Study manual

ESCAPE

European Study of Cohorts for Air Pollution Effects

ENV.2007.1.2.2.2. European cohort on air pollution

Coordinator: Bert Brunekreef, PhD
Institute for Risk Assessment Sciences
Utrecht University
The Netherlands

Version2008-1
October 3, 2008

Table of Contents

Preface

This manual contains the procedures for exposure assessment, data collection and data analysis of the ESCAPE project. The background for the project can be found in the Description of work. The actual air pollution measurements will be conducted according to standard operating procedures (SOP).

The text on exposure assessment in this manual is our proposal for the definitive text. This applies especially to the sections on the monitoring campaign (section 3.1). The sections on stochastic models and GIS predictor variables will be further developed in the first year of the project. The text about epidemiological analysis (Chapter 1) is a first draft – still skeleton-like – strategy for analyses in the ESCAPE project and will be extended.

1 Objectives

The study will investigate long-term effects on human health of exposure to air pollution in Europe. The background is that current estimates of the European health impact of especially fine particles in the air are large. However, available estimates are primarily based on exposure response relationships established in studies from North America. There is an urgent need to perform studies in Europe on recent and current exposures, and to use refined exposure assessment tools.

The objectives of the study are:

1. To develop a flexible methodology for assessment of long-term population exposure to air pollution focused primarily on fine particles, particle composition, and nitrogen oxides.
2. To apply the exposure assessment methodology on existing cohort studies of mortality and chronic disease in Europe that have been selected based on their potential to quantify relationships between long-term exposure and health response precisely.
3. Specifically, to investigate exposure-response relationships and thresholds for (a) adverse prenatal health outcomes, and development of diseases such as asthma in children; (b) respiratory disease endpoints in adults; (c) cardiovascular disease endpoints in adults; (d) all-cause and cause-specific mortality, and cancer incidence.
4. To develop a database for quantitative estimates of the health impacts of long-term exposure to air pollution for all of these health endpoints for the European population

2 Study design

The ESCAPE study will significantly increase knowledge on long-term health effects of air pollution by efficiently using health data already available from European cohort studies. These cohort studies included subjects across all ages, including the prenatal period up to the elderly. Several of these studies contain relevant biomarker data as well as genetic information so that susceptible sub-groups and gene-environment interactions can be addressed. Practically all studies include male as well as female participants so that effects can be addressed by gender. Several of the included studies have data on traffic noise exposure, and the study will utilize the noise maps as they will become available in the near future as per EU regulations. Additional exposures (for instance indoor pollution and environmental tobacco smoke) will be considered as many of the studies have data on these already. Socioeconomic factors will be addressed not only at the individual but also at the neighbourhood level. A single harmonized exposure assessment protocol will be used with a particular emphasis on understanding individuals' exposure to traffic-related pollution. Exposure models will be validated against actual measurements of air pollution. Epidemiologic data analyses will be further harmonized with respect to statistical and spatial modelling, specification of confounder models etc. Collaboration with a major research initiative (MESA-AIR) in North America has been ensured, and two Taiwanese cohorts are included as a model for capacity building and dissemination outside Europe.

Long-term exposure to air pollution has been associated with a range of adverse health effects. We have chosen for a broad representation of adverse health effects in this study. We included several studies in each of four 'effect' categories: a) adverse pregnancy outcome studies, and birth cohort studies of children for outcomes such as asthma and allergy; b) cohort studies of respiratory biomarker and morbidity endpoints; c) cohort studies of cardiovascular biomarker and morbidity endpoints; and d) cohort studies of non-accidental and cause specific mortality, and cancer incidence. The description of work provides the inclusion criteria.

The study will be of four year duration. Because of the policy interest in the study, we should anticipate that there is little room for delays. Careful planning is hence necessary.

3 Exposure assessment

The exposure assessment work package will develop tools for exposure assessment, and provide these to the four endpoint work packages for implementation. The work package will employ a mixture of measurements and modelling to estimate exposure of study participants to ambient air pollution. Specifically, we will perform spatially resolved measurements of PM10, PM2.5, the soot content of PM2.5 and NOx (section 3.1). Measurements will be used to develop stochastic models. This involves collection of predictor variables using GIS (section 3.2) and development / validation of stochastic models (section 3.3). Exposure estimates will be assigned to study participants on the basis of these LUR models (section 3.4). In some areas dispersion models are available and the results of these models will be validated with the monitoring data (section 3.5). Birth cohort studies need exposure metrics on a finer time scale (section 3.6). For some of the cohorts health measurements have been collected already. Hence historical exposures need to be estimated (section 3.7). Information about residential history may then also be interesting (section 3.8). Because both air pollution and noise may be correlated, noise assessment will be made (section 3.9), allowing an analysis of the potentially independent effects of air pollution and noise.

