormal distribution ($17.2 \pm 13.9\ \mu g\ m^{-3}$) was used for the traffic microenvironment. This represents the total microenvironment concentration, including the ambient fraction, in-vehicle sources (if any), and the additional fraction caused by the immersion in the traffic flow.

Infiltration Factors and Indoor Sources. Ambient $PM_{2.5}$ infiltration factors for residences and workplaces were calculated by use of the microenvironment measurements and elemental sulfur data. The population average $PM_{2.5}$ infiltration was estimated by use of the slope for home indoor–home outdoor regression. The sulfur indoor/outdoor (i/o) ratio for each home was scaled to the ambient $PM_{2.5}$ infiltration factor by use of the ratio of average $PM_{2.5}$ infiltration to average sulfur i/o ratio.[16]

A similar approach was used for the workplaces, except that there were no simultaneous outdoor measurements of elemental $PM_{2.5}$ composition available. Particulate sulfate is a smoothly distributed secondary pollutant, and it was assumed that its ambient concentration in Helsinki does not have any significant diurnal pattern. The 2-night residential outdoor sulfur and 2-day workplace indoor sulfur concentrations were used to estimate the sulfur i/o ratios for the workplaces. Ambient concentrations, measured simultaneously with the workplace indoor $PM_{2.5}$ concentrations, were used to obtain the average $PM_{2.5}$ infiltration factor as the slope for the workplace–ambient regression. The workplace sulfur i/o ratios were finally scaled to the ambient $PM_{2.5}$ infiltration factors by use of the ratio of these averages.

The indoor $PM_{2.5}$ concentration of ambient origin was calculated by use of the infiltration factors and residential outdoor concentrations. The difference between the total measured indoor concentration in homes without environmental tobacco smoke (ETS) and the estimated contribution of ambient $PM_{2.5}$ was used as the indoor source concentration.[16] A similar calculation was repeated for the non-ETS workplaces.

The method of matching moments[5] was used to fit distributions with identical means and SD on the data. A β distribution was used for the infiltration factor, for which the two-parameter β distribution completely covers the theoretical range from zero to one. Lognormal distributions were used for the indoor-source–generated particle concentrations. The lognormal distribution fits were truncated at the 99.9th percentile.

Correlations. The @Risk simulation program supports rank-order correlation matrixes for modeling of statistical dependencies between input variables. For example, the fraction of time variables have negative correlations, by definition, because a large fraction of time spent in one microenvironment (e.g., the residence) will reduce the fraction of time left for the other microenvironments. The ambient concentrations are also correlated across all microenvironments, depending on the spatial and temporal variability of the target pollutant and variations in the infiltration factors between different microenvironments. Behavioral factors might also lead to correlations of personal-activity–related indoor source concentrations, but these were not included in the models.

Spearman rank correlations were calculated with Stata software, version 5 (Stata Corp., College Station, TX) for the residential and occupational indoor concentrations (direct model), the three different ambient concentrations (nested model), and the time activities (both models; Table 2). The rank-order correlation of traffic measurements with the simultaneous ambient concentrations was 0.21.

Model and Measurement Errors

The model error is defined here as

$$y = x + \varepsilon_{\mathrm{model}} \tag{3}$$

where y is the modeled value (population mean, SD, or given percentiles in the current work), x is the corresponding true value, and $\varepsilon_{\mathrm{model}}$ is model error.

To calculate the model error, the true value (x) must be known. This is impossible, however, because of the measurement and other errors inherent to any empirical observations. The simple classical model for measurement error can be written as[26]

Table 2. Spearman rank correlation matrixes used in the simulation.

Time Activities				
	Working			
Nonworking	Home Indoor	Workplace	Traffic	Other
Home indoor	1	N/A[a]	−0.56	−0.71
Workplace	−0.39	1	N/A	N/A
Traffic	−0.27	−0.37	1	(0.0)[b]
Other	−0.42	−0.37	(−0.1)	1
Concentrations: Direct Model				
	Home Indoor	Workplace	A1h	Traffic
Home indoor	1			
Workplace	0.40	1		
A1h	0.83	0.57	1	
Traffic	(0.2)	(0.2)	(0.2)	1
Concentrations: Nested Model				
	A–N	A–D	A1h	Traffic
A–N	1			
A–D	0.69	1		
A1h	0.73	0.69	1	
Traffic	(0.2)	(0.2)	(0.2)	1

Notes: The time–activity matrix uses the lower half of the working and the upper half of the nonworking subpopulations. The concentration matrixes are identical for both subpopulations; [a]N/A = not applicable; A1h = ambient 1-hr; A–N = ambient night; A–D = ambient day; [b]Values in parentheses are not statistically significant ($P > 0.5$); all other are statistically significant ($P < 0.05$).

$$z = x + \varepsilon_{\mathrm{obs}} \tag{4}$$

where z is the observed value, x is the true value, and $\varepsilon_{\mathrm{obs}}$ is the measurement error.

The measurement error above may also include effects of the measurement on the behavior of the subject. Comparison of the model outputs (y) with corresponding observations (z) allows us to calculate the net error, a combination of the model and measurement errors:

$$y - z = (x + \varepsilon_{\mathrm{model}}) - (x + \varepsilon_{\mathrm{obs}}) = \varepsilon_{\mathrm{model}} - \varepsilon_{\mathrm{obs}} = \varepsilon_{\mathrm{net}} \tag{5}$$

The error introduced to distribution parameters (SD and percentiles) is much smaller than the measurement error itself. Because a random measurement error widens the observed distribution, the lower percentiles are biased downward and the upper percentiles and SD upward. The bias can be corrected for if the statistical properties of the measurement errors are known, e.g., from duplicate measurements. Each duplicate contains an unknown effect of the measurement error; thus, the difference of two duplicates is a sum of two measurement errors. Assuming nondifferential random measurement errors, the statistical properties can be estimated from a sample of duplicate measurements.

Using the *EXPOLIS* duplicate personal exposure measurements (n = 14 in Helsinki), we created a statistical measurement error model. The duplicates showed relative and absolute error components; at higher levels, relative errors caused probably mainly by the volumetric flow control of the personal monitor pump were apparent (coefficient of variation = 17%). At lower concentrations, the absolute errors caused by the weighing procedure and filter handling became dominant (SE = 0.92 μg m^{-3}). The final error model is a combination of these terms with the switch point concentration set to 5.4 μg m^{-3}, giving identical errors (Figure 4).

The effects of the measurement error on the observed exposure distribution was estimated by simulation. The true (but unknown) values were described by a lognormal maximum likelihood fit to the observations.[27] A random, nonbiased normally distributed error term with the properties described above was then added. Differences of the percentiles of the observed data and the error simulation were used as the measurement error bias. To estimate the measurement error bias uncertainty, we simulated two additional models: one with one-half and the other with double error. The simulations were conducted with 10,000 samples. Finally, the observed personal exposure distribution was corrected for the measurement error bias by subtracting the bias estimates from the corresponding

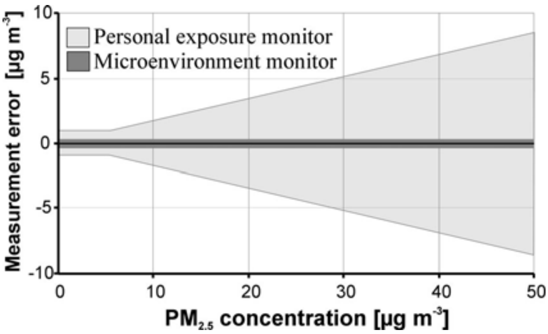

Figure 4. Comparison of measurement error models for personal and microenvironment monitors.

distribution values. The model errors were then calculated as differences between modeled and bias-corrected observed values.

RESULTS

Measurement Error Bias

The observed exposure distribution mean, SD, and main percentiles and the corresponding measurement error biases are shown in Table 3. In an absolute sense, the bias terms are small. The maximum value of 2 μg m^{-3} for the 99th percentile indicates that, from the point of view of exposure assessment, the measurement error bias can be mostly ignored. Relative bias values varied between −3 and +6% for the main percentiles, but in the lower tail, +10% was exceeded (Figure 5). Graphical inspection confirmed that the bias (the difference between the top of the gray area and the line with open circles) was notable only in the tails of the distribution.

Simulation Results

The exposure levels of the lowest and highest 1% of the population differed by ~10-fold (3–30 μg m^{-3}). From this point of view, both the direct and nested models captured the exposure variability quite well (Figure 5). The simulated mean exposure was overestimated by just 1% by the direct model and underestimated by 5% by the nested model (Table 3). Both models, however, underestimated the SD more clearly, by −9 and −23% by the direct and nested models, respectively. Both models overestimated values between the 20th and 50th percentiles and underestimated values in both tails. In the high tail, underestimation was larger in an absolute sense, but lower in relative terms. On examination of the model errors, the role of the measurement error bias becomes more apparent; e.g., underestimation of the 99th percentile in the direct model is reduced by 50% compared with the net error.

We ran two alternative nested simulations to demonstrate situations in which only 1- or 24-hr distribution of ambient concentrations is available (Figure 3). Use of only a single distribution for ambient concentrations led to effectively 100% correlation between the ambient concentrations used for calculation of residential and occupational indoor concentrations, and the distribution variances also reflected the averaging times slightly. For short averaging times, the variance increased and vice versa. Thus, as can be expected, the alternative simulation for 1-hr ambient concentrations produced slightly more variable exposure estimates (9.4 ± 5.2 μg m^{-3}), reducing underestimation of the SD. Similarly, the simulation with 24-hr data decreased exposure variability (9.2 ± 4.5

Table 3. Observed and simulated distributions and corresponding net errors and bias-corrected model errors.

Distribution	Mean	SD	Percentiles 25	50	75	90	95	99
EXPOLIS personal non-ETS exposure measurements								
Observed values (n = 126) (μg m^{-3})	9.8	6.4	5.5	7.7	12.1	16.7	19.7	33.6
Measurement error bias (μg m^{-3})	0.0	0.3	−0.14	−0.20	0.02	0.4	0.8	2.0
Relative bias (%)	+0	+5	−3	−3	+0.2	+2	+4	+6
Bias-corrected exposure (μg m^{-3})	9.8	6.1	5.6	7.9	12.1	16.3	18.9	31.6
Direct model	9.8	5.6	6.0	8.5	12.0	16.8	21.0	29.6
Net error (μg m^{-3})	0.1	−0.9	0.5	0.9	−0.1	0.1	1.2	−3.9
Model error (μg m^{-3})	0.1	−0.6	0.4	0.7	−0.1	0.5	2.0	−2.0
Relative model error (%)	+1	−9	+7	+8	−1	+3	+11	−6
Nested model	9.3	4.7	6.1	8.3	11.2	15.3	18.0	25.9
Net error (μg m^{-3})	−0.5	−1.7	0.7	0.6	−0.9	−1.3	−1.7	−7.7
Model error (μg m^{-3})	−0.5	−1.4	0.5	0.4	−0.9	−1.0	−0.9	−5.7
Relative model error (%)	−5	−23	+9	+5	−7	−6	−5	−18

μg m^{-3}), increasing the underestimation of SD. The overall changes in the simulated distributions, however, were smaller than model errors of the nested model in corresponding percentiles.

Model Error and Other Error Terms

The error terms quantified in the current work are depicted in Figure 6. In exposure assessment, the most significant error term is the sampling error (#1 in Figure 6; calculated from the observed data by the method described, e.g., by Small [28] and Campbell and Gardner[29]). It should be noted that this error term represents only the statistical uncertainty about the true distribution from which the sample has been drawn, assuming an ideal random sample. Any problems caused by nonrepresentativeness of the sample add to the sampling error. In the current work, sampling error was excluded by study design.

The measurement error for a single observation (#2) and the bias caused to the estimated percentiles (#3) are shown in the middle chart in Figure 6. Although the measurement error in the *EXPOLIS*-Helsinki measurements caused a 17% unsolvable relative uncertainty around individual personal exposures >5.4 μg m^{-3} and a 0.92 μg m^{-3} absolute uncertainty below that level, the uncertainty in the population percentiles was much smaller and could be corrected for. This level of measurement error is not very significant for exposure assessment, but it increases the apparent model errors (i.e., net errors) significantly; therefore, in a model validation study, it is reasonable to take the bias into account. The 0.5*x* and 2*x* measurement error bands, demonstrating the effects of uncertainty in the measurement error itself, indicate that if the measurement error can be reduced to one-half, it will no longer have notable effects. On the other hand, if bad measurement techniques or poor quality control lead to doubling of the measurement error, its effects increase significantly in the tails of the distribution.

Model errors (#4) of the direct and nested models are shown in the bottom chart of Figure 6. Model errors in the percentiles varied between −2 and +1 μg m^{-3} below the 80th percentile for the nested model and the 90th percentile for the direct model. Above these percentiles, the model errors increased suddenly, approaching the magnitude of the sampling error, indicating a more serious underestimation of exposures for the highest 10% of the population. This is a clear indication that some microenvironments and activities that affect exposures for the top 10% of the population are missing or not fully described in these models. In relative terms, the peak underestimation approached but did not exceed −20 and −30% for the direct and nested models, respectively.

If the models were evaluated with use of net errors (i.e., without correcting the measurement error bias), the errors were overestimated. Several model-validation studies have shown that simulation models tend to underestimate high exposures and exposure variances.[6,8,14] This finding may be explained in part by the effects of the measurement error bias.

The errors in the direct model were smaller than those in the nested model, but the overall shape of the errors was the same. The direct model is more accurate, but it requires measurements of microenvironment concentrations, whereas the nested model can be used when only ambient concentration data are available. The nested model requires infiltration factor inputs, but such data have recently become available from both Europe and the United States,[16,30] and the values do not seem to be highly variable among different cities or geographical areas.

A significant advantage of microenvironment-based exposure modeling is the fact that the measurement error is much smaller in microenvironment monitoring than in personal exposure measurements. Figure 4 shows the measurement error models built from the *EXPOLIS*-Helsinki duplicate data. The personal exposure measurement included a significant relative error term, whereas the microenvironment monitor duplicate data (n = 41) showed no such increase of measurement error as a function of the concentration. Because of the higher flow rate and larger samples, the absolute error term caused by filter

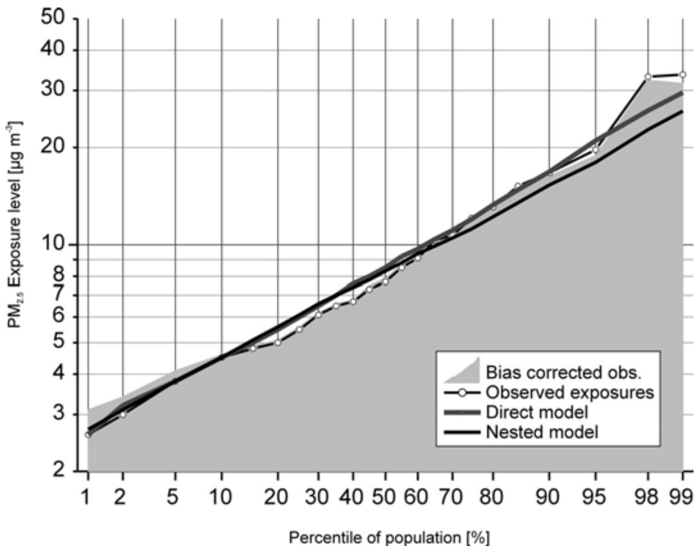

Figure 5. Graphical comparison of observed and simulated exposure distributions. The gray-shaded area represents the observed exposure distribution.

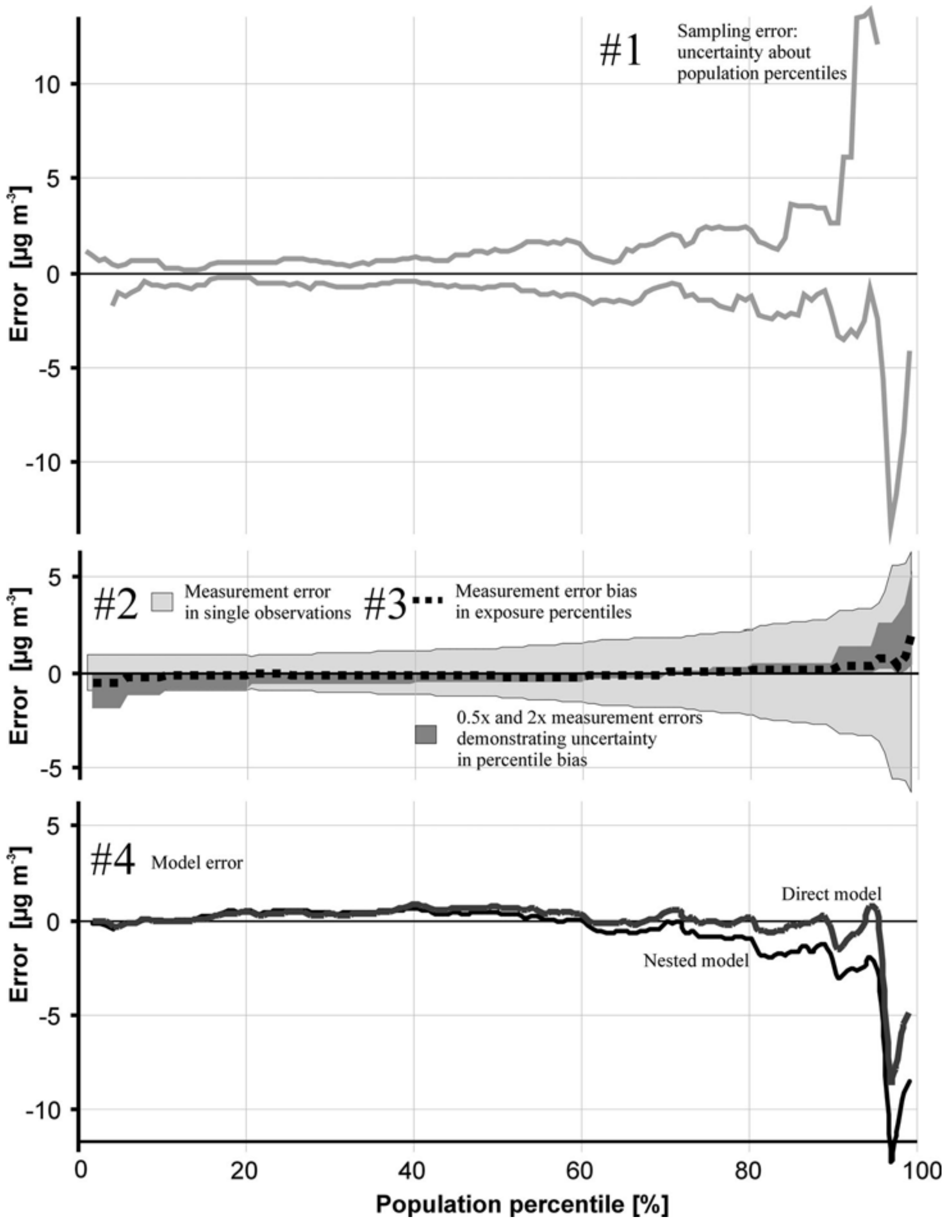

Figure 6. Sampling error (#1) is excluded here by study design. Measurement error in individual exposures (#2) is larger than the bias carried to sample percentiles (#3). Model error (#4) is comparable to measurement error bias below the 80th percentile and with sampling error above the 80th percentile.

weighing and handling was also lower (SE = 0.30 $\mu g\ m^{-3}$). Thus, when the model inputs are created from microenvironment and fixed monitoring station measurement data, one significant error source, the measurement error, is significantly reduced compared with personal exposure measurements.