Much of the texts below is based upon previous land-use regression studies we were involved in (e.g. SAVIAH, TRAPCA, GALEN). A review of land-use regression methodology prepared by IRAS, Imperial College and RIVM has recently been accepted and is available upon request.

In order to ensure exposure information that is useful for the endpoint WPs, regular contact with the WP leaders and the coordinator will be ensured.

3.1 Air pollution monitoring

The planned exposure measurements have the following aims:

- To document the spatial variation of long-term average ambient PM, PM composition and NO_x concentrations in a number of European cities/areas.
- To develop and validate site-specific stochastic models which link the measurements to geographic and land use data such as road network and usage, elevation, population density.
- To validate site-specific dispersion model calculations against ambient measurements.
- To collect and store ambient samples of PM10 and PM2.5 for later analysis of chemical composition and toxicity using state of the art in vitro methods.
- To establish the distribution of exposure for truly prospective analyses in ongoing cohort studies with planned observations of health and disease status in the future.

We will measure the spatial variation of annual average PM concentrations in 21 cities / areas. In all of these the spatial variation of NOx will be measured at the same locations plus additional locations. In a further

18 cities / areas we will measure the spatial variation of annual average NOx concentrations. Table 1 mentions the specific selected epidemiologic cohort studies as well as the planned measurements. The more intensive PM + NO_x monitoring will take place in the most informative areas (> 1 study, large populations and/or cities). Figure 1 shows the locations of ESCAPE monitoring areas in Europe. Tables 2 and 3 list the order of the measurements and the responsible institute.

Table 1. Study areas, and epidemiologic studies included in ESCAPE.

Study Area	Epidemiologic study acronym	Measurements
Györ, Hungary	APREG	PM+NOx
Florence, Italy	EPIC	NOx
Turin, Italy	EPIC, SIDRIA, ECRHS	PM+NOx
Varese, Italy	EPIC	NOx
Verona, Italy	ECRHS	NOx
Pavia, Italy	ECRHS	NOx
Rome, Italy	GASPII, SIDRIA	PM+NOx
Barcelona, Spain	ECRHS, INMA	PM+NOx
San Sebastian, Galdakao, Spain	EPIC, INMA, ECRHS	NOx
Huelva, Spain	ECRHS	NOx
Oviedo, Spain	ECRHS, INMA	NOx
Girona, Spain	REGICOR	PM+NOx
Athens, Greece	EPIC	PM+NOx
Heraklion, Greece	RHEA	PM+NOx
Oxford, Norfolk, Norwich, Ipswich, UK	EPIC, ECRHS, UK 1946 cohort	PM+NOx
Bradford, UK	BIB	NOx
Manchester, UK	MAAS, UK 1946 cohort	PM+NOx
Utrecht, Netherlands	EPIC	PM+NOx
Amsterdam, Netherlands	EPIC, ABCD	
Doetinchem, Netherlands	EPIC	
Maastricht, Netherlands	EPIC	PM+NOx
Rotterdam, Netherlands	PIAMA	
Antwerp, Belgium	ECRHS	
Heidelberg, Germany	EPIC	NOx
Erfurt, Germany	ECRHS	NOx
Ruhr Area, Germany	SALIA, RECALL	PM+NOx
Munich, Germany	LISA + GINI	PM+NOx
Augsburg, Germany	KORA	
Lugano, Switzerland	SAPALDIA	PM+NOx
Basel, Switzerland	SAPALDIA, ECRHS	NOx
Geneva, Switzerland	SAPALDIA	NOx
Vorarlberg, Austria	VHM&PP	NOx

Study Area	Epidemiologic study acronym	Measurements
Copenhagen, Denmark	DCH, National Birth Cohort	PM+NOx
Oslo, Norway	HUBRO, MOBA	PM+NOx
Stockholm, Sweden	BAMSE, TWINGENE, 60 YEAR OLDS	PM+NOx
Umea, Sweden	ECRHS, EPIC	NOx
Paris, France	ECRHS, EPIC, GAZEL, EGEA	PM+NOx
Grenoble, France	ECRHS, EGEA, GAZEL	NOx
Marseille, France	EPIC, EGEA, GAZEL	NOx
Lyon, France	EPIC, EGEA, GAZEL	NOx
Nancy, Poitiers, France	EDEN	NOx *
Helsinki, Turku, Finland	FINRISK	PM+NOx
Cracow, Poland	HAPIEE	PM+NOx
Kaunas, Lithuania	KANC	PM+NOx