Simulated Indoor Concentrations

The simulated residential indoor concentration distributions matched the corresponding observed distributions well for both the direct and nested models (Figure 7). For the occupational indoor concentrations, the direct and nested models differed more clearly. The direct model

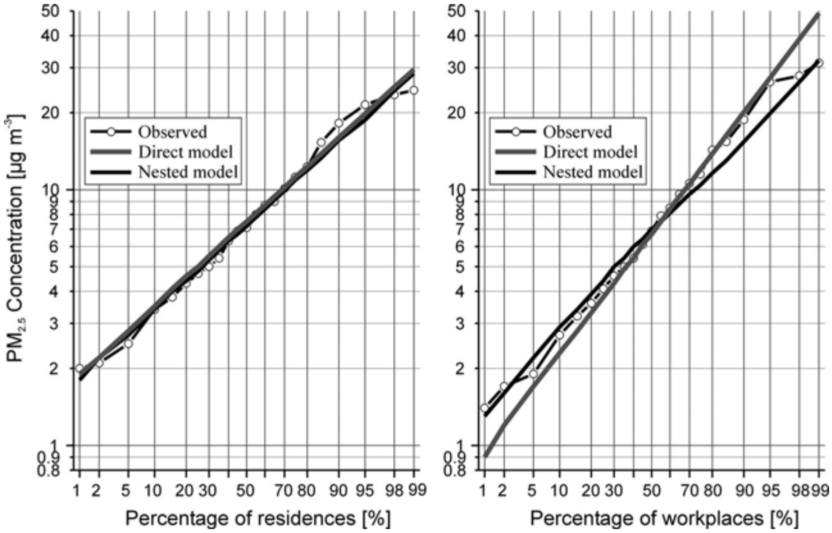

Figure 7. Comparison of observed residential and occupational indoor concentrations with the fitted lognormal distribution used in the direct model and the simulated distribution produced by the nested model.

captured the levels between the 50th and 95th percentiles well but overestimated the levels for the highest 5% and underestimated levels below the 30th percentile. On the other hand, the nested model matched the observations reasonably well below the 60th percentile and above the 98th percentile but underestimated levels in between. These slight problems in simulating the workplace concentrations probably indicate special occupational exposures that were not specifically included in the model. It must be also noted that the infiltration factors for the workplaces were estimated by use of residential leisure time outdoor measurements, which caused uncertainty in the infiltration factor distribution.

DISCUSSION

Sources of Model Errors

Sources of uncertainty in numerical models are typically grouped into three classes: (1) model uncertainty; (2) parameter uncertainty; and (3) scenario uncertainty. Model uncertainty concerns whether the conceptual model and its implementation as equations and algorithms really capture the essential causal relationships of the modeled system. Issues such as selecting the correct variables to describe the behavior of the system belong to this class. Measurement and sampling errors (including representativeness) and errors caused by proxy variables are responsible for parameter uncertainty. Scenario uncertainty arises in situations in which the modeled system does not exist (yet) and some or all its properties are based on assumptions. This situation is typical for models predicting the behavior of a system in the future.

In the current work, these classical sources of uncertainty can be interpreted in terms of being responsible for the model errors. The basic conceptual model in the current work defines the personal exposure as the time-weighted average concentration experienced by an individual. This is in fact the definition of the modeled variable and thus contains no model uncertainty. A similar example would be the area of a rectangle, which is calculated (both modeled and defined) as the product of the width and the length. Such a model will definitely include parameter uncertainty because of the imperfections of measurement of the width and the length, but model uncertainty would be involved only in cases when the assumption of the shape of the object (rectangle) would fail.

Use of the same known system, the *EXPOLIS*-Helsinki population sample, to collect the input data (on concentrations and time activities) and the personal exposures used for comparison removed scenario uncertainties and sampling errors from the study design. The exposure and input parameter measurements are independent of each other because they were conducted with different equipment. Thus the set-up really allows for quantification of model errors in a known system.

Although the overall match between the simulations and the observations was good, a significant remaining source of model error in the current setup is the definition of the microenvironments. In addition to residence, workplace, and traffic, all other environments were grouped together, including both outdoor environments and microenvironments with high concentrations attributable to local sources (e.g., parking garages). From the modeling point of view, this group of environments is problematic because of its heterogeneous nature. It would be a significant undertaking to conduct representative concentration measurements in all microenvironments belonging to this category, and in most cases, such data do not exist. The purpose of combining all of these environments together and using the hourly ambient concentration distribution to describe the exposures in them in the current work was to quantify the errors caused by such a simplification. Our conclusion in this respect is that when looking at the total $PM_{2.5}$ exposures, the effect of overlooking specific concentration distributions in various minor microenvironments is small. When looking at exposures to specific particle fractions (e.g., diesel particles), the separation of concentrations for time spent (e.g., in the vicinity of traffic arteries) versus time spent elsewhere might be necessary.

Underestimation of Variance

All current simulation models underestimate variance similarly to earlier studies.[6,8,14] The reasons for this include incomplete definition of high-concentration microenvironments; incomplete description of behavioral and environmental correlations, including concentrations of ambient origin in various microenvironments; behavioral factors affecting microenvironment concentrations as well as personal time–activity structures; and measurement error bias leading to slight overestimation of the sample variances.

The current work includes intervariable correlations obtained with the @Risk rank-correlation technique. The technique, using Spearman rank-correlation coefficients, partly reproduces the correlation structures. In the case of skewed distributions, such as lognormal ones, the high tail values dominate the Pearson linear correlation. When the simulation model reproduces correlations by use of the rank-correlation technique, part of the correlation is diluted in the low-tail values, which do not have a notable effect on exposures. Another reason for variance underestimation here could be the correlation between the time fraction spent in traffic and the concentrations experienced while in the traffic. Preliminary data analysis by Rotko et al.[31] show that men, who spend more time in traffic in general, also spend more time specifically in road traffic (diary categories: motorcycle, car/taxi, and bus/tram), particularly cars, in which the concentrations are higher than in rail traffic or when walking.

As reported earlier,[14] the SD of a simulated exposure distribution was sensitive to the single highest sample taken from lognormal input distributions. The earlier solution to truncate all lognormal distributions at the 99.9th percentile was followed, producing the best matching SD estimates in the earlier work. Truncation of the lognormal distributions does not suggest that under extreme conditions such high concentrations could not exist (although Ott[25] used a theoretical approach to show that the high concentrations fall short of the levels predicted by the lognormal density function). Truncation limits the concentrations to a range that is represented by the input data.

Weak correlations were observed in the infiltration factors and ambient concentrations in the *EXPOLIS* Helsinki data (data not shown). The infiltration factor varies according to the time of the year (e.g., keeping windows open during the summer increases ventilation rates). Because the ambient pollutant concentrations also vary among seasons, a correlation structure will emerge. In the present work, this correlation was so low that it did not appear to affect the simulations.

Time–Activity Modeling

Actual diaries randomly sampled from a time–activity database have been used in many models.[6,8-10] The main advantage of this approach is that it captures the complex autocorrelation structures of time–activity data. The current work, however, uses parametric distribution fits to describe fractions of time spent in different microenvironments. This β-fit approach allows very efficient description of the time–activity data. When the model is applied for alternative nonexisting scenarios, changes according to the estimated changes in population time–activity distributions are easy to make and to describe compared with modification of thousands of actual diaries.

Theoretically, the use of a probabilistic time–activity model instead of actual diaries could explain the bias of the results toward the mean value. In the earlier studies, however, in which actual diaries were used and the results were validated against observed data,[6,8] the underestimation of variance was larger than in the current work.

CONCLUSIONS

The current work interprets "model validation" as a demonstration that a model can predict the output variables in a known system with such accuracy that conclusions derived from the model are not driven by model errors. To address this problem quantitatively, we calculated model errors and compared them with the other errors affecting exposure assessments: random sampling error, random measurement error, and the bias in observed percentiles caused by the measurement error. These error terms, ranked in increasing order, were measurement error bias in sample distribution $<$ model error $<$ measurement error $<$ sampling error, except for the highest percentiles, for which model error approached the magnitude of sampling error. If the bias caused by the measurement errors is not corrected for, it increases the apparent model error (i.e., net error) for the highest exposure percentiles and the exposure variance.

Direct and nested simulation approaches were compared. The overall matches between the models and observations were good, indicating that such models can be used to characterize population exposures to $PM_{2.5}$. The model based directly on input distributions of indoor concentrations produced slightly better estimates of population parameters than the nested model using ambient concentrations and infiltration factors to model indoor concentrations. For both models, the errors for the mean exposure levels were slightly negative and for the SD clearly negative, indicating that the models underestimate these values. The suggested main reasons were as follows: (1) Correlation structures caused by behavioral and personal factors between concentrations in different microenvironments were not fully modeled. A rank-order correlation technique was used to include some of these correlations in the model, but the method loses some of its effect in correlating samples taken from the low end of the distributions, whereas the high concentrations drive variances of the exposures. (2) The available microenvironments and the concentration measurements in residences and workplaces did not capture all exposures, particularly those that occurred in proximity of local sources (e.g., exposures to traffic-generated particles while in areas affected directly by traffic).

ACKNOWLEDGMENTS

This work was supported by EU Contracts N ENV4-CT96-0202 (*EXPOLIS*; DG12-DTEE) and EVK4-CT-2002-00097 (FUMAPEX); Academy of Finland Contracts 36586, 40835 and 42610; and intramural funding by KTL, the National Public Health Institute of Finland, and other institutions in the other *EXPOLIS* centers.

REFERENCES

1. Oreskes, N.; Shraderfrechette, K.; Belitz, K. Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences; *Science* **1994**, *263,* 641-646.
2. Konikow, L.F.; Bredehoeft, J.D. Groundwater Models Cannot Be Validated; *Adv. Water Resour.* **1992**, *15,* 75-83.
3. Anderson, M.P.; Woessner, W.W. The Role of the Postaudit in Model Validation; *Adv. Water Resour.* **1992**, *15,* 167-173.
4. Cullen, A.C.; Frey, H.C. *Probabilistic Techniques in Exposure Assessment*; Plenum Press: New York, 1999.
5. Morgan, M.G.; Henrion, M. *Uncertainty*; Cambridge University Press, New York, 1990.
6. Ott, W.; Thomas, J.; Mage, D.; Wallace, L. Validation of the Simulation of Human Activity and Pollutant Exposure (Shape) Model Using Paired Days from the Denver, CO, Carbon Monoxide Study; *Atmos. Environ.* **1988**, *22,* 2101-2113.
7. Frey, H.C.; Rhodes, D.S. Characterizing, Simulating and Analyzing Variability and Uncertainty, an Illustration of Methods Using an air Toxics Emissions Example; *Hum. Ecol. Risk Assess.* **1996**, *2,* 762-797.
8. Law, P.L.; Lioy, P.J.; Zelenka, M.P.; Huber, A.H.; McCurdy, T.R. Evaluation of a Probabilistic Exposure Model Applied to Carbon Monoxide (PNEM/co) Using Denver Personal Exposure Monitoring Data; *J. Air & Waste Manage. Assoc.* **1997**, *47,* 491-500.
9. Freijer, J.I.; Bloemen, H.J.T., deLoos, S.; Marra, M.; Rombout, P.J.A., Steentjes, G.M.; van Veen, M.P. Modelling Exposure of the Dutch Population to Air Pollution; *J. Hazard. Mater.* **1998**, *61,* 107-114.
10. Burke, J.M.; Zufall, M.J.; Özkaynak, H. A Population Exposure Model for Particulate Matter: Case Study Results for PM2.5 in Philadelphia, PA; *J. Expo. Anal. Environ. Epidemiol.* **2001**, *11,* 470-489.
11. Yeh, S.; Small, M.J. Incorporating Exposure Models in Probabilistic Assessment of the Risks of Premature Mortality from Particulate Matter; *J. Expo. Anal. Environ. Epidemiol.* **2002**, *12,* 389-403.
12. Jantunen, M.J.; Hänninen, O.; Katsouyanni, K.; Knoppel, H.; Kuenzli, N.; Lebret, E.; Maroni, M.; Saarela, K.; Sram, R.; Zmirou, D. Air Pollution Exposure in European Cities: The "EXPOLIS" Study; *J. Expo. Anal. Environ. Epidemiol.* **1998**, *8,* 495-518.
13. Kruize, H.; Hänninen, O.O.; Breugelmans, O.; Lebret, E.; Jantunen, M. Description and Demonstration of the EXPOLIS Simulation Model: Two Examples of Modeling Population Exposure to Particulate Matter; *J. Expo. Anal. Environ. Epidemiol.* **2003**, *13,* 87-99.
14. Hänninen, O.O.; Kruize, H.; Lebret, E.; Jantunen, M. EXPOLIS Simulation Model: PM2.5 Application and Comparison With Measurements in Helsinki; *J. Expo. Anal. Environ. Epidemiol.* **2003**, *13,* 74-85.
15. Wilson, W.E.; Mage, D.T.; Grant, L.D. Estimating Separately Personal Exposure to Ambient and Nonambient Particulate Matter for Epidemiology and Risk Assessment: Why and How; *J. Air & Waste Manage. Assoc.* **2000**, *50,* 1167-1183.
16. Hänninen, O.O.; Lebret, E.; Ilacqua, V.; Katsouyanni, K.; Künzli, N.; Srám, R.J.; Jantunen, M.J. Infiltration of Ambient PM2.5 and Levels of Indoor Generated Non-ETS PM2.5 in Residences of Four European Cities; *Atmos. Environ.* **2004,** *38,* 6411-6423.
17. Dockery, D.W.; Spengler, J.D. Indoor-Outdoor Relationships of Respirable Sulfates and Particles; *Atmos. Environ.* **1981**, *15,* 335-343.
18. Ryan, P.B.; Spengler, J.D.; Letz, R. Estimating Personal Exposures to NO2; *Environ. Int.* **1986**, *12,* 395-400.
19. Rotko, T.; Oglesby, L.; Künzli, N.; Jantunen, M. Population Sampling in European Air Pollution Exposure Study, EXPOLIS: Comparisons between the Cities and Representativity of the Samples; *J. Expo. Anal. Environ. Epidemiol.* **2000**, *10,* 355-364.
20. Koistinen, K.J.; Kousa, A.; Tenhola, V.; Hänninen, O.O.; Jantunen, M.J.; Oglesby, L.; Künzli, N.; Georgoulis, L. Fine Particle (PM2.5) Measurement Methodology, Quality Assurance Procedures, and Pilot Results of the EXPOLIS Study; *J. Air & Waste Manage. Assoc.* **1999**, *49,* 1212-1220.
21. Mathys, P.; Stern, W.B.; Oglesby, L.; Braun-Fahrländer, C.; Ackermann-Liebrich, U.; Jantunen, M.J.; Künzli, N. Elemental Analysis of Airborne Particulate Matter by ed-xrf within the European EXPOLIS Study; *ICP Inf. Newsl.* **2001**, *27,* 29-34.
22. Hänninen, O.O.; Koistinen, K.J.; Kousa, A.; Keski-Karhu, J.; Jantunen, M.J. Quantitative Analysis of Environmental Factors in Differential Weighing of Blank Teflon Filters; *J. Air & Waste Manage. Assoc.* **2002**, *52,* 134-139.
23. Hänninen, O.O.; Alm, S.; Kaarakainen, E.; Jantunen, M. *The EXPOLIS Databases*. B13/2002. Publications of the National Public Health Institute. KTL: Kuopio, Finland, 2002.
24. Kousa, A.; Oglesby, L.; Koistinen, K.; Kunzli, N.; Jantunen, M. Exposure Chain of Urban air PM2.5—Associations between Ambient Fixed Site, Residential Outdoor, Indoor, Workplace and Personal Exposures in Four European Cities in the EXPOLIS-Study; *Atmos. Environ.* **2002**, *36,* 3031-3039.
25. Ott, W.R. A Physical Explanation of the Lognormality of Pollutant Concentrations; *J. Air & Waste Manage. Assoc.* **1990**, *40,* 1378-1383.
26. Wacholder, S. When Measurement Errors Correlate with Truth—Surprising Effects of Nondifferential Misclassification; *Epidemiology* **1995**, *6,* 157-161.
27. Ott, W.R. *Environmental Statistics and Data Analysis*; Lewis Publishers, CRC Press: Boca Raton, FL, 1995.
28. Small, M. Probability Distributions and Statistical Estimation. In *Uncertainty*; Morgan, M.G., Henrion, M., Eds. Cambridge University Press: New York, 1990, pp 73–101.
29. Campbell, M.J.; Gardner, M.J. Calculating Confidence Intervals for Some Non-Parametric Analyses. In *Statistics with Confidence*; Gardner, M.J.; Altman, D.G., Eds. British Medical Journal: London, 1989, pp. 71–79.
30. Özkaynak, H.; Xue, J.; Spengler, J.; Wallace, L.; Pellizzari, E.; Jenkins, P. Personal Exposure to Airborne Particles and Metals: Results from the Particle TEAM Study in Riverside, California; *J. Expo. Anal. Environ. Epidemiol.* **1996**, *6,* 57-78.
31. Rotko, T.; Hänninen, O.; Jantunen, M. Time Used in Traffic by Sociodemographic Groups in EXPOLIS-Helsinki. In *Proceedings of the Annual Conference of ISEA/ISEE*; ISEA/ISEE 2002 Secretariat: Vancouver, Canada, 2002.

About the Authors

Otto O. Hänninen, M.Sc., and Jouni T. Tuomisto, M.D., are researchers and Matti J. Jantunen, Ph.D., is a research professor in the Centre of Excellence for Environmental Health Risk Analysis in the Division of Environmental Health, the Finnish National Public Health Institute (KTL), Kuopio, Finland. Erik Lebret, Ph.D., is the head of the Department of Environmental Health Research at the National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. Address correspondence to: Otto Hänninen, Division of Environmental Health, the Finnish National Public Health Institute (KTL), P.O. Box 95, FIN-70701 Kuopio, Finland; e-mail: otto.hanninen@ktl.fi.

VI

Hänninen, O.O., Palonen, J., Tuomisto, J., Yli-Tuomi, T., Seppänen, O., Jantunen, M.J., 2005. **Reduction potential of urban $PM_{2.5}$ mortality risk using modern ventilation systems in buildings**. *Indoor Air*. In press (published as *OnlineEarly* in the Internet).

Indoor Air 2005
www.blackwellpublishing.com/ina
Printed in Singapore. All rights reserved

INDOOR AIR
doi:10.1111/j.1600-0668.2005.00365.x

Reduction potential of urban $PM_{2.5}$ mortality risk using modern ventilation systems in buildings

Abstract Urban $PM_{2.5}$ (particulate matter with aerodynamic diameter smaller than 2.5 μm) is associated with excess mortality and other health effects. Stationary sources are regulated and considerable effort is being put into developing low-pollution vehicles and environment-friendly transportation systems. While waiting for technological breakthroughs in emission controls, the current work assesses the exposure reductions achievable by a complementary means: efficient filtration of supply air in buildings. For this purpose infiltration factors for buildings of different ages are quantified using Exposures of Adult Urban Populations in Europe Study (EXPOLIS) measurements of indoor and outdoor concentrations in a population-based probability sample of residential and occupational buildings in Helsinki, Finland. These are entered as inputs into an evaluated simulation model to compare exposures in the current scenario with an alternative scenario, where the distribution of ambient $PM_{2.5}$ infiltration factors in all residential and occupational buildings are assumed to be similar to the subset of existing occupational buildings using supply air filters. In the alternative scenario exposures to ambient $PM_{2.5}$ were reduced by 27%. Compared with source controls, a significant additional benefit is that infiltration affects particles from all outdoor sources. The large fraction of time spent indoors makes the reduction larger than what probably can be achieved by local transport policies or other emission controls in the near future.

O. O. Hänninen[1], J. Palonen[2], J. T. Tuomisto[1], T. Yli-Tuomi[1], O. Seppänen[2], M. J. Jantunen[1]

[1]KTL, Centre for Environmental Health Risk Analysis, Kuopio, [2]Laboratory of Heating, Ventilating and Air Conditioning, Helsinki University of Technology, Espoo, Finland

Key words: Exposure assessment; Infiltration; Modeling; PM_2; Public health risk; Supply air filtration; Ventilation.

Otto O. Hänninen
KTL, Centre for Environmental Health Risk Analysis, PO Box 95, FIN-70701 Kuopio, Finland
Tel.: +358 17 201 171
Fax: +358 17 201 184
e-mail: otto.hanninen@ktl.fi

Received for review 7 October 2004. Accepted for publication 7 April 2005.

Practical Implications
It has been suggested that indoor concentrations of ambient particles and the associated health risks can be reduced by using mechanical ventilation systems with supply air filtering in buildings. The current work quantifies the effects of these concentration reductions on population exposures using population-based data from Helsinki and an exposure model. The estimated exposure reductions suggest that correctly defined building codes may reduce annual premature mortality by hundreds in Finland and by tens of thousands in the developed world altogether.

Introduction

Epidemiologists have shown that urban fine particulate matter ($PM_{2.5}$: particulate matter with aerodynamic diameter smaller than 2.5 μm) concentration is associated with increased risk of premature mortality (e.g. Pope et al., 2002). The observed risk ratios translate to hundreds of thousands of annual excess deaths in the developed world at the prevailing $PM_{2.5}$ levels. Although successful restrictions have been set on industrial and energy production emissions and a lot of work has been done in developing low-emission motor vehicles to reduce exposures to particles from these sources, significant exposures still remain. Besides the remaining emissions from these sources, particles are generated by sources that are more difficult to control by local policies, like natural sources and distant sources contributing to long-range transport. In Helsinki it has been estimated that up to 76% of ambient $PM_{2.5}$ originates from long-range transport (Karppinen et al., 2004; Koistinen et al., 2004; Vallius et al., 2003).

Many studies have shown that personal PM exposures correlate poorly with ambient concentrations (Koistinen et al., 2001; Pellizzari et al., 1999) and that indoor sources make remarkable contributions to personal exposures (Clayton et al., 1993; Koistinen et al., 2004; Wallace, 1996). The health effects observed in the epidemiological studies, however, must be caused by ambient PM (or some factor closely associated with it), and not by exposures to indoor-generated particles, which do not correlate with the ambient pollution levels (Wilson et al., 2000). The additional personal exposures caused by individual behavior and independent indoor sources may, of course, be responsible for additional health effects that are not associated with ambient concentrations.

It has been suggested that ventilation systems in buildings could protect people from ambient particles (Fisk et al., 2002). In the warm and humid climate areas in the US, where sealed and air conditioned buildings are most common, the dose–response rate for PM_{10}-induced morbidity was found to be lower than in the milder climate areas, where open windows are used more for ventilation, indicating a safety factor created by the sealed building envelopes (Janssen et al., 2002). Similarly, in Canada residents of new energy efficient homes experienced less air quality-related symptoms than the control group members (Leech et al., 2004). People in developed countries spend a majority of their time indoors (Clayton et al., 1993; Hänninen et al., 2003) and thus filtration of ambient pollution by building envelopes can be expected to be an important exposure modifier. In residential buildings, where mechanical ventilation systems have been rare, outdoor particles penetrate indoors very efficiently (penetration factors close to unity) (Özkaynak et al., 1996; Wallace, 1996), but in buildings with two-way mechanical ventilation particle removal by supply air filters has been identified as the most significant particle removal process (Thornburg et al., 2001). Indoors particles are slowly removed from the air due to deposition and other decay processes even in houses with no supply air filtering (Hänninen et al., 2004; Wallace, 1996). In mechanical ventilation systems particle removal can be accelerated by recirculating indoor air through the filters (Fisk et al., 2002).

In Helsinki metropolitan area $< 1\%$ of homes built before 1990 have supply air filters, but these are becoming increasingly common in new buildings. The recently renewed National Building Code of Finland (section D2, 2003) requires mechanical ventilation with heat recovery and efficient fine particle filtration of supply air in urban areas. Since 2000 a majority of single-family houses have been equipped with mechanical supply and exhaust ventilation system with supply air filtration. Mechanical supply and exhaust ventilation system with supply air filtration was used in 78% of the existing office buildings in Helsinki already in 1990 (Jaakkola and Miettinen, 1995) and 83% of office employees were working in such buildings. Since then all new office buildings have been equipped with mechanical supply and exhaust air ventilation systems.

Fisk et al. (2002) estimated performances of various supply air filters on indoor particle concentrations using a mass-balance model. According to their results, up to 80% reductions in indoor concentrations of ambient fine particles can be achieved with realistic filter efficiencies and flow rates. Such a modeling study, however, is based on assumptions on filter efficiencies, air leaks, particle penetration rates through the building envelopes (Airaksinen et al., 2004), and indoor particle decay rates. In reality also the behavior of the inhabitants affects the indoor concentrations; efficiency of even the best filtration system is reduced when windows or doors are kept open. Therefore the theoretical estimates calculated by Fisk et al. (2002) must be validated by using real life observations.

The objective of the current work is to compare the theoretical reductions estimated by Fisk et al. (2002) with the values observed in the Helsinki metropolitan area building stock in the Exposures of Adult Urban Populations in Europe Study (EXPOLIS) (Hänninen et al., 2004a; Jantunen et al., 1998). In addition, to support air pollution exposure control policy optimizations, a probabilistic simulation model is used to estimate how much the mechanical ventilation systems with supply air filtration, if assembled to the whole building stock, residential and occupational, could reduce population exposure to ambient $PM_{2.5}$.

Material and methods

The conceptual exposure model used in this work is shown in Figure 1. The adult population in Helsinki metropolitan area spends on average 87% of their time in indoor environments; approximately 8% in traffic (including walking) and only 5% in non-traffic outdoor environments. Therefore decreasing infiltration of particles indoors significantly reduces overall exposure levels to particles of ambient origin.

Scenarios

The current work defines two exposure scenarios. The current scenario is based on the prevailing situation in 1996–97 when the population-based EXPOLIS study was conducted in the Helsinki metropolitan area. A random sample of adults was drawn and exposures and concentration in the residences and workplaces of the subjects were measured. Infiltration factors for the ambient $PM_{2.5}$ were calculated using indoor and outdoor measurements of $PM_{2.5}$ concentrations and corresponding PM-bound elemental sulfur levels (Hänninen et al., 2004). In the alternative scenario the infiltration properties of the future building stock of the 21st century are approximated by using the infiltration factors observed in the newest occupational buildings built in the 1990s, which were captured in the EXPOLIS workplace sample, i.e. existing buildings that all use mechanical ventilation systems with F7 or F8 class supply air fine particle filters with 80–95% collection efficiencies for 0.4 μm particles. A probabilistic simulation model (Hänninen et al., 2003, 2005;

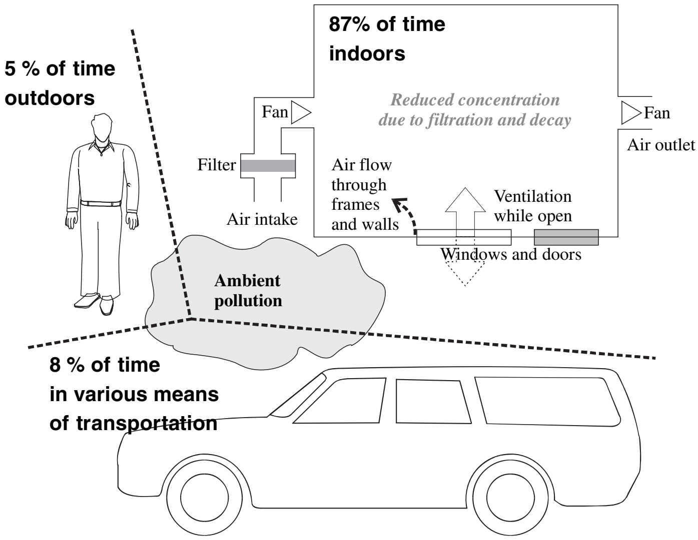

Fig. 1 Schematic diagram of the exposure model used in this study: a major fraction of the population exposure to ambient $PM_{2.5}$ occurs indoors. The effect of supply of air filtration, which is an efficient means to reduce these exposures, is quantified for the existing building stock in Helsinki

Kruize et al., 2003) is applied to estimate the population distributions of 48-h exposures for these two scenarios.

Simulation model

The simulation model used is based on microenvironment approach (Duan, 1982; Letz et al., 1984; Ryan et al., 1986) and probabilistic simulation (Law et al., 1997; Ott et al., 1988). The model defines personal exposure level (E) as the time-weighted average concentration (C) over the microenvironments (indexed by i) visited. According to Equation 1, time weighting is done using personal time activities as fractions of time (f_i) spent in each microenvironment, implicitly defining the averaging time:

$$E = \sum_i f_i \times C_i. \quad (1)$$

The simulation model has been validated for $PM_{2.5}$ exposures in two steps. First, the model was used in microenvironment mode, where the concentrations in the microenvironments are directly defined with parameters of log-normal distributions (Hänninen et al., 2003). In the second step the indoor microenvironment concentrations (C_i) were modeled from ambient concentration according to Equation 2:

$$C_i = F_{\mathrm{inf}} \times C_{\mathrm{a}} + \sum_j C_{Sj}, \quad (2)$$

where F_{inf} is the infiltration factor and C_{a} the ambient $PM_{2.5}$ concentration. The additional concentrations (C_{Sj}) caused by various sources (indexed by j) within the microenvironment are then added to the concentration of ambient origin ($F_{\mathrm{inf}} \times C_{\mathrm{a}}$). Infiltration factor can be estimated as the slope of indoor–outdoor concentration regression (Hänninen et al., 2005).

The simulations were run using four microenvironments: (i) residential indoors, (ii) workplace indoor (working subpopulation only), (iii) in traffic, and (iv) all other environments grouped together (Hänninen et al., 2005).

Input data

The model inputs were calculated from the EXPOLIS database (Hänninen et al., 2002). EXPOLIS study was conducted in seven European cities in 1996–2000, including Helsinki, Finland. Fine PM exposures, corresponding residential and occupational concentrations and exposure-related characteristics of the residences, workplaces and time activities of the subjects were measured from a random sample of the adult urban populations. The study design has been described by Jantunen et al. (1998), the collection of the PM data by Koistinen et al. (1999), the X-ray-induced fluorescence analysis of the $PM_{2.5}$ samples by Mathys et al. (2001) and the calculation of the infiltration factors by Hänninen et al. (2004). Elemental sulfur had no notable indoor sources (i.e. indoor–outdoor ratios above unity) in the data and the sulfur indoor–outdoor ratio was assumed to represent the effective infiltration factor for those fine particles that have a similar size distribution as the sulfur-containing particles. The sulfur infiltration factors were corrected for the slightly different size distribution of $PM_{2.5}$ particles using the ratio of corresponding indoor–outdoor regression coefficients (Hänninen et al., 2004). For occupational buildings simultaneous outdoor sulfur measurements were not available; these data were substituted with corresponding residential outdoor concentrations. It was assumed that as a secondary long-range transported pollutant the sulfur concentrations do not have significant spatial or diurnal patterns and that the two-night average residential concentration is a reasonable estimate for the 2-day occupational outdoor concentration.

Distribution of the ambient $PM_{2.5}$ concentration was formed from hourly ambient $PM_{2.5}$ concentrations, monitored by the Helsinki Metropolitan Area Council (YTV). The 6854-h time series data was measured during the study field phase at Vallila monitoring station, located approximately 3.5 km north-east from the Helsinki downtown area, using β-radiation absorption-based Eberline FH 62 I-R analyzer. Non-positive data (182 h) were discarded before fitting the log-normal distribution to the concentration data using method of matching moments (i.e. using mean and standard deviation values). Indoor concentrations in residences and workplaces were probabilistically modeled using the ambient concentration distribution and Equation 2. Residential and occupational concentrations of indoor sources were estimated from the EXPOLIS data and modeled assuming log-normal distributions (Hänninen et al., 2004, 2005). Log-normal traffic concentration distribution was simulated using the 37 in-transport measurements conducted during the EXPOLIS study and the population time activities (Hänninen et al., 2005). The ambient concentration distribution described above was used directly for the other microenvironment. The model input values are listed in Table 1.

Time activities of the working and non-working adult populations were modeled separately. The time activity data for the 11 microenvironments in the EXPOLIS Helsinki database for 434 subjects was grouped into four microenvironment categories and transformed into fractions of time spent in each during the 48-h diary collection period. In the model time activity values were sampled from beta distributions for each microenvironment and scaled for the sum of unity for each simulated individual.

Four simulation models were run. For the current scenario a model was run for the total $PM_{2.5}$ exposures, including exposures from non-ETS (environmental tobacco smoke) indoor sources (model 1) and for the exposures to ambient $PM_{2.5}$ (model 2). Similar models were run for the alternative scenario (models 3 and 4 respectively). The total non-ETS exposures for the current scenario (model 1) were simulated for validation purposes and compared with the personal exposure distribution observed in the EXPOLIS study.

Table 1 Model input distributions and parameters used in the simulations. Models columns indicate in which models (1–4) each input was used

Input category	Data distribution	Parameters: Mean	s.d.	Obs[a] (n)	Models: 1	2	3	4
Time-activity (fractions of time, %)								
Working subpopulation (86.2%)								
Home indoors	beta	57	8	374	×	×	×	×
Workplace	beta	28	9	374	×	×	×	×
Traffic	beta	8	6	374	×	×	×	×
Others	beta	6	7	374	×	×	×	×
Non-working subpopulation (13.8%)								
Home indoors	beta	85	13	60	×	×	×	×
Traffic	beta	9	13	60	×	×	×	×
Others	beta	7	7	60	×	×	×	×
$PM_{2.5}$ concentrations ($\mu g/m^3$)								
Ambient 1-h	log-normal	9.6	6.8	7036	×	×	×	×
Traffic	log-normal	17.2	13.9	37	×	×	×	×
Infiltration factors (fractions)								
Current building stock scenario								
Homes	beta	0.64	0.20	98	×	×		
Workplaces	beta	0.47	0.24	94	×	×		
Building stock 1990s scenario								
Homes[b]	beta	0.35	0.12	n/a			×	×
Workplaces	beta	0.35	0.12	9			×	×
Indoor sources for $PM_{2.5}$ ($\mu g/m^3$)								
General home source	log-normal	2.48	3.18	78	×		×	
General work source	log-normal	4.18	4.98	41	×		×	

[a]Number of observations used in parameter estimation.

[b]Workplace data used also for residences.

The natural negative autocorrelations of time fractions and correlation between the ambient concentration and concentration experienced while in traffic were modeled using the rank correlation technique provided by the @Risk software (Palisade, Newfield, NY). The rank correlation values varying between −0.1 and −0.7 for time fractions and between 0.2 and 0.7 for concentrations were analyzed from the EXPOLIS data and have been reported in detail earlier (Hänninen et al., 2005).

Results

The infiltration factors for ambient $PM_{2.5}$ in the residential buildings are higher (mean ± s.d.: 0.64 ± 0.20) than those in the occupational buildings (0.47 ± 0.24, Figure 2, Table 2). More efficient filtration of ambient particles in the occupational buildings is presumably caused by the facts that supply air filtering is more common in office buildings and that ventilation by opening windows is more common in residential buildings. The 90-day running averages (Figure 2) show a slight seasonal pattern for both types of buildings, following the average seasonal temperatures. For both building types there are some outliers above the theoretical upper limit of 1.0, caused by (i) indoor sources of sulfur (especially in two workplaces with $PM_{2.5}$ infiltration values of 2.8 and 3.6, which were excluded from the analysis), (ii) time delay from outdoor PM via infiltration to indoor levels, (iii) by

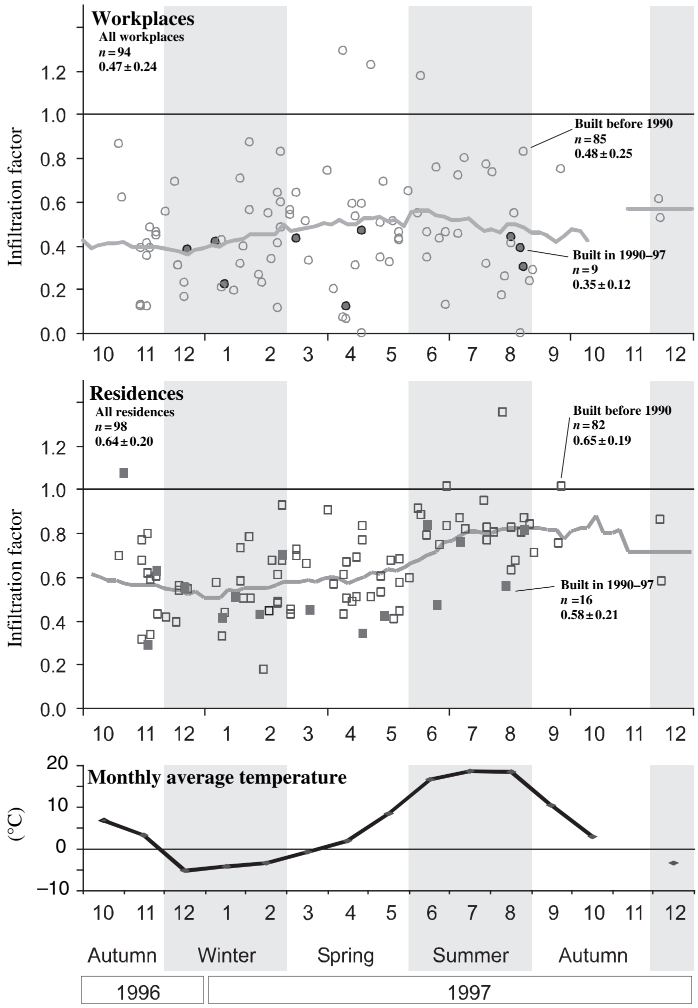

Fig. 2 Comparison of $PM_{2.5}$ infiltration factors for workplaces and residences (mean ± s.d.). Solid markers indicate newer buildings built between 1990 and the study field phase in 1996–97. The gray solid lines represent 90-day running averages for all buildings. Monthly average temperatures are shown in the bottom chart as an important modifier for building ventilation adjustments

Table 2 Infiltration factors observed in different EXPOLIS building categories and values used to describe scenarios

	Construction before 1990		Construction 1990–97		
	Filtering prevalence (%)	Observed infiltration (fraction)	Filtering prevalence (%)	Observed infiltration (fraction)	Reduction (%)
Residences	<1	0.65	n/a	0.58	11
Workplaces	78	0.48	100	0.35	27
	(1) No filtering used		(2) Filtering in all buildings		
Scenario values	<1	0.65	100	0.35	46

n/a, not available.

measurement errors, and (iv) PM concentration difference between the outdoor monitoring site and actual air intake location. Despite of these minor shortcomings, the overall distributions of infiltration factors are plausible.

A log-normal fit to the observed ambient fixed monitoring station concentration data was used in the simulations (Figure 3). The adjusted coefficient of determination (R^2) calculated from the observed concentration data using values from the fitted log-normal function with identical z-score values, was 0.98, i.e. 98% of the observed variation in the ambient concentration could be modeled indicating a very good fit. The same ambient concentration model was used for both scenarios.

Besides the building infiltrations, population time activity is the most important factor affecting the exposure reduction potential modeled here. The more people spend time in indoor environments, the larger effect the building filtration properties have on their exposures. On individual level the time activity is very variable, as can be seen in Figure 4a. The histograms in these charts describe the distribution of the observed fractions of time spent indoors, outdoors, and in traffic according to the 434 time activity diaries collected in the EXPOLIS study in Helsinki. The population average for the fraction of time spent in indoor environments is 87%. The overlaid beta distribution in each chart depicts the technique used to model the time activity distributions in the simulations; in the simulations the number of microenvironments was four for the working and three for the non-working subpopulations (totaling seven time activity classes; parameters of these distributions are listed in Table 1).

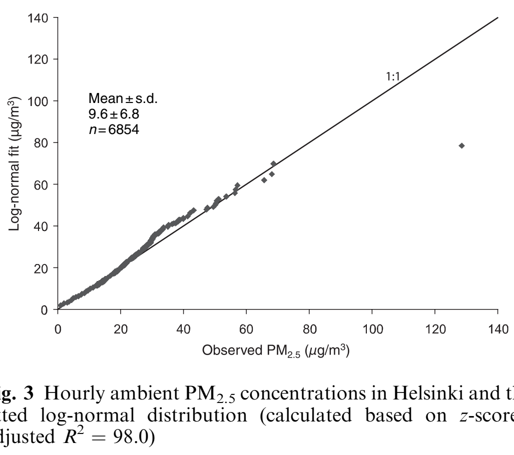

Fig. 3 Hourly ambient $PM_{2.5}$ concentrations in Helsinki and the fitted log-normal distribution (calculated based on z-scores; adjusted $R^2 = 98.0$)

From the point of view of generalizing the Helsinki results to other cities in Europe or elsewhere, it is important to look at the differences in the state of the art of building construction and ventilation technology for residential and occupational buildings, including the infiltration properties, and the population time activity patterns. To demonstrate that the time use differences between urban populations in Europe are small, the population averages for indoors, outdoors and in traffic fractions of times observed in the EXPOLIS study are compared in Figure 4b. The average fraction of time spent indoors varies from 0.86 in Athens (Greece) to 0.89 in Grenoble (France) and Milan (Italy), being thus nearly constant. Therefore it can be concluded that if there are differences between geographical areas in the efficiency of the suggested approach to reduce exposures, they must be driven by the differences in buildings and occupant behavior.