* Measurements funded locally (conducted in 2007)

Figure 1. Locations of ESCAPE monitoring areas in Europe

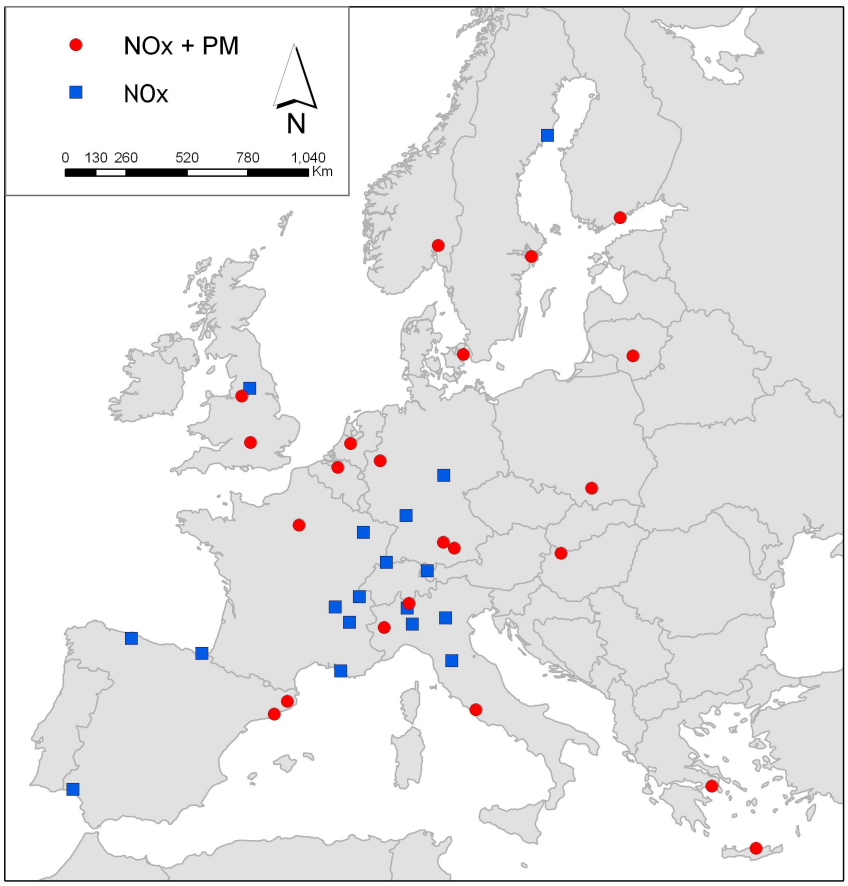

Table 2. Planning of PM10, PM2.5 and NO_x measurements

First year (Oct 2008 – July 2009)			Second year (Oct 2009 – July 2010)	
Study area	**Partner**		**Study area**	**Partner**
Netherlands	IRAS		Copenhagen	Cancer Institute
Belgium / Netherlands	IRAS		Athens	NKUA
Barcelona	CREAL		Helsinki	KTL
Girona	CREAL		Rome	ASL Rome
Manchester	Uniman		Turin	ASL Rome
Munich / Augsburg	Helmholtz		Oxford	Imperial
Stockholm	Karolinska		Cracow	Helmholtz
Ruhr area	IUTA		Paris	INVS
Lugano	UNIBAS		Kaunas	Uni Kaunas
Oslo	NIPH		Györ	NIEH
Heraklion	U Crete			

Table 3. Planning of NO_x only measurements

First year (Oct 2008 – July 2009)			Second year (Oct 2009 – July 2010)	
Study area	**Partner**		**Study area**	**Partner**
San Sebastian	CREAL		Vorarlberg	Unibas
Huelva	CREAL		Bradford	Imperial
Oviedo	CREAL		Florence	ASL
Erfurt	Helmholtz		Grenoble	INVS
Heidelberg	Unibas		Marseille	INVS
Basel	Unibas		Lyon	INVS
Geneva	Unibas		Nancy*	INVS
Umea	U Umea		Varese + Verona	Unibas
			Pavia	Unibas

* resolved in 2008/2009 whether measurements are necessary

PM measurements will be conducted at 20 monitoring sites per city / area; NOx measurements will be conducted at 40 monitoring sites per city / area. Measurements will be conducted for three periods of two weeks per site in the cold, warm and one intermediate temperature season. At one additional background site, PM and NO_x will be measured using the same instruments continuously during a complete year (starting in October/November) so that the discontinuous site specific measurements can be adjusted to the true long term average for the observation period. Hence, there will be 21 PM sites and 41 NO_x for each city / area. The data from the additional background site will only be used to adjust the discontinuous site

measurements to the true long-term average for the observation period. The data from the additional background site will not be used for model development. These methods have been developed and validated in the TRAPCA project.
Six units will be made available for each area so that 5 sites can be monitored simultaneously. PM_{10} and $PM_{2.5}$ will be measured with Harvard impactors according to SOPs used in previous studies (TRAPCA, ULTRA, RUPIOH). Weighing of all filters will be performed in one central laboratory (IRAS). The reflectance of all $PM_{2.5}$ filters will be measured in one central laboratory (IRAS). NO_x will be measured with Ogawa badges. Chemical analysis of all badges will be performed by IRAS. Sampling and site selection is the responsibility of the local partner. To harmonize site selection and field sampling, a two-day training workshop will be organized in the Netherlands in September 2008 for centres that measure in the first round. A similar training workshop for centres that measure in the second round will be organized in September 2009 (although they are also welcome at the workshop in September 2008). These training workshops are especially meant for field work coordinators.

We will thus use the same number of sites in each study area, independent of the size and complexity of the study area. Whereas this approach may not do justice to local complexities, for PM measurements this is governed by logistical constraints; for NOx measurements (which are far easier to perform), some more flexibility is possible.

3.1.1 Site selection

Sites will be selected to represent the anticipated spatial variation of air pollution at home addresses of participants in the epidemiologic studies. Hence a key input is the distribution of study homes over the study area. Sites should be broadly distributed proportional to cohort distribution. The study population especially determines the borders of the study area. Further, sites should be selected such that the anticipated spatial variation in air pollution is covered with sufficient measurements. If part of the study area has not much air pollution variation, fewer sites are needed in that area, even if a large part of the study population is living in that area. This choice can be made 'informally', that is without performing the actual GIS calculations of the different variables in different buffers around measurement points. If it is feasible to perform such GIS calculations (e.g. measures of urbanization, traffic intensities in areas), it is preferable however. Formal procedures for allocating sites exist (Kanaroglou et al. 2005), but are quite complex and require a significant amount of prior data on spatial variability and population distribution. To ensure a high level of comparability in choosing measurement sites, maps of proposed sites will be produced and submitted to a small working group (consisting of Gerard Hoek, Rob Beelen, Kees Meliefste, David Briggs, Mark Nieuwenhuijsen, Sally Liu, Christian Madsen, Josef Cyrys) within the exposure assessment work package for feedback.

The long term average outdoor concentration is a function of the regional background, additional pollution from all urban sources (resulting in an urban background) and pollution from very local sources (such as

traffic on nearby streets). Further, emissions from industries or ports may contribute to the long-term average concentration. For some cohorts the regional component is likely important when cohorts are spread out over large areas. Table 4 lists the variables that have been used in previous land use regression studies as predictor variables. In some regions wood smoke might also be a potential predictor variable.

Table 4. Overview of significant predictor variables in LUR studies

	Traffic [a]	Population [b]	Land use	Physical geography
Briggs et al, 1997	Road length, distance, intensity		Composite	Altitude
Briggs et al. 2000	Intensity	Housing		Altitude
Stedman et al. 1997	Emission		Urban	
Hoek et al. 2001	Intensity, road length	Housing		Region country
Carr et al. 2002	Intensity, congestion			
Brauer et al. 2003	Intensity, road length	Housing		Region country
Gonzales et al. 2005	Distance			Altitude, distance border
Gilbert al. 2005	Distance, intensity, road length	Population	Open space	
Smith et al. 2006	Intensity	Population		Altitude, distance border
Briggs et al. 2005b	Road length	Housing	Multiple	Altitude, distance sea, climate, topex
Hochadel et al. 2006	Intensity	Housing		
Ross et al. 2006	Intensity, road length			Distance to sea
Sahsuvaroglu et al. 2006	Intensity, distance		Industry, open space	Distance lake
Beelen et al. 2007	Intensity	Population	Urban	Region country
Morgenstern et al. 2007	Road length	Housing	Composite	
Rosenlund et al. 2007	Distance	Population	City zone	Altitude
Madsen et al. 2007	Road length			Altitude
Jerrett et al. 2007	Road length, distance, intensity	Housing	Industry	Geographical coordinate
Henderson et al. 2007	Road length,	Population	Commercial	Geographical coordinate, altitude
Aguilera et al. 2008	Road length		Composite	Altitude
Moore et al. 2007	Intensity		Industry, government	
Ross et al. 2007	Intensity	Population	Industry	
Ryan et al. 2007	Intensity, road length			Altitude
Briggs et al. 2007	Intensity		Composite	
Wheeler et al. 2008	Distance, road length			