Simulated total exposures in current scenario (model 1) compare well with the observations (Figure 5a). For the highest percentiles the model underestimates the levels slightly. The observed mean exposure level is 9.8 $\mu g/m^3$ and simulated 9.3 $\mu g/m^3$. Thus the overall underestimation is 0.5 $\mu g/m^3$, or 5%. The corresponding standard deviation values were 6.4 and 4.7 $\mu g/m^3$, respectively, having larger underestimation in both absolute and relative terms. This could be expected, because standard deviation of a skewed distribution is more sensitive to underestimation of the high-tail values and consequently is not a very stable statistic for such distributions. The overall match between the two distributions is reasonable: the model is capable of catching 95% of the population exposures and can thus be considered valid for the following analyses.

Modeled mean exposure levels to ambient $PM_{2.5}$ were 6.9 and 5.0 $\mu g/m^3$ for the current and alternative

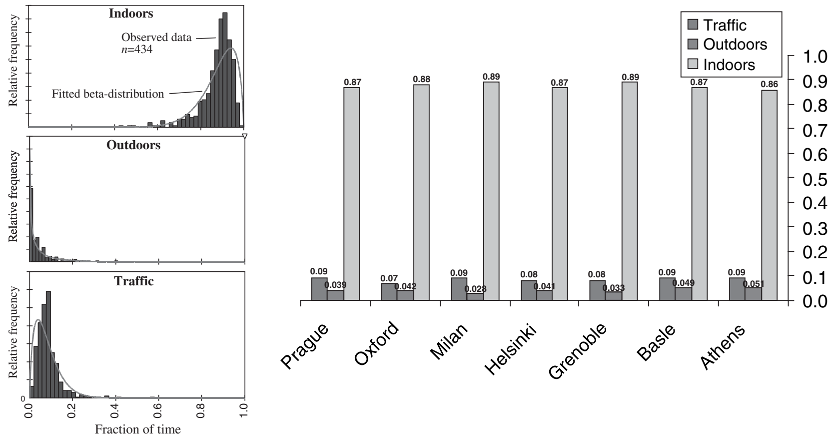

Fig. 4 Histograms of population variability of time-activity in Helsinki (a) and comparison different EXPOLIS cities (b). While the within city variability between individuals is significant, the differences in city averages are almost negligible, especially for the time fraction spent indoors. For the other two categories the difference is relatively larger

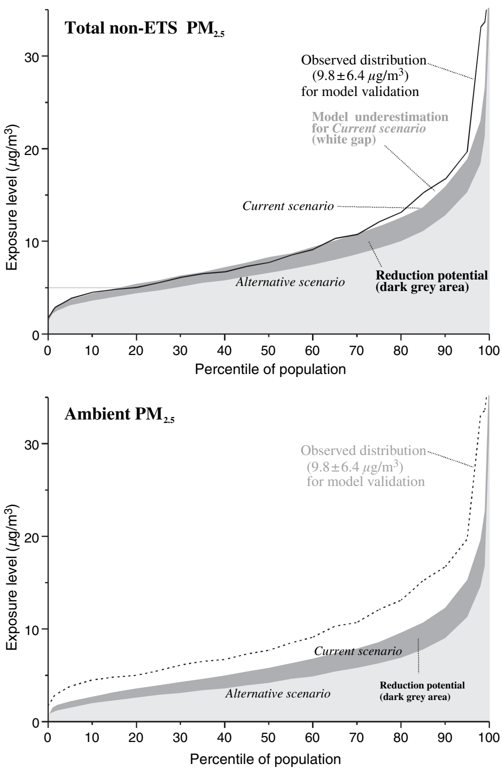

Fig. 5 The observed non-ETS $PM_{2.5}$ exposure distribution, corresponding simulated exposure, and estimated exposure reduction potential (a) and same for the exposures to ambient $PM_{2.5}$ particles (b). Dark gray area represents the reduction potential; the top edge of the gray area is the model result for the current and the bottom edge for the alternative scenario

scenarios, respectively, indicating a 27% reduction potential (Table 3). This main result of the current work is graphically depicted in Figure 5b, where the difference between the scenarios is shown in darker shade of gray. As both axes are printed on linear scales, the areas under the curves are proportional to the corresponding risks. Reduction affects all percentiles as can be expected, but the absolute reduction is largest around the 70th to 90th percentiles, i.e. where the exposure levels are rather high. This can be considered as an advantage: exposures can be reduced efficiently by using filtration systems in buildings in polluted areas. For the highest percentiles the effectiveness gets smaller, corresponding to relatively rare personal activities that lead into high exposures.

Current approach assumes that concentrations of indoor-generated particles would not be affected in the alternative scenario. While this assumption is reasonable when focusing on the ambient exposures to which health effects have been mostly associated, the indoor-generated concentrations can also be lowered with changes in the ventilation system, e.g. by using indoor air recirculation through filters. Simultaneously with

Table 3 Simulated mean population exposures to ambient, indoor-generated and total $PM_{2.5}$, and the corresponding risk reduction estimates (%) based on the linear exposure-response factor

Exposure fractions	Current scenario ($\mu g/m^3$)	Alternative scenario ($\mu g/m^3$)	Exposure reduction (%)
Ambient $PM_{2.5}$	6.9	5.0	−27
Indoor sources	2.5	2.5	0
Total $PM_{2.5}$ exposure	9.3	7.5	−20
Indoor % of ambient	36	49	–

the lowering ambient exposures in alternative scenario, the relative magnitude of indoor-generated non-ETS exposure increases from 36 to 49% (Table 3). If the indoor-generated particles turn to be toxic at all, their role in the PM question will become more important as the ambient part is alleviated.

The simulated exposure results can be translated to reduction in the ambient $PM_{2.5}$-associated health risks by using the generally adopted no-threshold linear dose–response relationship (WHO, 2000). This assumption suggests that a reduction in the health risk, e.g. mortality, is proportional to the reduction in the exposure. When looking at the main focus of the current work, the ambient exposures, the current scenario exposure level 6.9 $\mu g/m^3$ reduces to 5.0 $\mu g/m^3$ in the alternative scenario, a 27% reduction in the exposure and thus potentially a similar risk reduction. Taking the World Health Organization estimate that the annual number of deaths associated with ambient $PM_{2.5}$ levels in Europe is 102,000–368,000 (WHO, 1999), the estimated reduction would turn to be in the order of 27,000–100,000 deaths per year in Europe.

Discussion

To compare the theoretical reductions of ambient $PM_{2.5}$ in indoor air obtainable with supply air filters as estimated by Fisk et al. (2002) with respective observations, the buildings in the EXPOLIS sample were classified into two age categories divided by construction before or after 1 January 1990. The technical specifications of ventilation systems of the EXPOLIS buildings were not collected, but over three quarters of office buildings constructed before 1990 already had mechanical ventilation with supply air filtration (Jaakkola and Miettinen, 1995). Some of the EXPOLIS workplaces were not located in office buildings, so it can be expected that the prevalence of supply air filtering in the EXPOLIS workplaces is somewhat lower. Less than 1% of residences built before 1990 use supply air filtering. Residences built in the 1990s started to introduce mechanical ventilation with supply air filters and all office buildings built in 1990s were designed with mechanical ventilation with supply air filters. Consequently, the old residences (built before 1990) represent a reference building stock, where filtration systems are practically absent. The old occupational buildings (built before 1990) and the newer residences built in 1990s represent mixed building stocks, and in the occupational buildings built in 1990s a vast majority uses mechanical ventilation with supply air filtration.

In the EXPOLIS Helsinki sample there were nine occupational buildings built after 1 January 1990 and 16 corresponding residential buildings. For both building types the newer buildings had smaller infiltration factor values than the pre-1990 buildings, but the difference was much larger for the workplace buildings (Table 2). Fisk et al. (2002) estimated that the levels of ambient $PM_{2.5}$ could be reduced approximately by 23, 51 and 80% when using fine particle filters with classification ASHRAE 45, 65 and 85% (efficiencies as defined in standard ASHRAE, 1992), respectively, compared with ventilation without filter. In their base case they assumed 1 h^{-1} mechanical outside air ventilation, 0.25 h^{-1} unfiltered ventilation and 4 h^{-1} indoor air recirculation through the filters, representing a North American one-family house with forced air heating system. The estimate for ASHRAE 65% class filters (51%) is close to the observed reduction of 46% for the building categories 'all with filters' versus 'none with filters' (Table 2). Out of this reduction potential, the current building stock in 1996–97 had already established reductions of 2 and 28% for residences and workplaces, respectively, calculated as the proportion of current building stock infiltration values to that of the reference building stock of old residences. In comparison, the theoretical maximum of 80% reported by Fisk et al. (2002) indicates that with the building technology to be developed in the 21st century, significant benefits remain to be achieved.

The $PM_{2.5}$ fraction responsible for the observed excess mortality has not been identified conclusively yet regardless of the significant effort put to study this problem. $PM_{2.5}$ is composed of fractions including long-range transported particles of different types, tailpipe particles from local traffic, combustion particles from local stationary sources, crustal particles generated and/or re-suspended by road traffic and natural processes, salt particles associated both with natural processes as well as road de-icing in colder climates. In the current situation the mean population exposure level to ambient fine particles, observed as $PM_{2.5}$ mass concentration, is still the most widely accepted health-relevant PM measure. Primary combustion-generated particles from local sources are very small, typically smaller than 100 nm in diameter These ultrafine particles behave differently in the filtering and ventilation systems. Especially their removal rate in indoor air is lower than that for the larger particles which comprise a majority of $PM_{2.5}$ mass. It has been suggested that the ultrafine particles have health effects different from those of $PM_{2.5}$; it should be noted that the results obtained here for $PM_{2.5}$ particles are not representative for the ultrafines.

Filtration by the building envelopes reduces exposures to particles from all ambient sources. The filters in mechanical ventilation systems are capable of removing $PM_{2.5}$ particles with high efficiency. When windows or doors are kept open, suspended particles of all sizes penetrate indoors with equal efficiency ($\sim$100%). Only when outdoor air penetrates indoors through small cracks, holes, and fibrous insulation materials in the building envelope does the infiltration result in particle size-dependent losses. For larger particles the dominant mechanisms are sedimentation and impaction, for the smallest interception and diffusion. Accumulation mode particles have the highest penetration efficiency (Kulmala et al., 1999; Raunemaa et al., 1989; Tung et al., 1999). The same physical phenomena reduce the PM concentrations after they have penetrated into indoor spaces. When the air leaks are minimized, the exposure reduction affects particles of all sizes and a risk reduction can be expected regardless of future findings of the role of different $PM_{2.5}$ fractions in causing the premature mortality associated with ambient $PM_{2.5}$.

The current work simulated the exposure reduction for the active working age population. The suggested approach, however, affects the exposures of all residents without any behavioral changes. Susceptible population groups, like the newborns and the elderly, spend more of their time indoors and less in traffic compared with the working age population; thus they would benefit the most from exposure reduction affecting indoor environments. Because buildings are designed, built, and renewed one by one, the ventilation system specifications reducing $PM_{2.5}$ exposures can be targeted to selected buildings, geographical areas, and population groups.

Renewing of the urban building stock is expensive and occurs gradually along the natural renovation and re-construction process. The same, however, is more or less true also for most local outdoor source control alternatives. People concerned about air pollution can act accordingly and select residences in sealed building envelopes and with good filtration systems. To support this, information on the filtration properties of houses should be made available. However, ventilation systems themselves can become sources of pollution (Pasanen et al., 1994) and therefore it is important also to maintain the ventilation systems properly.

Enhancements of city transportation system and changes of local traffic emissions and population time activity affect mainly exposures to local traffic particles. Based on published data (Koistinen et al., 2004; Vallius et al., 2003) we estimate that in Helsinki particles from local traffic contribute approximately 10–20% to the total $PM_{2.5}$ exposures. Compared with the exposure reduction potential estimated in the current work, the tailpipe $PM_{2.5}$ emissions from local traffic should be totally eliminated to obtain similar reductions in the total $PM_{2.5}$ exposures. Battery- or fuel cell-operated vehicles might eliminate traffic tailpipe emissions in the decades to come, but even then exposure to re-suspended soil particles and to industry and energy production-generated long-range particles would not be affected. In contrast, filtration by building envelope affects particles from local and regional sources as well as long-range transport, and its potential is not limited to our simulation results, which only reflect the ongoing business as usual policy.

The risk reduction potential is estimated using data from Helsinki, a city with northern location and population of 1 million. Because of the northern climate, triple glazing is standard in most buildings and the current building stock may also be in other ways tighter than buildings e.g. in the Mediterranean area, Central Europe or Southern states in the US. Thus it can be expected that the infiltration of $PM_{2.5}$ is similar or larger in most parts of the developed world and that the reduction potential could thus be even larger. Janssen et al. (2002) looked at the relationships between the health outcomes, including chronic obstructive pulmonary disease, cardiovascular disease and pneumonia, associated to ambient PM_{10} and prevalence of air conditioning systems in 14 cities in the US. In comparison with open window ventilation, a sealed building with air-conditioning considerably reduces PM infiltration. Consequently they found out that the prevalence of air conditioning reduced the concentration–response slope, especially for cardiovascular diseases, suspected to be the most common primary cause of premature death linked to PM. This result indicates that the reduced exposures in mechanically ventilated sealed buildings indeed do reduce morbidity and mortality, and supports the idea that the building envelope and ventilation system design can be used to reduce $PM_{2.5}$ risks also in warmer climates than Helsinki.

Slower air exchange rates lead to decreased infiltration due to the longer air residence times and particle decay processes. However, it is known that low air exchange rates lead to poor indoor air quality caused by indoor sources of CO_2 and other compounds (Lin and Deng, 2003; Thornburg et al., 2004; Wong and Huang, 2004). The concentrations caused by indoor sources are proportional to the air exchange rate and would be increased if air exchange rates would be reduced. Although the exposures to pollution of ambient origin would be reduced due to lower infiltration rates in such situation, the net effect could be worsening of total exposures and potentially increasing health risks. Moreover, poorly designed building structures can lead to moisture condensation and consequent mold problems, having both economical and health consequences. All these issues must be carefully considered when planning exposure reduction policies.

Conclusions

Engineering buildings and their ventilation systems in a way that minimizes the infiltration of fine particles indoors is an efficient way to reduce population exposures to PM and corresponding health risks. In the EXPOLIS Helsinki data the $PM_{2.5}$ infiltration efficiencies for the residential and office buildings built after 1990 were clearly lower than for the older buildings and especially for the occupational buildings, where the mechanical ventilation systems with supply air filters became standard in 1990s. If the non-ETS-exposed working age population in Helsinki lived and worked in buildings with similar filtration efficiencies as the occupational buildings built after 1990, their $PM_{2.5}$ exposures would be reduced by 27% in comparison with the current situation.

Advantages of filtration by ventilation systems compared with other local exposure reduction alternatives include:

- Exposures to particles from all ambient sources are reduced;
- The reduction can be targeted to susceptible sub populations;
- Making building filtration property information available so that people can select their residences according to their concern for air pollution;
- The benefits of reducing ambient air concentrations indoors can be further amplified by indoor air recirculation.

Based on the generally accepted no-threshold linear dose–response model for the ambient $PM_{2.5}$, any exposure reduction will lead to a proportional reduction in $PM_{2.5}$-induced mortality and other health effects. The public health benefit potential can be tens of thousands saved lives per year in both Europe and North America. Improvement of the ambient air quality, however, is necessary and the primary means to reduce these exposures in the long run.

Acknowledgements

This work has been supported by EU Contracts N ENV4-CT96-0202 (EXPOLIS, DG12-DTEE) and EVK4-CT-2002-00097 (FUMAPEX), Academy of Finland Contracts 36586, 40835 and 42610, and intramural funding by KTL (Finnish National Public Health Institute) and other institutions in other centers.

References

Airaksinen, M., Kurnitski, J., Pasanen, P. and Seppanen, O. (2004) Fungal spore transport through a building structure, *Indoor Air*, **14**, 92–104.

ASHRAE (1992) *Gravimetric and Dust-Spot Procedures for Testing Air Cleaning Devices Used in General Ventilation for Removing Particulate Matter*. ANSI/ASHRAE Standard 52.1. Atlanta, GA, American Society of Heating, Refrigerating, and Air Conditioning Engineers.

Clayton, C.A., Perrit, R.L., Pellizzari, E.D., Thomas, K.W., Whitmore, R.W., Wallace, L.A., Özkaynak, H. and Spengler, J.D. (1993) Particle total exposure assessment methodology (PTEAM) study: distributions of aerosol and elemental concentrations in personal, indoor and outdoor air samples in an southern California company, *J. Expo. Anal. Environ. Epidemiol.*, **3**, 227–250.

Duan, N. (1982) Models for human exposure to air pollution, *Environ. Int.*, **8**, 305–309.

Fisk, W.J., Faulkner, D., Palonen, J. and Seppanen, O. (2002) Performance and costs of particle air filtration technologies, *Indoor air*, **12**, 223–234.

Hänninen, O.O., Alm, S., Kaarakainen, E. and Jantunen, M. (2002) *The EXPOLIS Databases*. B13/2002. Kuopio, KTL, Publications of the National Public Health Institute.

Hänninen, O.O., Kruize, H., Lebret, E. and Jantunen, M. (2003) EXPOLIS Simulation Model: PM2.5 application and comparison with measurements in Helsinki, *J. Expo. Anal. Environ. Epidemiol.*, **13**, 74–85.

Hänninen, O.O., Lebret, E., Ilacqua, V., Katsouyanni, K., Künzli, N., Srám, R.J. and Jantunen, M.J. (2004a) Infiltration of ambient PM2.5 and levels of indoor generated non-ETS PM2.5 in residences of four European cities, *Atmos. Environ.*, **38**, 6411–6423.

Hänninen, O.O., Lebret, E., Tuomisto, J.T. and Jantunen, M.J. (2005) Characterization of model error in the simulation of PM2.5 exposure distributions of the working age population in Helsinki, Finland, *J. Air Waste Manage. Assoc.*, **55**, 446–457.

Jaakkola, J.J.K. and Miettinen, P. (1995) Type of ventilation system in office buildings and sick building syndrome, *Am. J. Epidemiol.*, **141**, 755–765.

Janssen, N.A.H., Schwartz, J., Zanobetti, A. and Suh, H.H. (2002) Air conditioning and source-specific particles as modifiers of the effect of PM10 on hospital admissions for heart and lung disease, *Environ. Health. Perspect.*, **110**, 43–49.

Jantunen, M.J., Hänninen, O.O., Katsouyanni, K., Knöppel, H., Künzli, N., Lebret, E., Maroni, M., Saarela, K., Srám, R.J. and Zmirou, D. (1998) Air pollution exposure in European cities: The EXPOLIS-study, *J. Expo. Anal. Environ. Epidemiol.*, **8**, 495–518.

Karppinen, A., Härkönen, J., Kukkonen, J., Aarnio, P. and Koskentalo, T. (2004) Statistical model for assessing the portion of fine particulate matter transported regionally and long range to urban air, *Scand. J. Work Environ. Health*, **30**, 47–53.

Koistinen, K.J., Kousa, A., Tenhola, V., Hänninen, O.O., Jantunen, M.J., Oglesby, L., Künzli, N. and Georgoulis, L. (1999) Fine particle (PM2.5) measurement methodology, quality assurance procedures, and pilot results of the EXPOLIS study, *J. Air Waste Manage. Assoc.*, **49**, 1212–1220.

Koistinen, K.J., Hänninen, O.O., Rotko, T., Edwards, R.D., Moschandreas, D. and Jantunen, M.J. (2001) Behavioral and environmental determinants of personal exposures to PM2.5 in EXPOLIS-Helsinki, Finland, *Atmos. Environ.*, **35**, 2473–2481.

Koistinen, K., Edwards, R., Mathys, P., Ruuskanen, J., Künzli, N. and Jantunen, M. (2004) Sources of PM2.5 in personal exposures and residential indoor, outdoor and workplace microenvironments in EXPOLIS-Helsinki, Finland, *Scand. J. Work Environ. Health*, **30** (Suppl. 2), 36–46.

Kruize, H., Hänninen, O.O., Breugelmans, O., Lebret, E. and Jantunen, M. (2003) Description and demonstration of the EXPOLIS simulation model: two examples of modeling population exposure to particulate matter, *J. Expo. Anal. Environ. Epidemiol.*, **13**, 87–99.

Kulmala, M., Asmi, A. and Pirjola, L. (1999) Indoor air aerosol model: the effect of outdoor air, filtration and ventilation on indoor concentrations, *Atmos. Environ.*, **33**, 2133–2144.

Law, P.L., Lioy, P.J., Zelenka, M.P., Huber, A.H. and McCurdy, T.R. (1997) Evaluation of a probabilistic exposure model applied to carbon monoxide (pNEM/CO) using Denver personal exposure monitoring data, *J. Air Waste Manage. Assoc.*, **47**, 491–500.

Leech, J.A., Raizenne, M. and Gusdorf, J. (2004) Health in occupants of energy efficient new homes, *Indoor Air*, **14**, 169–173.

Letz, R., Ryan, P.B. and Spengler, J.D. (1984) Estimated distributions of personal exposure to respirable particles, *Environ. Monit. Assess.*, **4**, 351–359.

Lin, Z. and Deng, S. (2003) The outdoor air ventilation rate in high-rise residences employing room air conditioners, *Build. Environ.*, **38**, 1389–1399.

Mathys, P., Stern, W.B., Oglesby, L., Braun-Fahrländer, C., Ackermann-Liebrich, U., Jantunen, M.J. and Künzli, N. (2001) Elemental analysis of airborne particulate matter by ED-XRF within the European EXPOLIS Study, *ICP Inf. Newsl.*, **27**, 29–34.

Ott, W., Thomas, J., Mage, D. and Wallace, L. (1988) Validation of the simulation of human activity and pollutant exposure (Shape) model using paired days from the Denver, CO, carbon monoxide study, *Atmos. Environ.*, **22**, 2101–2113.

Özkaynak, H., Xue, J., Spengler, J., Wallace, L., Pellizzari, E. and Jenkins, P. (1996) Personal exposure to airborne particles and metals: results from the particle TEAM study in Riverside, California, *J. Expo. Anal. Environ. Epidemiol.*, **6**, 57–78.

Pasanen, P.O., Teijonsalo, J., Seppanen, O., Ruuskanen, J. and Kalliokoski, P. (1994) Increase in perceived odor emissions with loading of ventilation filters, *Indoor Air*, **4**, 106–113.

Pellizzari, E.D., Clayton, C.A., Rodes, C.E., Mason, R.E., Piper, L.L., Fort, B., Pfeifer, G. and Lynam, D. (1999) Particulate matter and manganese exposures in Toronto, Canada, *Atmos. Environ.*, **33**, 721–734.

Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K. and Thurston, G.D. (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *JAMA*, **287**, 1132–1141.

Raunemaa, T., Kulmala, M., Saari, H., Olin, M. and Kulmala, M.H. (1989) Indoor air aerosol model – transport indoors and deposition of fine and coarse particles, *Aerosol Sci. Technol.*, **11**, 11–25.

Ryan, P.B., Spengler, J.D. and Letz, R. (1986) Estimating personal exposures to NO2, *Environ. Int.*, **12**, 395–400.

Thornburg, J., Ensor, D.S., Rodes, C.E., Lawless, P.A., Sparks, L.E. and Mosley, R.B. (2001) Penetration of particles into buildings and associated physical factors. Part I: model development and computer simulations, *Aerosol Sci. Technol.*, **34**, 284–296.

Thornburg, J.W., Rodes, C.E., Lawless, P.A., Steven, C.D. and Williams, R.W. (2004) A pilot study of the influence of residential HAC duty cycle on indoor air quality, *Atmos. Environ.*, **38**, 1567–1577.

Tung, T.C., Chao, C.Y. and Burnett, J. (1999) A methodology to investigate the particulate penetration coefficient through building shell, *Atmos. Environ.*, **33**, 881–893.

Vallius, M., Lanki, T., Tiittanen, P., Koistinen, K., Ruuskanen, J. and Pekkanen, J. (2003) Source apportionment of urban ambient PM2.5 in two successive measurement campaigns in Helsinki, Finland, *Atmos. Environ.*, **37**, 615–623.

Wallace, L. (1996) Indoor particles: a review, *J. Air Waste Manage. Assoc.*, **46**, 98–126.

WHO (1999) Overview of the Environment and Health in Europe in the 1990ies. Background Document. Copenhagen, World Health Organization, Regional Office for Europe.

WHO (2000) Air Quality Guidelines for Europe. 2. European Series No. 91. Geneva, WHO Regional Publications.

Wilson, W.E., Mage, D.T. and Grant, L.D. (2000) Estimating separately personal exposure to ambient and nonambient particulate matter for epidemiology and risk assessment: why and how, *J. Air Waste Manage. Assoc.*, **50**, 1167–1183.

Wong, N.H.N. and Huang, B. (2004) Comparative study of the indoor air quality of naturally ventilated and air-conditioned bedrooms of residential buildings in Singapore, *Build. Environ.*, **39**, 1115–1123.

VII

Hänninen, O.O., Alm, S., Katsouyanni, K., Künzli, N., Maroni, M., Nieuwenhuijsen, M.J., Saarela, K., Srám, R., Zmirou, D., Jantunen, M.J., 2004. **The *EXPOLIS* Study: Implications for exposure research and environmental policy in Europe**. *Journal of Exposure Analysis and Environmental Epidemiology*, 14: 440-456.

Journal of Exposure Analysis and Environmental Epidemiology (2004) 14, 440–456

www.nature.com/jea

The *EXPOLIS* study: implications for exposure research and environmental policy in Europe

OTTO O. HÄNNINEN,[a] SARI ALM,[a] KLEA KATSOUYANNI,[b] NINO KÜNZLI,[c] MARCO MARONI,[d] MARK J. NIEUWENHUIJSEN,[e] KRISTINA SAARELA,[f] RADIM J. SRÁM,[g] DENIS ZMIROU[h] AND MATTI J. JANTUNEN[a]

[a]*KTL, POB 95, FI-70701 Kuopio, Finland*
[b]*Medical School, University of Athens, Athens, Greece*
[c]*Institute of Social and Preventive Medicine, University of Basel, Basel, Switzerland*
[d]*Institute of Occupational Health, University of Milan, Milan, Italy*
[e]*Imperial College of Science, Technology & Medicine, London, UK*
[f]*VTT, Chemical Technology, Espoo, Finland*
[g]*Inst. Exp. Medicine AS CR and Health Inst. of Central Bohemia, Prague, Czech Republic*
[h]*Medical School, University Joseph Fourier, Grenoble, France*

Exposure analysis is a crucial part of effective management of public health risks caused by pollutants and chemicals in our environment. During the last decades, more data required for exposure analysis has become available, but the need for direct population based measurements of exposures is still clear. The current work (i) describes the European *EXPOLIS* study, designed to produce this kind of exposure data for major air pollutants in Europe, and the database created to make the collected data available for researchers (ii) reviews the exposure analysis conducted and results published so far using these data and (iii) discusses the implications of the results from the point of view of research and environmental policy in Europe. Fine particle (with 37 elements and black smoke), nitrogen dioxide, volatile organic compounds (30 compounds) and carbon monoxide inhalation exposures and exposure-related questionnaire data were measured in seven European cities during 1996–2000. The *EXPOLIS* database has been used for exposure analysis of these pollutants for 4 years now and results have been published in approximately 30 peer-reviewed journal papers, demonstrating the versatility, usability and scientific value of such a data set. The multipollutant exposure data from the same subjects in the random population samples allows for analyses of the determinants, microenvironments and sources of exposures to multipollutant mixtures and associations between the different air pollutants. This information is necessary and useful for developing effective policies and control strategies for healthier environment.
Journal of Exposure Analysis and Environmental Epidemiology (2004) **14,** 440–456. doi:10.1038/sj.jea.7500342
Published online 17 March 2004

Keywords: *inhalation exposures, exposure database, exposure simulation,* $PM_{2.5}$*, particulate matter, elements, black smoke, volatile organic compounds, nitrogen dioxide, carbon monoxide.*

The goals, design and methods used in the *EXPOLIS* study

Owing to the lack of population-based information on personal exposures to air pollution in Europe, the *EXPOLIS* study (*Air Pollution Exposure Distributions within Adult Urban Populations in Europe*) was launched in 1996 as a part of the European Community (EC) Framework program IV for Research and Technological Development. Additional funding was provided by the EC for the study in Czech Republic and by national funders in all centers. The main goals of the study were: (i) to measure personal exposures of population samples of European urban populations to major air pollutants; (ii) to analyze the personal and environmental determinants of these exposures and (iii) to create a European database of these exposures and exposure-related data for exposure analysis and simulation of population exposures in the current and future scenarios (Jantunen et al., 1998).

Population samples of adult urban populations were drawn in seven selected cities or metropolitan areas, representing different city sizes and geographical locations over Europe. The study areas were Athens (Greece), Basle (Switzerland), Grenoble (France), Helsinki (Finland), Milan (Italy). Oxford (Great Britain) and Prague (Czech). The field measurements were carried out during 1996–2000 in each center over an approximately 1-year period to integrate over the seasonal variations in environmental concentrations and in population behavior and time activity. The following paragraphs give a short overview of the main features of the study design.

Population Sampling

In each of the cities, personal exposures, microenvironment concentrations and personal time activities were measured

1. Address all correspondence to: O.O. Hänninen, KTL, POB 95, FI-70 701 Kuopio, Finland.
E-mail: otto.hanninen@ktl.fi
Received 24 April 2003; accepted 26 November 2003; published online 17 March 2004

from a population sample. A *Primary population* sample was randomly drawn in each city. Based on a short mailed *Screening questionnaire*, two smaller subsamples were created. The *Exposure sample* was recruited for exposure and microenvironment monitoring, including all questionnaires. The *Diary sample* participated for time activity diary and questionnaire application without exposure or microenvironment monitoring (Figure 1). In Athens, Basle, Helsinki, Milan and Prague, the *Primary samples* were based on a random draw from the working age (25–55 years) city inhabitants (in Prague from limited *Region V* area only). In Oxford, the primary population sample was drawn from a larger ongoing epidemiological study. In Milan, the *Exposure* sample was selected from office workers and in Prague both the *Exposure* and *Diary samples* were selected from the municipality employees. In Grenoble, an ongoing study on the $PM_{2.5}$ exposures and daily symptoms of 40 volunteering asthmatics, 20–60 years of age, was adapted to yield $PM_{2.5}$ exposure results, which can be related to the data from other *EXPOLIS* centers. The response rates and the representativeness of the population samples were analyzed in detail by Rotko et al. (2000b).

Measurements

Weekday personal exposures of the *Exposure sample* and microenvironment concentrations at the subjects' home indoors, home outdoors and in workplace were monitored for 48 h. The measured air pollutants included fine particulate matter ($PM_{2.5}$), its elemental composition (37 elements in total) and black smoke (BS) concentration, carbon monoxide (CO), volatile organic compounds (VOC; 30 target compounds analyzed from all samples and approximately 250 other compounds identified when present in notable amounts) and nitrogen dioxide (NO_2; in Basle, Helsinki, Oxford and Prague). Only personal $PM_{2.5}$ exposures and their composition were monitored in Grenoble.

The air pollutants were selected based on their health effects, environmental concerns and available reliable monitoring techniques. CO originates especially from traffic and indoor sources and a monitoring technique with continuous

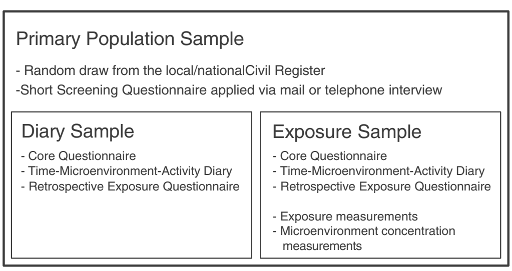

Figure 1. Diagram of the various population samples and the collected environmental data.

logging of levels in 1-min interval is available; thus CO is suitable for representing short-term variations in exposures to traffic exhausts and indoor combustion sources. Many VOC compounds are known to be carcinogenic, odorous and irritating, but also precursors for tropospheric ozone (O_3), and useful markers for various emission sources. Fine particles ($PM_{2.5}$) have the greatest current health concern, and no $PM_{2.5}$ exposure studies on representative population samples were reported in Europe so far. $PM_{2.5}$ samples also allow for analysis of their elemental composition and assessment of exposures to nonvolatile toxic elements.

Each subject carried a personal exposure monitoring (PEM) case and her/his home inside and outside and workplace were equipped with microenvironment monitors (MEM) for 48 h. The workplace concentrations were measured for the normal working hours at the actual work spot of the subject. The home inside and outside concentrations were monitored from the time when the subject would normally return from work to the time when she/he would normally leave home for work. CO was measured using electrochemical detection and continuous logging of exposures with 1-min interval. CO was not separately measured in the microenvironments. VOC were sampled on a Tenax TA absorbent (Carbotrap in Basle) using a restricted side-flow from the PM sampling line. NO_2 was measured using a passive sampling technique, producing a 48-h average concentration. Weekend exposures were not measured.

To facilitate the analysis of $PM_{2.5}$ composition and source attribution, its BS levels were determined optically and elemental composition by energy dispersive X-ray fluorescence (ED-XRF). The PM sampling and analysis techniques are described in more detail by Koistinen et al. (1999) and Hänninen et al. (2002b); the elemental analysis of the filter samples by Mathys et al. (2001), and VOC sampling by Jurvelin et al. (2001a).

Questionnaires

Four exposure questionnaires/diaries were collected from the study participants: (i) *Short Screening Questionnaire*, (ii) *Core Questionnaire*, (iii) *Time-Microenvironment-Activity-Diary* (TMAD) and (iv) *Retrospective Exposure Questionnaire*. The Short Screening Questionnaire evaluated the subjects' suitability and intention for participation before the actual field phase of the study. The other questionnaires were applied during the field campaign. The Core Questionnaire covered indoor air quality related characteristics of each subject's home and workplace, as well as commuting, socioeconomic and some exposure-related personal characteristics, such as smoking.

The TMAD defined the *EXPOLIS* microenvironments and it was used to assess the subject's time use and activities while their personal exposures and microenvironment concentrations were measured. The subject was asked to mark for each 15-min of the day the appropriate microenvironment and activity category. The 11 microenvironment categories in this TMAD were classified as "in transfer" (*walk/bike, motor cycle, car/taxi, bus/tram* and *metro/train*) and "not in transfer" (*home in* and *out, work in* and *out, other in* and *out*), and activities (*cooking, self-smoking* and *someone smoking in same room*). Multiple entries were allowed for each 15 min. The Retrospective Exposure Questionnaire was filled in at the end of the 48-h measurement period addressing specific activities, which may influence personal exposure to some compounds during the measurement period.

Quality Assurance

A quality assurance program was used to minimize any differences between the centers, affecting the comparability of the results, and specifically to ensure quantitatively reliable data. The field procedures were carefully planned, tested and documented in the pilot phase. Quality assurance methods included (i) using identical sampling equipment and (ii) questionnaires according to (iii) standard operating procedures (SOPs) in all centers, (iv) training the field researchers together in common workshops and (v) encouraging daily communication between them during the field phase including the use of cellular phones acquired partly for also this purpose. All international communication and project documentation was conducted in English.

Data Management

All the collected data was stored together with the corresponding ambient pollution and meteorological data in the local *EXPOLIS* database for further statistical analyses. The *EXPOLIS* data management procedures were developed in KTL, Finland, in collaboration with other partners. Data management objectives included the following: (i) all data items affecting the final calculated results are stored, (ii) data from all centers are stored, (iii) data storage structure is flexible, allowing later any analyses necessary, (iv) correctness of the data is maximized, (v) data entry tools and procedures are provided and (vi) privacy of study subjects is protected.

A common relational database structure called *EXPOLIS* Access Database (EADB) was developed using Microsoft (Seattle, WA, USA) Access 7.0 (also known as version -95). Relational database model was selected especially to allow maximum flexibility for data processing. Microsoft Access was selected because of its visual development and end user friendly environment, low software cost and easy availability as part of the most common office software package.

A local database was created for each center. The local database consisted of Access database files for storing data from local Civil Register and other national registers, *EXPOLIS* time-activity diaries, questionnaires, monitors, laboratory analyses, calibration procedures and environmental conditions as well as data from urban air quality and meteorological measurements covering the field study periods. All data was stored in its primary form and all calculations were performed using the primary data dynamically.

The local data was grouped to be stored in separate database files (Table 1). Population sample management, questionnaire data and concentration sampling were stored into the local main database. Time-activity diaries were stored in a 15-min time series database, CO data in 1-min time series database, meteorological data in 3-h resolution time series database and ambient air quality fixed station data in 1/24-h database. The averages of environmental variables from the meteorological and fixed station databases were calculated into the *Fixedruns* database for periods corresponding the microenvironment and personal sampling.

To facilitate updating of the data processing algorithms without changes to the data files, the local database files were split into two functional groups. (i) *Data files* contained all data tables but no queries, forms or Visual basic modules; (ii) these data processing tool elements were stored in *Tool files*. The tool databases were then linked to the data files using Access Linked Table Manager.

A data integrity protocol was established according to the data security requirements of EU Directive on Protection of

Table 1. Database files in the local and the combined international databases

Local database files in each center		
Data files	Tool file	Description
HELSINKI.MDB	EADBTOOL.MDB	Main local database: questionnaires, exposures, concentrations, etc.
TMAD15 min.MDB	TMAD15 minTOOL.MDB	Time-activity diaries (15-min resolution) 15-min avg. personal CO data
CO1 min.MDB	CC1 minTOOL.MDB	1-min CO exposures and TMAD data
FIXED.MDB	AmbientTOOL.MDB	Hourly ambient air quality data
MET.MDB		3-hourly meteorological data
FIXEDRUNS.MDB		*EXPOLIS* sample sampling period averages of ambient and met data: all stations
Combined International database files		
CIDB_Sep02_Access95.mdb		Access 95 (version 7) format
CIDB_Sep02_Access97.mdb		Access 97 (version 8) format

Individuals with Regard to Processing Personal Data in Medical and Epidemiological Research. According to this protocol, persons were identified using codes, which cannot be translated back to identity. The database files were secured with user identification and password control and the staff working with the databases in all centers were trained in several common workshops.

After the field phase and local data cleaning in each center, the local databases were collected in KTL and the main results were then collected from the local databases and put into the Combined International *EXPOLIS* Database (CIDB) (Figure 2). Unlike the local databases, CIDB contains only the tables of calculated concentrations. The local databases were run solely in Access-95 environment. The CIDB was created in Access-95 format, but then converted also to Access-97 format. While the local databases require user identification and are password protected, the CIDB is not, allowing easy access to the data. All data allowing identification of subjects have been removed from the CIDB.

The first version of the international database was compiled in late 1999 (v. November 1999) and delivered to the *EXPOLIS* centers in CD-ROM format. The particulate matter data was updated and the database version two released in 2000 (December 2000). The elemental data was updated again in summer 2002 and the database version three was released in September 2002 (September 2002). The complete set of *EXPOLIS* databases contains database documentation (Hänninen et al., 2002a, available also from http://www.ktl.fi/expolis/bb.html), the CIDB database in the two versions as well as copies of all local databases in one CD-ROM disk. Sizes of the data files on the disk are listed in Table 2.

Review of the published results of the *EXPOLIS* study

Using a variety of statistical methods, the *EXPOLIS* database has been used to analyze (i) the statistical associations between exposures to different air pollutants, (ii) the contributions of different air pollution sources to air pollution exposures and (iii) the relationships of geographic,

Table 2. *EXPOLIS* database sizes, including the local databases

Component	Number of files	Number of folders	Uncompressed size (MB)
CIDB	2		73
\Documents	15	1	6.4
\Additional data	6	1	3.6
\Expolis www	212	10	6.3
\Local databases			
\Athens	10	1	35
\Basle	10	1	122
\Grenoble	5	1	13
\Helsinki	10	1	168
\Milan	10	1	68
\Oxford	10	1	83
\Prague	10	1	39
Complete CD-ROM	301	19	614

Compressed size of the whole CD-ROM is 233 MB (compression ratio 64%).

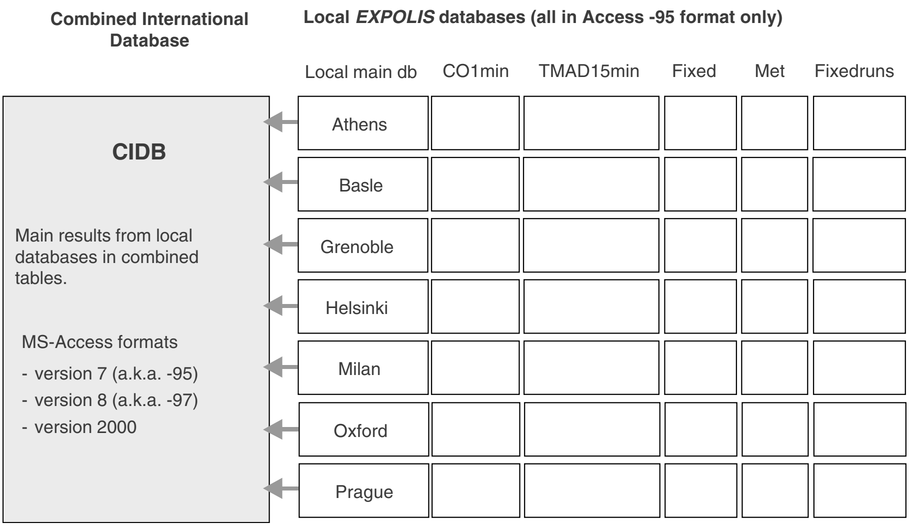

Figure 2. Local databases and the Combined International *EXPOLIS* database (CIDB).

housing, occupation and commuting related, behavioral and socioeconomic factors to air pollution exposures. Further, using the *EXPOLIS* database, probabilistic simulation models have been developed to assess the population exposure distributions for specific subpopulations, specific urban areas and selected future scenarios. These analysis have been published, besides numerous conference abstracts, doctoral theses and other publications, in approximately 30 papers in peer-reviewed scientific journals. These papers are shortly reviewed below first for each pollutant and then for the nonpollutant specific topics. Each analysis combines various selected data from questionnaires and personal, microenvironment and ambient measurements. Some of the papers look at differences between the *EXPOLIS* cities, others perform more detailed analyses within a single city. The reviewed papers and *EXPOLIS* data subsets used are summarized in chronological order of publication in Table 3. Pollutant specific data from each city used in the publications are summarized in Table 4.

$PM_{2.5}$

Boudet et al. (1998) analyzed the roles of ambient air and time spent in traffic for personal $PM_{2.5}$ exposures measured in Grenoble. A total of 40 adult asthmatic volunteers carried a personal exposure monitor (PEM) case for 48 h. The Grenoble study deviated from the *EXPOLIS* design by using two PEM pumps and no microenvironment monitoring. One of the PEM pumps was used to collect average 48-h samples, while the subject manually stopped the other one whenever he/she went to outdoors (including traffic). Each subject

Table 3. Summary published data analyses performed using the *EXPOLIS* database

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	Design & QA/QC					Data analysis																								
	Jantunen et al., 1998	Koistinen et al., 1999	Rotko et al., 2000a	Jurvelin et al., 2001a	Hänninen et al., 2002	Boudet et al., 1998	Oglesby et al., 2000a	Oglesby et al., 2000b	Rotko et al., 2000b	Boudet et al., 2001	Edwards and Jantunen 2001	Koistinen et al., 2001	Kousa et al., 2001	Rotko et al., 2001	Edwards et al., 2001a	Edwards et al., 2001b	Jurvelin et al., 2001b	Georgoulis et al., 2002	Götschi et al., 2002	Kousa et al., 2002	Rotko et al., 2002	Hänninen et al., 2003	Hänninen et al., 2004	Koistinen et al., 2004	Kruize et al., 2003	Bruinen de Bruin et al., 2004a	Bruinen de Bruin et al., 2004b	Lai et al., 2004		
Civil register	D		A			A	A		A								A													
Questionnaires																														
Short screening Q	D		A														A													
Core Q	D		A				A		A	A		A	A	A	A	A	A		A		A	M	M			A				
TMAD	D		A			A					A	A	A	A			A	A				M	M			A		A		
Retrospective exposure Q	D		A									A	A	A				A	A							A				
Concentrations																														
Exposure data	D			A				A	A	A	A	A	A	A	A	A	A	A		A	A	M	M	A	M	A	M	A		
Microenvironment concentrations	D			A				A			A	A	A		A	A	A		A	A	A	M	M	A	M	A	M	A		
Ambient air quality	D					A		A		A		A	A	A				A		A	A		M			A				
Meteorology	D				A											A								A						
Pollutants																														
$PM_{2.5}$	D	A			A	A		A	A	A		A							A	A	A	M	M	A	M			A		
$PM_{2.5}$ elements								A						M										A				A		
$PM_{2.5}$ black smoke																			A											
Target VOC (30 compounds)	D			A							A				A	A												A		
Other VOC																	A													
NO_2	D					A								A							A							A		
CO	D																	A						A			M	A		
Cities																														
Athens	D	A	A	A														A	A	A	A				M					
Basle	D	A	A	A			A	A					A					A	A	A	A				M					
Grenoble	D		A			A				A																				
Helsinki	D	A	A	A	A		A		A		A	A	A	A	A	A	A	A	A	A	A	M	M	A	M					
Milan	D		A	A														A			A					A	M			
Oxford																					AA									
Prague	D		A	A									A					A	A	A	A				M					

D = description, A= analysis, M= modeling

Table 4. Summary of published data analyses for each pollutant in each city

	Athens	Basle	Grenoble	Helsinki	Milan	Oxford	Prague	Total
$PM_{2.5}$	2, 19, 20, 21, 25	2, 8, 19, 20, 21, 25	6, 10	2, 5, 9, 12, 19, 20, 21, 22, 23, 24, 25	19, 20, 21	21, 28	19, 20, 21, 25	15
$PM_{2.5}$ elements		8		23, 24		28		4
$PM_{2.5}$ BS	19	19		19	19		19	1
Target VOC	4	4	N/A	4, 11, 15, 16	4	28	4	5
Other VOC			N/A	17				1
NO_2	21	13, 21	N/A	13, 14, 21	21	21, 28	13, 21	4
CO	18	18	N/A	18	18, 26, 27	28	18	4
Total	7	10	2	19	7	2	7	-

N/A = data not available.
The listed numbers refer to references in Table 3, except in the total column and row where they are counts of papers.

completed the 15-min resolution *EXPOLIS* time-activity diary, indicating times spent outdoors and in various means of transportation. The results showed that 33% of $PM_{2.5}$ mass exposures occurred while outdoors or in traffic. The average exposure levels were 7.3 $\mu g/m^3$ indoors and 29 $\mu g/m^3$ outdoors and in traffic. The average time spent outdoors or in traffic was 11%.

The relationship of ambient PM_{10} levels and personal $PM_{2.5}$ exposures in Grenoble was analyzed by Boudet et al. (2001). PM_{10} levels were available from one urban background station and one traffic-oriented station. According to the geographical home and workplace locations and the traffic density in the corresponding nearby streets, six proximity models were created for the difference of ambient PM_{10} and personal $PM_{2.5}$ levels.

Götschi et al. (2002) analyzed the residential indoor and outdoor $PM_{2.5}$ and BS levels in four *EXPOLIS* cities, Athens, Basle, Helsinki and Prague. $PM_{2.5}$, and BS levels were lowest in Helsinki, moderate in Basle and remarkably higher in Athens and Prague. In each city, Spearman rank correlation coefficients of indoors *versus* outdoors were higher for BS than for $PM_{2.5}$.

In a linear regression model (data from all cities), outdoor BS levels explained clearly more of indoor variation (86%) than in the corresponding $PM_{2.5}$ model (59%). The authors conclude that BS captures the traffic-, especially diesel, related elemental carbon particles better than $PM_{2.5}$ measurement and thus can be used as a cheap additional analysis method to assess concentration of particles of traffic origin.

Koistinen et al. (2001) used statistical methods to analyze $PM_{2.5}$ exposures and exposure determinants in *EXPOLIS*—Helsinki. The most important single factor affecting exposures was found to be exposure to environmental tobacco smoke (ETS); mean exposure level of ETS-exposed subjects was almost double compared to those not (16.6 vs. 9.6 $\mu g/m^3$). The mean exposure level of active smokers (exposure to smoking assessed only as ETS) was 31 $\mu g/m^3$.

The mean residential indoor concentrations of non-ETS-exposed subjects were lower than those outdoors (levels were 8.2 and 9.5 $\mu g/m^3$, respectively). In simple linear regression models, residential indoor concentrations were the best predictors of personal exposure concentrations, even though the residential concentrations were measured in the *EXPOLIS* protocol only during the leisure time. Coefficients of determination (R^2) between personal exposures of all participants and residential indoor, workplace indoor, residential outdoor and ambient concentrations were 0.53, 0.38, 0.17 and 0.16, respectively.

Multiple regression, using residential indoor and workplace concentrations and traffic density in the nearest street from home as independent variables, explained 77% of the exposure variance of non-ETS-exposed subjects. Stepwise regression without residential and workplace indoor concentrations explained 47% of the exposure variance using ambient concentration and home location as predictors of personal exposure. Wilcoxon and ANOVA tests identified the time of windows kept open during the 48-h measurement period and home location (classified as downtown/suburban high-rise/suburban single houses) as statistically significant determinants of personal exposures. Time spent in traffic, home to work distance, cooking or stove type used at home were not statistically significantly associated to the exposure levels.

Rotko et al. (2000a) used similar techniques to analyze relationships between *EXPOLIS*-Helsinki $PM_{2.5}$ exposures and sociodemographic factors. Variation in personal exposures between sociodemographic subgroups was best described by differences in occupational status, education level and age. Lower occupational status, less-educated and younger participants have greater exposures than upper occupational status, more-educated and older participants. Differences in workplace concentrations explained most of the occupational socioeconomic differences. Differences in personal day and night exposures and residential indoor concentrations explained the exposure differences between age groups. Men had higher exposure levels and higher differences between sociodemographic groups than women. No gender, socioeconomic or age differences were observed in residential outdoor concentrations between groups.

In a following work, Koistinen et al. (2004) used principal component analysis (PCA) and $PM_{2.5}$ elemental analysis data to identify PM sources from *EXPOLIS*-Helsinki concentrations. Then, mass reconstruction techniques was used to quantify source contributions for residential indoor and outdoor and workplace indoor concentrations and personal 48-h exposures. Inorganic secondary particles contributed 31 (personal)–46 (outdoors) percents to the $PM_{2.5}$ levels. Second highest were primary particles (28–35%) followed by soil particles (16–27%). Besides these source categories, also sea/road salt and detergent sources were identified. Resuspension of soil particles in indoor environments was found to be notable. As the authors concluded, the use of fixed site data in epidemiological studies might lead to underestimation of true exposure–response relationship and respective health effects. Kousa et al. (2002) analyzed the exposure chain of $PM_{2.5}$ levels from ambient levels to residential outdoor, indoor and workplace concentrations and personal exposures in Athens, Basle, Helsinki and Prague. Ambient $PM_{2.5}$ data was available only from Helsinki, where ambient levels correlated quite well with residential outdoor concentrations ($r = 0.90$). Highest correlations were found between leisure time personal exposures and residential indoor concentrations. When ETS exposures were excluded, these correlation coefficients varied between 0.72 (Prague) and 0.92 (Basle) between the cities.

Linear regression model built using log-transformed non-ETS residential indoor concentrations from all cities predicted 76% of variation in personal leisure time exposures. Similar model predicted 66% of the day time exposure variation with workplace indoor concentration. Leisure time and workday exposures correlated with each other quite poorly. In the absence of ETS and other significant indoor or personal sources and for nonworking, noncommuting subjects, ambient fixed station levels explained approximately 50% of personal exposure variation.

Oglesby et al. (2000a) used *EXPOLIS*-Basle $PM_{2.5}$ levels and elemental data to study validity of ambient PM concentrations as surrogates for personal exposures of ambient origin from different sources. Elemental data was used to estimate $PM_{2.5}$ fraction from long-range transport, traffic and crustal origin. Personal $PM_{2.5}$ mass exposures were not correlated to corresponding residential outdoor levels (rank correlation 0.07). Long-range fractions of residential outdoor concentrations correlated much better with corresponding personal exposure fractions (rank correlation 0.85) than the traffic and crustal fractions (varying from element to element from 0.12 to 0.53). The finding was consistent with the spatially homogeneous regional pollution and higher spatial variability of traffic and crustal indicators. Thus, the authors conclude that for some source-specific exposures, ambient fixed site data is not the optimal measure.

VOC

Edwards and Jantunen (2001) and Edwards et al. (2001a, b) analyzed the Helsinki VOC data from several aspects. Edwards and Jantunen (2001) focused on benzene exposures only. Observed median levels were for the personal exposures 2.5 $\mu g/m^3$ for nonsmokers, 2.9 $\mu g/m^3$ for ETS-exposed subjects and 3.1 $\mu g/m^3$ for active smokers. Residential indoor levels were 3.1 and 1.9 $\mu g/m^3$ for environments with and without tobacco smoke, respectively. Residential outdoor level was 1.51 $\mu g/m^3$ and workplace concentrations were 3.6 and 2.1 $\mu g/m^3$ (with and without tobacco smoke, respectively).

Multiple stepwise regression identified indoor benzene concentrations as the strongest predictor for personal benzene exposures of those not exposed to tobacco smoke, followed sequentially by time spent in a car, time in the indoor environment, indoor workplace concentrations and time in the home workshop. Relationships between indoor and outdoor microenvironment concentrations and personal exposures showed considerable variation between seasons. Automobile use-related activities were significantly associated with elevated benzene levels in personal and indoor measurements when tobacco smoke was not present.

Edwards et al. (2001b) looked at the 30 measured target VOC compounds measured in *EXPOLIS*-Helsinki. Residential indoor levels were found to be higher than outdoor levels for all other compounds but hexane. Personal exposure levels were lower and workplace indoor concentrations even still lower for compounds that had strong residential indoor sources (D-limonene, alpha pinene, 3-carene, hexanal, 2-methyl-1-propanol and 1-butanol).

ETS-exposed participants had significantly higher personal exposures to benzene, toluene, styrene, *m*, *p*-xylene, *o*-xylene, ethylbenzene and trimethylbenzene. ETS-free workplace concentrations were higher than ETS-free personal exposure concentrations for styrene, hexane and cyclohexane. Personal exposures of participants not exposed to ETS were approximately equivalent to time-weighted ETS-free indoor and workplace concentrations, except for octanal and compounds associated with traffc, which showed higher personal exposure concentrations than any microenvironment (*o*-xylene, ethylbenzene, benzene, undecane, nonane, decane, *m*, *p*-xylene, and trimethylbenzene). The observed concentration levels varied from below 1 $\mu g/m^3$ to few hundreds or few thousands of $\mu g/m^3$. Highest single levels were observed for *m*, *p*-xylene, 2-butoxyethanol and cyclohexane (2779, 2421 and 1512 $\mu g/m^3$, respectively).

In their follow-up work, Edwards et al. (2001a) used principal component analysis to identify VOC sources from the Helsinki microenvironment concentration and personal exposure data. Variability in VOC concentrations in residential outdoor microenvironments was dominated by compounds associated with long-range transport of pollutants, followed by traffic emissions, emissions from trees and household product emissions. Variability in VOC concentrations in ETS-free residential indoor environments was dominated by compounds associated with indoor cleaning products, followed by compounds associated with traffic emissions, long-range transport of pollutants and household product emissions. The median indoor/outdoor ratios for compounds typically associated with traffic emissions and long-range transport of pollutants exceeded 1, in some cases quite considerably, indicating substantial indoor source contributions.

Jurvelin et al. (2001b) analyzed a carbonyl data set collected in *EXPOLIS*-Helsinki besides the standard *EXPOLIS* measurements. Using Sep-Pak DNPH-Silica cartridges, formaldehyde and acetaldehyde exposures and concentrations in the standard *EXPOLIS* microenvironments were measured for 15 subjects. Observed mean personal exposure levels were 21.4 ppb for formaldehyde and 7.9 ppb for acetaldehyde. Personal exposures were systematically lower than residential indoor concentrations for both compounds, and ambient air concentrations were lower than both residential indoor concentrations and personal exposure levels. The mean workplace concentrations of both compounds were lower than mean residential indoor concentrations. This indicated that residential indoor concentrations are better estimates of personal exposures to these compounds than the ambient concentrations.

In their follow-up work, Jurvelin et al. (2003) looked at all the 16 carbonyl compounds measured in this substudy. Findings for the remaining 14 compounds were similar than those presented in the previous paper for formaldehyde and acetaldehyde; residential indoor concentrations were higher than personal exposures and other microenvironment concentrations, thus driving the personal exposures.

CO

The CO exposures have been analyzed by Georgoulis et al. (2002) in five of the *EXPOLIS* cities and Bruinen de Bruin et al. (2004a) in detail in Milan. Georgoulis et al. (2002) used two different approaches in the statistical analyses. First, the determinants of log-transformed average 48-h exposures were analyzed using analysis of variance (ANOVA) and multiple linear regression techniques. Secondly, the CO personal exposure corresponding to the specific 15-min time periods when different activities (i.e. "*in transfer*", "*under ETS exposure*", "*during use of gas appliances*") were taking place was calculated and compared using nonparametric tests.

The geometric mean 48-h exposure levels of nonsmoking subjects were 1.68, 0.82, 0.45, 2.17 and 1.50 mg/m^3 in Athens, Basle, Helsinki, Milan and Prague, respectively. Levels for smokers were slightly or substantially higher in all cities but Helsinki. Proportion of smokers was restricted in the population sampling process in all but Helsinki and Milan. The Spearman rank correlation coefficients between ambient and personal 48-h levels varied from 0.33 (Helsinki) to 0.77 (Milan).

The coefficient of determination (adjusted R^2) in regression models, using the log-transformed 48-h personal exposure as the dependent variable and the independent variables were the ambient concentrations, ETS exposure, exposure to gas appliances and the time spent in traffic, varied from 0.08 (Helsinki) to 0.59 (Milan).

The analysis of short-term (15-min) exposure levels showed that in all cities the time spent in traffic corresponded to the highest personal exposure events. Exposure during time spent outdoors was second in Athens, Helsinki and Milan, but not in Basle and Prague. Time spent indoors resulted on average in the lowest exposure.

Bruinen de Bruin et al. (2004a) studied the CO measurements conducted in Milan in more detail focusing on the contribution of indoor sources on the microenvironment concentration and personal exposure levels. Bruinen be Bruin also calculated running 1- and 8-h average exposures and compared the personal running maxima to corresponding ambient levels. For the 1-h running average, the personal exposures were found to be higher than the ambient levels, indicating that short-term exposure peaks cannot be seen in ambient data.

Bruinen de Bruin et al. (2004a) also calculated proportional contributions of microenvironments to the 48-h personal exposures. It was found that exposures in indoor environments contributed approximately 82% of the total CO exposures. While only 7.5% of time was spent in traffic, the contribution to 48-h exposures was clearly higher, 16%, indicating higher CO levels in traffic. Both ETS and gas cooking were statistically significantly connected to microenvironment concentrations ($P < 0.05$). Multiple linear regression models, using ambient levels and the presence of indoor sources (ETS and gas cooking, yes/no) as independent variables, explained 49, 36 and 89% (adjusted R^2) of the microenvironment concentrations in home indoor, work indoor and other indoor, respectively.

NO_2

Kousa et al. (2001) analyzed NO_2 levels and exposures in Basle, Helsinki and Prague. The mean residential indoor and outdoor levels were lowest in Helsinki (24 and 18 μg/m^3, respectively), intermediate in Basle (36 and 27 μg/m^3) and highest in Prague (61 and 43 μg/m^3). Workplace levels were highest in Basle, followed by Prague and Helsinki (36, 30 and 27 μg/m^3, respectively). Time-weighted average microenvironment exposure model predicted 74% of the personal exposure variation. Regression models using log-transformed residential outdoor and ambient levels and home and workplace characteristics (work location, use of gas appliances and keeping windows open) as predictors explained 48 and 37% of the personal exposure variation. Regression model using only ambient monitoring explained 19% of exposure variation for all centers and only 11% in Helsinki with the largest data set.

Rotko et al. (2001) studied microenvironment, behavioral and sociodemographic factors in relationship to personal NO_2 exposures in Helsinki. Differences in exposures were analyzed by comparing subpopulations created by grouping exposures according to behavioral, socioeconomic and demographic factors. Factors associated with statistically significant differences between the population groups were work and residence location, housing characteristics, traffic volume near residence, season and keeping windows open. Exposure to ETS and use of gas stove were also associated with elevated NO_2 exposures, although the latter were rare in Helsinki. Increased education associated with decreasing exposures. Employed men had lower exposures on the average than unemployed men, but otherwise the occupational status did not link to exposure levels.

Participation Bias

Oglesby et al. (2000b) analyzed the characteristics of the *EXPOLIS* participants in Basle and in Helsinki for participation bias. Participants of intensive exposure monitoring (exposure sample) were compared to subjects that completed only the questionnaire study (diary sample). The comparison was based on home locations and traffic densities on the nearby street. In Basle exposure study, participants were more likely to live along streets with low traffic volume. Adjusted for sex, age and nationality, an increase of 100 cars per hour was associated with 14% decrease in participation. In Helsinki, the corresponding finding was qualitatively similar but not statistically significant.

Air Pollution Annoyance

Rotko et al. (2002) compared the perceived air quality to measured $PM_{2.5}$ and NO_2 exposures in six cities (Athens, Basle, Helsinki, Milan, Oxford and Prague). The measured microenvironment concentrations and personal exposures were compared to the annoyance levels reported in the questionnaires for home, work and traffic.

A considerable proportion of the adults surveyed was annoyed by air pollution. Female gender, self-reported respiratory symptoms, downtown living and self-reported sensitivity to air pollution were directly associated with high air pollution annoyance score for exposure in traffic, but association for smoking status, age or education level were statistically significant. Population level annoyance averages correlated with the city average exposure levels of $PM_{2.5}$ and NO_2. A high correlation was observed between the personal 48-h $PM_{2.5}$ exposure and perceived annoyance at home as well as between the mean annoyance at work and both the average work indoor $PM_{2.5}$ and the personal work time $PM_{2.5}$ exposure. With the other determinants (gender, city code, home location) and home outdoor levels, the model explained 14% ($PM_{2.5}$) and 19% (NO_2) of the variation in perceived air pollution annoyance in traffic. Compared to Helsinki, in Basle and Prague the adult participants were more annoyed by air pollution while in traffic.

Reporting to Study Participants

Helm et al. (2000) compared reporting procedures used in the German Environmental Survey (Seifert et al., 2000) and the Helsinki part of the *EXPOLIS* study. Both independently reported personal results in a similar fashion. Apparently, a lot of thought and planning was found to be necessary to produce reports containing a quantity and depth of expert information that is easily and correctly understood by laymen.

Simulation of Population Exposures

Development of probabilistic simulation modeling technique was one of the original goals of the *EXPOLIS* study (Jantunen et al., 1998). The Dutch Institute for Public Health and the Environment (RIVM) coordinated the development of worksheet-based framework for building population exposure simulation models. The framework development is described by Kruize et al. (2003). Kruize et al. demonstrate the modeling environment by presenting two examples.

The first example was built around the *EXPOLIS* database; $PM_{2.5}$ exposures are simulated for Athens, Basle, Helsinki and Prague using a simple microenvironment approach. The model inputs were formed directly from the *EXPOLIS* measurements in residential indoor, outdoors and workplace. Time-activity distributions were created from the *EXPOLIS* time-microenvironment-activity diaries; the *EXPOLIS* population was not divided into any subpopulations. The model outputs roughly predicted the mean exposures in each city and the simulated mean exposures ranked into same order as the observed exposures. The differences between model output and observed distribution were bigger for standard deviation (SD) estimates.

The second example modeled PM_{10} exposures of the whole Dutch population. Rural and urban populations were modeled separately and each of them was divided into four age/occupation categories. Indoor concentrations are modeled using constant effective penetration factor. ETS exposures were modeled as an indoor source in each microenvironment. Model outputs were presented for the current situation (including ETS exposures) and for the hypothetical scenario where the ETS exposures are excluded. The results showed that ETS exposures contribute remarkably to the population exposure distribution as was to be expected. The exposures of the highest quartile were approximately doubled with ETS.

Hänninen et al. (2003) used the *EXPOLIS* simulation framework to perform detailed validation and model component evaluation tests using Helsinki $PM_{2.5}$ data. Four models were built using different approaches to microenvironment and subpopulation definitions. All models were based on microenvironment concentration distributions observed in the *EXPOLIS* study. The two simple models did not exclude ETS affected environments from the data and modeled the target population time activity with single beta distributions for each microenvironment. The two more detailed models excluded ETS cases from the data and modeled time activities of employed and nonemployed subpopulations separately. The model outputs were compared to corresponding observed distributions both graphically and numerically. All models compared reasonably with the validation data, but the two detailed models were clearly closer to the observed values especially in the higher percentiles of the population distribution. The population averages were quite close to observed values (e.g. for model 4 both levels were $9.2\,\mu g/m^3$), but the SDs were slightly underestimated by the simpler models (23–35%). The results showed, as expected, that the microenvironment modeling approach accurately predicted the exposures when the true microenvironment concentration distributions were known.

In their follow up work, Hänninen et al. (2004) developed the microenvironment-based modeling approach further, using the effective penetration factor approach to model indoor concentrations from ambient $PM_{2.5}$ concentrations in Helsinki. Three different approaches to model ambient concentrations were tested. Penetration factors were analyzed from the *EXPOLIS* elemental data using sulfur as a marker for particle fraction of ambient origin. Fourth model was calculated including ETS indoor source. The non-ETS models predicted the mean population exposure level within 5–6% of the observed value; the ETS-included model underestimated the mean by 15%. All models underestimated the highest percentiles slightly and thus the modeled SDs were approximately 30% lower than the observed values.

Bruinen de Bruin et al. (2004b) used the simple microenvironment model approach demonstrated by Kruize et al. (2003) and validated by Hänninen et al. (2003) for $PM_{2.5}$ to model CO exposures in Milan. Bruinen de Bruin et al. (2004b) tested different levels of grouping of the *EXPOLIS* time-activity diary categories to simulate CO exposures with different averaging times. The most detailed microenvironment model used the original 11 diary microenvironments directly. The second approach grouped traffic microenvironments and all stationary outdoor microenvironments together producing five microenvironments (home, work, other indoors, outdoors and traffic). The simplest microenvironment model used only home and workplace microenvironments. Each of these three microenvironment models were run for running maximum exposures for 24-, 8- and 1-h averaging times.

All models predicted the mean population exposure within $\pm 11\%$ of the observed value; results for the 24-h averaging time were closest for the 5- and 11-microenvironment models. All models underestimated the population SD by 3–25%. The results demonstrated that the modeling approach can be used even for averaging times below 24 h and that the model is not very sensitive to the number of microenvironments included even in the case of CO, which has much steeper within-city concentration gradients than $PM_{2.5}$.

Oxford Results

Lai et al. (2004) reported a summary of results for all pollutants measured in Oxford, UK. They found that the exposure levels were in general higher than those observed in *EXPOLIS*-Helsinki, but lower than those in the all other *EXPOLIS* cities. They looked also at the correlations for exposure levels of different pollutants; the only statistically significant correlation was found between TVOC and $PM_{2.5}$ (for log-transformed data $r = 0.41$, $P < 0.05$). They concluded that various pollutants cannot be used as markers for each other.

Implications for policy and research in Europe

Numerous epidemiological studies have connected air pollutant levels to adverse health effects, including premature mortality (e.g. Laden et al., 2000; Pope et al., 2002). While there has been much controversy about the accuracy of the exposure estimates used in such studies, and thus in the actual value of the dose–response factors, there is no doubt about the finding itself: current urban levels of air pollution have a statistical connection with complications of health.

It is possible to observe this connection to public health using merely proxies of true exposures—such as ambient air pollution levels measured at fixed monitoring stations within the urban areas under study—but any health effect caused by air pollution to a specific individual at a specific time must be caused by the actual personal exposure of this individual.

Development of science-based policies for promotion of public health requires careful analysis of exposures within the population, including emission sources, exposure routes, behavioral determinants and population groups at risk. When information about these critical factors accumulates, also more specific dose–response factors for various pollutants are needed. The strength of the epidemiological dose–response factors is in the fact that they represent real population in an existing exposure scenario, but they often lack information on the differences between population groups and specificity to causal agents. Since exposures to a specific pollutant vary from subpopulation to another, and various policy options affect these exposures with largely different efficacies, exposure and risk analysis should be carried out in population group level. In the case of particulate matter, the pollutant itself consists of different fractions, with presumably different toxicities, and thus in this case also the dose–response factor should be determined for each of these fractions to allow efficient policy optimizations.

Value of Exposure Databases

Databases with representative data on exposure factors on population level as well as data on actual exposures are necessary to analyze population exposures for efficient management. The published data analyses based on the *EXPOLIS* database show that the data collection and storage were successful, and that the database has made it possible to extract and combine these data in many different useful ways. The data analysis of the *EXPOLIS* data is by no means complete. This can been seen clearly from the summaries in Tables 3 and 4, but also the analyses already performed are not comprehensive and leave room for more detailed or focused analyses of the same data even in scientific sense. Exposure databases can be valuable administrative tools in exposure management even after they have been scientifically completely exploited, as shown earlier by THERdbASE (Hern et al., 1997).

Documentation of complex data systems is always a challenge. Scientists and analysts need information on the data structures, units of measure and many other things that are not necessarily self-evident from the data itself. Thus, there is a clear need to produce and make available project documents containing this information, allowing the scientists to use the existing and available data to its maximum. Documentation of the *EXPOLIS* databases, including description of data entry, quality control and processing tools, is available in the Internet (Hänninen et al., 2002a, http://www.ktl.fi/expolis/bb.html).

Indoor Air Pollution

For some pollutants, the personal exposures are driven mainly by pollution of ambient origin. For these, controlling of ambient sources and ambient concentrations is the most effective approach to protect public health. For many others, the indoor pollutant levels are higher than the ambient levels and modify the personal exposures to such extent that it is not possible to protect public health by only looking at ambient environment. These pollutants, like many VOCs, carbon monoxide or fine particles, have significant indoor sources, which raise the concentrations in closed compartments with limited air ventilation to high levels even when the pollutant emission indoors would be small compared to ambient emissions.

The European Commission is currently starting to develop methodologies to control indoor exposures to chemical compounds. Analysis of the total exposures is required to select the pollutants reasonably and new way of thinking is needed to develop means for controlling them. European industry has also expressed increasing concern for conducting high-quality scientific exposure analysis, as it is in their interests to avoid health relevant exposures that may lead to expensive interventions. An example of such research funded by the European industry is the *EXPOLIS-INDEX* project, where the *EXPOLIS* time activity, indoor concentration and personal exposure data are used to analyze exposure determinants especially for the indoor environments.

ETS

ETS is known to be an important source for a large number of pollutants. Approximately half of the gross population exposure to fine particles can be attributed to ETS—and this is not even taking into account direct inhalation of tobacco smoke by active smokers. Technically speaking, by far the most efficient way to reduce population exposures to $PM_{2.5}$—and to many other air pollutants as well—would be to abandon smoking. Good results have been achieved by controlling exposures to ETS in North European countries by setting restrictions to smoking in public spaces and workplaces. Problems in exposures of special groups, like restaurant workers or children of smoking parents, still mostly remain.

Exposure Levels Against Current Guidelines

European Commission (EC) has set limit values for pollutant concentrations in the ambient air based on the Framework Directive 1996/62/EC (http://europa.eu.int/comm/environment/air/ambient.htm). Actual limit values have been set for NO_2, SO_2, Pb and PM_{10} in Daughter Directive 1999/30/EC and limit values for CO and benzene in Daughter Directive 2000/69/EC. Limit values for arsenic (As), cadmium (Cd), mercury (Hg), nickel (Ni) and polycyclic aromatic hydrocarbons (PAH) have been suggested, but they are not set yet. WHO (2000) has set guidelines for slightly larger number of pollutants (Table 5). In overlapping cases, the values are identical or close to each other in most cases.

For NO_2, *EXPOLIS* data is available from four cities (Basle, Helsinki, Oxford and Prague). Population average of personal exposures to NO_2 exceeds the annual EC limit value ($40\,\mu g/m^3$) in Prague (Figure 3). The hourly limit value ($200\,\mu g/m^3$) is not exceeded by maximum 48-h exposure in any of the four cities with NO_2 data, but in Basle the observed maximum is very close ($184\,\mu g/m^3$), indicating a high probability of hourly limit value exceedance, and also in Oxford and Prague it is more than half of the limit value—levels which must be considered high when the much longer averaging time in the measurement is taken into account.

Table 5. List of pollutants with EU limit or WHO guideline values and corresponding personal exposures observed in the *EXPOLIS* study

Pollutant	Unit	WHO guidelines (WHO, 2000)								EU limit values				EXPOLIS personal exposures							
		Averaging time								Averaging time					Athens	Basle	Grenoble	Helsinki	Milan	Oxford	Prague
		10-min	15-min	30-min	1-h	8-h	24-h	Week	Annual	1-h	8-h	24-h	Annual	Hours	avg (max)	avg (max)	avg (max)	avg (max)	avg (max)	avg (max)	avg (max)
Organic																					
Benzene	$\mu g/m^3$												5	48	18 (217)	5 (49)		3 (69)		5 (28)	10 (38)
1,2-Dichloroethane ($C_2H_4Cl_2$)	mg/m^3						0.7														
Dichloromethane (CH_2Cl_2)	mg/m^3						3	0.45													
Formaldehyde (HCHO)	mg/m^3			0.1																	
Styrene	mg/m^3							0.36						48	0.005 (0.07)	0.001 (0.01)		0.001 (0.02)		0.002 (0.02)	0.003 (0.01)
Tetrachloroethylene	mg/m^3								0.25												
Toluene	mg/m^3							0.26						48	0.13 (1.33)	0.03 (0.19)		0.03 (0.49)		0.03 (0.32)	0.08 (0.45)
Inorganic																					
Carbon monoxide (CO)	mg/m^3		100		60	30	10				10			1	15 (118)	5 (39)		4 (84)	9 (21)	5 (24)	8 (24)
Carbon monoxide														8	6 (40)	2 (18)		2 (18)	4 (10)	2 (8)	4 (12)
Carbon disulfide (CH_2)	$\mu g/m^3$						100														
Hydrogen sulfide (HS)	$\mu g/m^3$						150														
Nitrogen dioxide (NO_2)	$\mu g/m^3$				200				40	200[a]			40	48		33 (184)		25 (60)		29 (125)	44 (107)
Ozone (O_3)	$\mu g/m^3$					120															
Particulate matter (PM[b])												50	40[e]	48	37 (119)	30 (289)	32 (111)	15 (177)		18 (77)	35 (145)
Sulfur dioxide (SO_2)	$\mu g/m^3$	500					125		50	350[c]		125[d]	20								
Elements																					
Arsenium (As)	ng/m^3												5[f]	48	20 (53)	5 (11)	11 (37)	4 (13)		5 (18)	
Cadmium (Cd)	ng/m^3								5				6[f]	48							
Lead (Pb)	$\mu g/m^3$								0.5				0.5	48	0.09 (02)	0.02 (0.05)	0.05 (0.14)	0.01 (0.04)		0.03 (0.17)	
Manganese (Mn)	$\mu g/m^3$								0.15					48	0.19 (5.43)	0.003 (0.01)	0.02 (0.11)	0.01 (0.11)		0.003 (0.02)	
Mercury (Hg)	$\mu g/m^3$								1					48	0.06 (0.99)	0.04 (0.28)	0.003 (0.06)	0.04 (0.6)		0.002 (0.01)	
Nickel (Ni)	ng/m^3												20[f]	48	5.2 (6.4)	5.2 (6.1)	2.7 (8.4)	4.9 (7.2)		16 (57)	
Vanadium (V)	$\mu g/m^3$						1							48	0.005 (0.027)	0.001 (0.002)	0.005 (0.016)	0.002 (0.015)		0.002 (0.005)	

[a] 18 annual exceedances allowed.
[b] EU limit values for PM_{10}, *EXPOLIS* measurements for $PM_{2.5}$.
[c,d] 24 and 3 annual exceedances allowed, respectively.
[e] Limit value drops to 20 $\mu g/m^3$ after 2005.
[f] Suggested limit values.

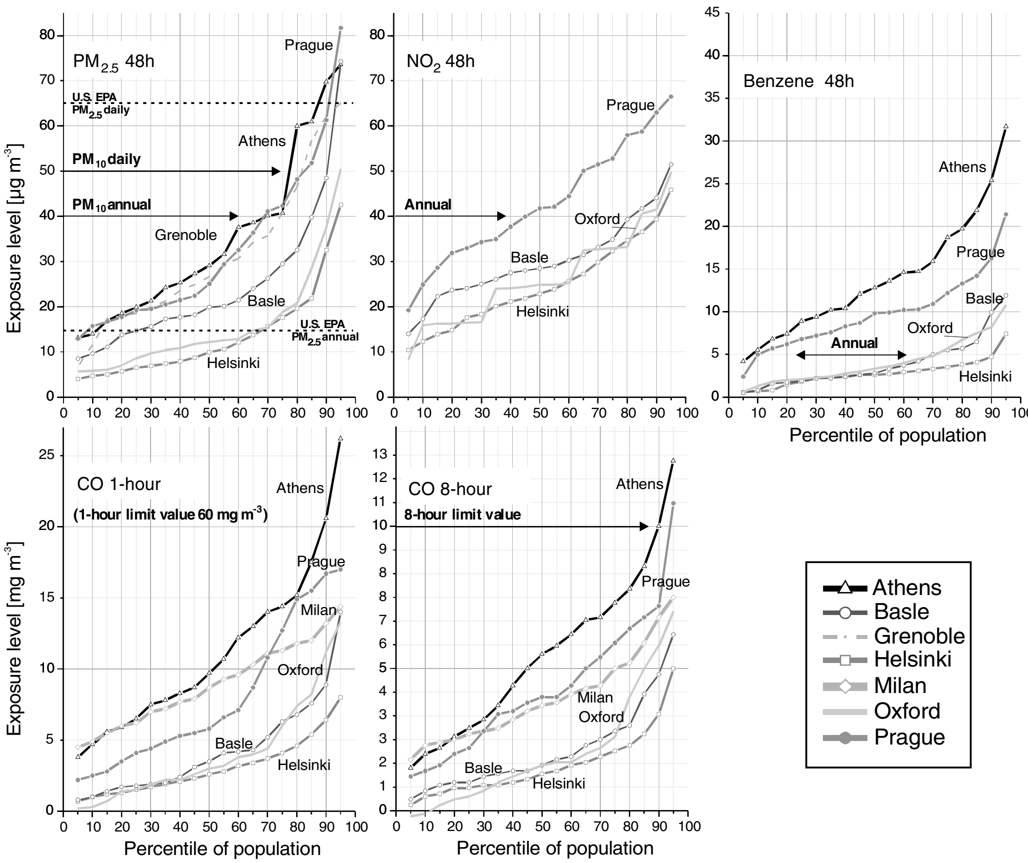

Figure 3. Exposure distributions of selected air pollutants observed in the *EXPOLIS* study compared to the EU limit values (horizontal arrows) and US EPA guidelines (dashed lines; $PM_{2.5}$ only).

CO data is available from six cities (Athens, Basle, Helsinki, Milan, Oxford and Prague). The 8-h EC limit value (10 mg/m^3) is exceeded by the observed 95th percentile in Athens and Prague (Figure 3) and by the maximum observed 8-h exposure level in Basle and Helsinki (Table 5). Even in Milan and Oxford the highest 1-h exposure level is more than double compared to the 8-h limit value. The WHO 1-h guideline (60 mg/m^3) is exceeded by corresponding maximum exposures in Athens and Helsinki (Table 5). CO is a good example of an air pollutant, for which the highest exposures are not at all related to the ambient concentrations.

Benzene exposures were measured in five cities (Athens, Basle, Helsinki, Oxford and Prague). Population average exceeded EC limit value for annual average benzene concentration (5 μg/m^3) clearly in Athens and Prague, and just slightly in Basel. Levels in Helsinki and Oxford are quite close to the limit value as well. Highest observed personal exposures exceed the limit value five-fold in Oxford and more in the other cities (Table 5).

Particulate matter limit value has been set only to PM_{10} particles in Europe, although also $PM_{2.5}$ limit values are under consideration. USA Environmental protection agency has set guidelines for $PM_{2.5}$ as 65 μg/m^3 (24-h) and 15 μg/m^3 (annual). The EC annual limit value for PM_{10} is 40 μg/m^3, which is not exceeded by the average population $PM_{2.5}$ exposure in any of the five cities with personal data (Athens, Basle, Helsinki, Oxford and Prague). Only microenvironment levels are available from Milan, where both the average residential indoor and the average workplace indoor air concentration exceeded the annual limit value. Highest observed personal $PM_{2.5}$ exposure levels are two to five times higher than the 24-h PM_{10} limit value (50 μg/m^3) in all other cities except in Oxford, where the observed maximum is 77 μg/m^3 (being only 1.5 times higher than the limit value; Table 5). These results show that a fraction of personal $PM_{2.5}$

(and thus also PM_{10}) exposures in all cities exceed the limit value set for the ambient air. It must be noted that when comparing $PM_{2.5}$ levels to PM_{10} limit values, a higher observed level indicates a sure exceedance, while a level below the limit value does not guarantee that the PM_{10} level would not in fact reach the limit. This is due to the fact that by definition all $PM_{2.5}$ particles are also PM_{10} particles (i.e. smaller than 10 μm in aerodynamic diameter).

The EC limit value for annual PM_{10} concentration is lower than the US EPA guideline value for $PM_{2.5}$ and thus the limit value requirement is much tighter. On the average, approximately half of the PM_{10} particle mass is caused by particles smaller than 2.5 μm (i.e. by $PM_{2.5}$ particles). This approximation would imply that the annual EC limit value would correspond to $PM_{2.5}$ level of 25 $\mu g/m^3$, less than half of the US guideline. The EC limit value for 24-h PM_{10} levels is set first at 40 $\mu g/m^3$, but then reduced to 20 $\mu g/m^3$ between 2005 and 2010. The latter level would correspond to approximately 10 $\mu g/m^3$ when translated to $PM_{2.5}$ particles, a level also tighter than the corresponding US guideline (15 $\mu g/m^3$). Tight limit values support efficient public health protection, but the observed exposure levels, when compared against these values, indicate that in reality large fraction of the urban European population is exposed to much higher concentrations.

Elemental composition of $PM_{2.5}$ is available from five cities (Athens, Basle, Grenoble, Helsinki and Oxford). Annual limit value has been set to lead (Pb; 0.5 $\mu g/m^3$). Neither average nor maximum observed exposure level exceeded this in any of the cities. In the future, limit value will probably be set also for arsenic (As), cadmium (Cd) and nickel (Ni) (suggested limit values 6, 5 and 20 ng/m^3, respectively). The average population exposure level exceeds the suggested arsenic limit value in two cities (Athens and Grenoble) and maximum personal exposure level in all five cities. The average exposure levels exceed the suggested cadmium limit value in three cities and the maximum personal exposures exceed it 6–10-fold in all cities. There was a large variation in the cadmium blank values and it is possible that the observed differences between the cities are caused by the variability in the blank filter contamination and thus the cadmium results were not included in Table 5. It seems probable, however, that the suggested cadmium limit value is exceeded in many of the cities. The average exposure levels are below the suggested nickel limit value in all cities, but the maximum personal level exceeded it in Oxford.

All these exceedances indicate that during the 1996–1999 situation, population exposures to these substances were considerable and that there indeed is a need to control these exposures. Analysis of total exposures and corresponding new control strategies are needed besides ambient air guidelines and limit values to ensure also safe and healthy indoor environment for the urban populations.

Traffic as a Emission Source and an Exposure Microenvironment

During the last few decades tremendous progress has been achieved in lowering industrial and energy production emissions. At the same time the traffic, especially road traffic has continued to increase. While auto industry has been able to continuously provide new models with lower emission rates, the increase in the traffic volume and ageing car fleet have kept the total emission levels quite high. Since traffic is by nature distributed evenly to the areas where most people are, the emission-to-exposure ratio, the so-called *intake fraction* (Bennett et al., 2002), is high. The determinant analysis conducted this far on $PM_{2.5}$, its elemental composition, BS and VOC exposures in the *EXPOLIS* study have been able to indicate traffic originating fractions of the exposures. The $PM_{2.5}$ exposure level while in traffic seems to be in average two times higher than the overall average exposure level. About half of the total exposure to traffic-generated fine PM appears to be acquired while commuting. Thus time spent in road traffic is an important determinant of personal exposure levels.

Exposure to traffic-generated pollutants occurs of course also while persons are not in the traffic themselves. As the time spent in traffic is typically something like 5–10% of the total daily time, a substantial fraction of traffic-originating exposures occurs outside traffic. Traffic-generated pollutants infiltrate indoors and this can be controlled by building design, ventilation systems and separating the occupied indoor spaces from direct influence of vehicles by, for example, using detached garages.

The role of long-range transported pollution has been acknowledged for a long time. The contribution of traffic to the formation of the long-range transported pollution, however, has not been clearly separated from industrial and natural emissions. As more than half of the $PM_{2.5}$ concentrations in Europe are caused by long-range transportation, it is very important to attribute this exposure fraction to sources. When emissions are looked at local scale, it appears to be clear that industrial emissions—with high emission height and efficient emission controls—are no longer important for exposures locally, but this might not be true for long-range-transported pollution. Also, the role of emissions in developing countries or in Eastern Europe might prove to be significant contributors to exposures in Europe and they might be among the most cost-effective targets for exposure controls.

Sensitive Population Groups

Exposure analysis is achieving the required level of sophistication to produce exposure estimates for specific population groups, including elderly, infants and persons suffering from a specific disease, like lung or heart conditions. When alternative and often expensive environmental policy scenarios are compared, it is essential to look at their efficiencies in reducing population exposures especially in those subpopulations where the burden of adverse health effects is the highest. If this analysis is based on centrally monitored ambient air quality data only, and dose–response factors obtained for the general population, non-optimal policy may be selected. To allow the risk analyst to use subpopulation-based exposure analysis efficiently, they should have available dose–response factors that would be specific to the target groups. The epidemiological studies are coming out more and more with this kind of information, but still much remains to be done in both epidemiology and toxicology.

Mixtures of Pollutants

During the past decades, air pollutant control mechanisms have been established using guideline levels for concentrations in occupational and ambient environments. Such guidelines might prove useful also for nonoccupational indoor environments, but because of the diversity of such environments, controlling the prevailing levels is not straightforward. Thus probably also other indoor air quality management approaches must be developed.

The guidelines are set for specified pollutants (WHO, 2000). When they are used, the concentration of the target pollutant is compared to the guideline level, taking into account the correct averaging time. This procedure does not account for health stress from multiple pollutants at the same time. Multiple stressors are taken into account in the epidemiological studies, where the total observed health effect is attributed to the explaining variables; there is no guarantee that the effects would be attributed to the correct causes, but the total effect of multiple stressors is seen. When guidelines are set, the scientific evidence used as background information may well include data on effects caused together with coexisting pollutants to some extent. In a given situation, however, dozens of air pollutants occur together and contribute to the health stress in synergetic or additive ways. For example, the analysis of the *EXPOLIS* data shows that the same subjects are typically exposed to high levels of many pollutants. New techniques and approaches are needed to study the effects of these multiple stressor exposures. New information is needed to create control mechanisms to protect the public in situations where the concentration of each single pollutant is below the guideline, but the combination of these pollutants poses a risk to the health.

Conclusions

Production and availability of population-based exposure data for exposure analysis has been a high priority goal for exposure research during the last decade. Such data is needed for development of efficient exposure control and reduction strategies: when a pollutant has multiple routes of exposures, the dominating route (s) should be controlled first. In cases of multiple emission sources, reductions should be focused on the sources contributing most to the exposures. The roles of different routes and sources should always be analyzed before making decisions about costly—or otherwise harmful—interventions.

The TEAM studies (e.g. Wallace et al., 1987; Özkaynak et al., 1996) and the NHEXAS research program (Lioy and Pellizzari, 1995; Pellizzari et al., 2001) have produced such data in the US and databases have been made available to support maximal use of these data collected using public funds. The current work describes the development and content of the first European multipollutant, multicenter exposure database focusing on inhalation exposures. A European database was created, combining questionnaire, air concentration and exposure data from seven cities collected during 1996–1999. The database structures are described in the project document Hänninen et al. (2002a) and in the current paper. The *EXPOLIS* databases are collected on a single CD-ROM disk containing both the local databases from each of the study centers as well as the combined CIDB with the main results.

The multipollutant exposure data from the same subjects in the random population samples allows for analyses of the determinants, microenvironments and sources of exposures to multipollutant mixtures. Almost 30 papers have been published in peer-reviewed journals presenting data analysis results of the *EXPOLIS* data. These papers, reviewed shortly in the current paper, prove the usability of the *EXPOLIS* database and demonstrate many aspects of data analysis that can be conducted using the data.

EC pursues to develop guidelines for new pollutants, including $PM_{2.5}$, and methodologies to control exposures to pollutants and chemicals with significant indoor sources. The collected exposure data in the *EXPOLIS* database should, can and will be used to support these processes among other available tools and exposure analysis techniques.

Acknowledgements

The *EXPOLIS* Centers are KTL (coordinating center) in Helsinki (Finland), University of Athens (Greece), University of Basle (Switzerland), University Joseph Fourier in Grenoble (France), University of Milan (Italy), Inst. Exp. Medicine ASCR and Health Inst. of Central Bohemia (Czech), VTT in Helsinki (Finland), RIVM in Bilthoven (The Netherlands) and Imperial College (United Kingdom). This work has been supported by EL contracts ENV4-CT96-0202 (DB12-DTEE) and ERB IC20-CT96-0061 (Prague), Academy of Finland contracts Nos. 36586, 40835 and 42610 (Helsinki), Swiss Ministry for Education and Science contract BBW No, 95.0894 (Basel), French National Environment Agency (ADEME), Union Routière de France and Grenoble Communauté de Communes (Grenoble), DEFRA contract EPG1/5/106 (Oxford), other national research funds and intramural funding from participating institutes. We express our gratitude also to the noncited persons participating the conduction of the study: Alcini D., Balducci F., Bollini D., Braiati S., Braun T., Bürer K., Caparis M., Ghalandar R., Halonen T., Heeb D., Hummel T., Lovato L., Manninen K., McMullen M., Piippo A-M., Pipinen T., Radaelli G., Roth B., Sandroni R., Stambej T., Szegvary T., Tähtinen M., Tountas Y., Vercelli F., Visigalli F., Vouros P., Wigzell E.

References

Bennett D.H., Mckone T.E., Evans J.S., Nazaroff W.W., Margni M.D., Jolliet O., and Smith K.R. Defining intake fraction. *Environ Sci Technol* 2002: 36(9): 206A–211A.

Boudet C., Zmirou D., and Vestri V. Can one use ambient air concentration data to estimate personal and population exposures to particles? An approach within the EXPOLIS study. *Science of the Total Environment* 2001: 267: 141–150.

Boudet C., Dechenaux J., Balducci F., Masclet P., and Zmirou D. Total personal exposure to fine particulate matter in an urban adult population: the role of ambient air and transport. *Int J Vehicle Des* 1998: 20(1–4): 30–38.

Bruinen de Bruin Y., Carrer P., Jantunen M., Hänninen O., Scotto di Marco G., Kephalopoulos S., Cavallo D., and Maroni M. Personal carbon monoxide exposure levels; contribution of local sources to exposures and microenvironmental concentrations in Milan. *JEAEE* 2004a, in press.

Bruinen de Bruin Y., Hänninen O., Carrer P., Maroni M., Kephalopoulos S., Scotto di Marco G., and Jantunen M. Simulation of working population exposures to carbon monoxide using EXPOLIS-Milan microenvironment concentration and time activity data. *JEAEE* 2004b: 14(2): 154–163.

Edwards R., and Jantunen M. Benzene exposure in Helsinki, Finland. *Atmos Environ* 2001: 35(8): 1411–1420.

Edwards R.D., Jurvelin J., Koistinen K., Saarela K., and Jantunen M. VOC source identification from personal and residential indoor, outdoor and workplace microenvironment samples in EXPOLIS-Helsinki, Finland. *Atmos Environ* 2001a: 35: 4829–4841.

Edwards R.D., Jurvelin J., Saarela K., and Jantunen M.J. VOC concentrations measured in personal samples and residential indoor, outdoor and workplace microenvironments in EXPOLIS-Helsinki, Finland. *Atmos Environ* 2001b: 35: 4531–4543.

Georgoulis L.B., Hänninen O.O., Samoli E., Katsouyanni K., Künzli N., Sram R., Bruinen de Bruin Y., Alm S., and Jantunen M. Personal carbon monoxide exposures in five European cities and its determinants. *Atmos Environ* 2002: 36: 963–974.

Götschi T., Oglesby L., Mathys P., Monn C., Manalis N., Koistinen K., Jantunen M., Hänninen O.O., Polanska L., and Künzli N. Comparison of black smoke and $PM_{2.5}$ levels in indoor and outdoor environments of four European cities. *EST* 2002: 36(6): 1191–1197.

Hänninen O., Lebret E., and Jantunen M. Simulating personal $PM_{2.5}$ exposures using ambient concentrations as inputs, manuscript, 2004.

Hänninen O.O., Alm S., Kaarakainen E., and Jantunen M. *The EXPOLIS Databases*, B13/2002. Publications of the National Public Health Institute, Kuopio, KTL, 2002a.

Hänninen O.O., Koistinen K.J., Kousa A., Keski-Karhu J., and Jantunen M.J. Quantitative analysis of environmental factors in differential weighing of blank Teflon filters. *J Air Waste Manage Assoc* 2002b: 52(2): 134–139.

Hänninen O.O., Kruize H., Lebret E., and Jantunen M. EXPOLIS simulation model: $PM_{2.5}$ application and comparison with measurements in Helsinki. *J Exp Anal Environ Epidemiol* 2003: 13: 74–85.

Helm D., Jantunen M., and Rotko T. Reporting personal results to participants of exposure studies. *Sci Tot Env* 2000: 262(1–2): 191–195.

Hern S.C., Robertson G.R., Butler L.C., Engelmann W.H., Kantor E.J., Quackenboss J.J., Behar J.V., and Pandian M.D. Reaching "THERdbASE" in human exposure research, Abstracts of Papers of the American Chemical Society, 1997 Vol. 214, 134-ENVR.

Jantunen M.J., Hänninen O.O., Katsouyanni K., Knöppel H., Künzli N., Lebret E., Maroni M., Saarela K., Sram R., and Zmirou D. Air pollution exposure in European cities: The EXPOLIS-study. *J Exp Anal Environ Epidemiol* 1998: 8(4): 495–518.

Jurvelin J., Edwards R., Saarela K., Laine-Ylijoki J., De Bortoli M., Oglesby L., Schläpfer K., Georgoulis L., Tischerova E., Hänninen O.O., and Jantunen M. Evaluation of VOC measurements in the EXPOLIS study. *J Environ Monit* 2001a: 3: 159–165.

Jurvelin J., Vartiainen M., Jantunen M., and Pasanen P. Personal exposure levels and microenvironmental concentrations of formaldehyde and acetaldehyde in Helsinki metropolitan area, Finland. *J Air Waste Manage Assoc* 2001b: 51: 17–24.

Jurvelin J.A., Edwards R.D., Vartiainen M., Pasanen P., and Jantunen M.J. Residential indoor, outdoor, and workplace concentrations of carbonyl compounds: relationships with personal exposure concentrations and correlation with sources. *J Air Waste Manage Assoc* 2003: 53(5): 560–573.

Koistinen K., Edwards R., Mathys P., Ruuskanen J., Künzli N., and Jantunen M. Sources of $PM_{2.5}$ in personal exposures and residential indoor, outdoor and workplace microenvironments in EXPOLIS-HELSINKI, FINLAND. *Scand J Work, Environ Health* 2004, in press.

Koistinen K.J., Hänninen O.O., Rotko T., Edwards R.D., Moschandreas D., and Jantunen M.J. Behavioral and environmental determinants of personal exposures to $PM_{2.5}$ in EXPOLIS-Helsinki, Finland. *Atmos Environ* 2001: 35(14): 2473–2481.

Koistinen K.J., Kousa A., Tenhola V., Hänninen O.O., Jantunen M.J., Oglesby L., Künzli N., and Georgoulis L. Fine particle ($PM_{2.5}$) measurement methodology, quality assurance procedures, and pilot results of the EXPOLIS study. *J Air Waste Manage Assoc* 1999: 49: 1212–1220.

Kousa A., Monn C., Rotko T., Alm S., Oglesby L., and Jantunen M. Personal exposures to NO_2 in the Expolis-study: relation to residential indoor, outdoor and workplace concentrations in Basel, Helsinki and Prague. *Atmos Environ* 2001: 35(20): 3405–3412.

Kousa A., Oglesby L., Koistinen K., Kunzli N., and Jantunen M. Exposure chain of urban air $PM_{2.5}$—associations between ambient fixed site, residential outdoor, indoor, workplace and personal exposures in four European cities in the EXPOLIS-study. *Atmos Environ* 2002: 36: 3031–3039.

Kruize H., Hänninen O.O., Breugelmans O., Lebret E., and Jantunen M. Description and demonstration of the EXPOLIS simulation model: two examples of modeling population exposure to particulate matter. *J Exp Anal Environ Epidemiol* 2003: 13(2): 87–99.

Laden F., Neas L.M., Dockery D.W., and Schwartz J. Association of fine particulate matter from different sources with daily mortality in six U.S. cities. *Environ Health Perspect* 2000: 108(10): 941–947.

Lai H.K., Kendall M., Ferrier H., Lindup I., Alm S., Hänninen O., Jantunen M., Mathys P., Colvile R.N., Ashmore M.R., Cullinan P., and Nieuwenhuijsen M.J. Personal exposures and microenvironment concentrations of $PM_{2.5}$, VOC, NO_2 and CO in a medium-sized town, Oxford, UK, manuscript, 2004, Submitted.

Lioy P.J., and Pellizzari E. Conceptual framework for designing a national survey of human exposure. *J Exp Anal Environ Epidemiol* 1995: 5: 425–444.

Mathys P., Stern W.B., Oglesby L., Braun-Fahrländer C., Ackermann-Liebrich U., Jantunen M.J., and Künzli N. Elemental analysis of airborne particulate matter by ED-XRF within the European EXPOLIS Study. *ICP Inform Newslett* 2001: 27(3): 29–34.

Oglesby L., Künzli N., Röösli M., Braun-Fahrländer C., Mathys P., Stern W., Jantunen M., and Kousa A. Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution. *J Air Waste Manage Assoc* 2000a: 50: 1251–1261.

Oglesby L., Rotko T., Krütli P., Boudet C., Kruize H., Jantunen M.J., and Künzli N. Personal exposure assessment studies may suffer from exposure relevant selection bias. *J Exp Anal Environ Epidemiol* 2000b: 10(3): 251–266.

Özkaynak H., Xue J., Spengler J., Wallace L., Pellizzari E., and Jenkins P. Personal exposure to airborne particles and metals: results from the particle team study in Riverside, California. *J Exp Anal Environ Epidemiol* 1996: 6: 57–78.

Pellizzari E.D., Smith D.J., Clayton C.A., Michael L.C., and Quackenboss J.J. An assessment of the data quality for NHEXAS—Part I: exposure to metals and volatile organic chemicals in Region 5. *J Exp Anal Environ Epidemiol* 2001: 11(2): 140–154.

Pope C.A., Burnett R.T., Than M.J., Calle E.E., Krewski D., Ito K., and Thurston G.D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J Am Med Assoc* 2002: 287(9): 1132–1141.

Rotko T., Koistinen K., Hänninen O.O., and Jantunen M. Sociodemographic descriptors of personal exposure to fine particles ($PM_{2.5}$) in EXPOLIS-Helsinki. *J Exp Anal Environ Epidemiol* 2000a: 10(4): 385–393.

Rotko T., Oglesby L., Künzli N., and Jantunen M. Population sampling in European air pollution exposure study, EXPOLIS: comparisons between the cities and representativity of the samples. *J Exp Anal Environ Epidemiol* 2000b: 10(4): 355–364.

Rotko T., Kousa A., Alm S., and Jantunen M. Exposures to nitrogen dioxide (NO_2) in Expolis-Helsinki: microenvironment, behavioral and sociodemographic factors. *J Exp Anal Environ Epidemiol* 2001: 11: 216–223.

Rotko T., Oglesby L., Künzli N., Carrer P., Nieuwenhuijsen M.J., and Jantunen M. Determinants of perceived air pollution annoyance and association between annoyance scores and air pollution ($PM_{2.5}$, NO_2) concentrations in the European EXPOLIS study. *Atmos Environ* 2002: 36(29): 4593–4602.

Seifert B., Becker K., Hoffmann K., Krause C., and Schulz C. The German Environmental Survey 1990/1992 (GerES II): a representative population study. *J Exp Anal Environ Epidemiol* 2000: 10(2): 103–114.

Wallace L.A., Pellizzari E.D., Hartwell T.D., Sparacino C., Whitmore R., Sheldon L., Zelon H., and Perrit R. The TEAM Study: personal exposures to toxic substances in air, drinking water and breath of 400 residents of New Jersey, North Carolina and North Dakota. *Environ Res* 1987: 43: 290–307.

WHO. *Air Quality Guidelines for Europe*, European Series No. 91, 2. WHO Regional Publications, Copenhagen, 2000.