a) Distance to major roads / freeways, traffic intensity or length of road types without traffic intensity data

b) Population of housing density

Monitoring sites will be selected with help of local expertise as well as available maps (local topographic maps, CORINE land cover maps) and satellite data (e.g. Google Earth). These sites should be divided among several well-chosen regional background, urban background and street type sites (in some areas, possibly sites representing other sources than road traffic). In the measurement program, traffic exposed sites will be overrepresented as these are expected to be the most informative when it comes to developing stochastic exposure models, and to validating results of dispersion models.
An urban background site is defined as a site that is not influenced substantially by sources in the 'direct vicinity' of the site. In a circle of 50 meter around the site no more than 3000 vehicles should pass daily. An estimate of the daily traffic intensity for a road can be made by counting of the traffic intensity for 15-minutes between 9.00 AM and 3.00 PM on a weekday (see for a more detailed description: Van Roosbroeck et al. Atmos Environ 2007;41:3381-3394). Within a circle of 100 meter around the site no other important sources of PM and NOx should be present (combustion sources, construction works, small industries, district heating plant, parking lot/garages etc.). The distance to large (industrial) sources should be considerably larger than 100 meter, depending on the specific emission characteristics (chimney height, temperature of flue gases). In order to obtain sufficient variability in the predictor variables, sites should be distributed well over the city, that is some sites in the city centre and several sites more in the suburbs of the city.
A street site is a site representing exposure to traffic pollution. Somewhat arbitrarily, a traffic intensity above 10,000 vehicles per day is considered as a traffic site. Preferably, no other sources should be present. In order to obtain sufficient variation among the predictor variables streets that cover the full range of traffic intensities and, possibly, composition in the city of interest should be selected. For example, include some streets with large fractions of buses or lorry traffic because this may diminish the correlation between total and heavy-duty vehicle counts. Also, include some 'open' and some 'street canyon' type streets as well. Do not select only streets with the highest traffic intensity, but also more intermediate streets. It is important that sites cover locations with varying air pollution concentrations, traffic intensities and different land uses. A traffic site should be located at the road side, and should be at ground level or at the first floor. Preferably do not use lamp posts as a monitoring location at traffic sites, because such a location is too close to a road (in case you might use lamp posts, hang them facing away from the street). A drain pipe is a better option. For urban and regional background locations lamp posts can be used.
At least one measurement site should be located at an existing routine monitoring site for calibration purposes. Previous monitoring locations (e.g. TRAPCA) are also attractive candidates for sites, as they also allow assessment of historic trends of spatial pollution patterns.

Based on these broad guidelines the local partner will make a proposal to be discussed with the aforementioned working group who will attempt to harmonize site selection across centres.

Measuring at the continuous monitoring location should start in October/November (even if you start measuring during the winter campaign). Before starting with the measurements at the continuous monitoring site, send a description of this location (see section 3.1.2) to Rob Beelen (r.m.j.beelen@uu.nl) AND

Gerard Hoek (g.hoek@uu.nl). They will distribute the description to the exposure assessment working group.
Do not start with your continuous measurement before the exposure assessment working group has approved the location of the continuous site!

Guidelines for the selection of sites are in summary:

1. Visit the study area and get a general understanding of the range of air pollution environments that need to be monitored and modelled.
2. Sites should be broadly distributed proportional to cohort distribution. Take into account the distribution of homes across the study area (% living in rural areas, % living near major roads). Make also sure that there is enough variation in air pollution concentrations and predictor variables at the selected sites.
   Each cohort should be characterized with respect to locations of homes. Make sure that some monitoring sites are included near the boundaries of the study areas, such that the developed models can be applied in the entire study area.
3. Assess which previously used predictor variables are likely important in your study area. Urbanization and traffic may be important key variables, and therefore sites could be divided in regional background, urban background and traffic sites.
4. Select sites that provide contrast in predictor variables (e.g. provide a range of traffic intensities rather than one extremely high and others with low intensities; provide also a range in urbanization). Over-sampling is needed in areas where pollution is higher and more variable (e.g. traffic areas) compared with other areas.
5. Select sites that are representative for homes of your study population, and, where applicable, also schools and day care centres. Try to avoid extreme situations, it is thus not a good idea to attach samplers to lampposts in the middle of major roads (too close to traffic) when these sites do not represent homes of the study population (a site should be at least 2 meters from the roadside, if there is more than 2m distance between road and building). Public buildings may be appropriate if their location to roads is comparable to homes (for background locations more flexibility is present).

It may be convenient to choose home addresses of participants of the cohort study. However this is not necessary as all we need to ensure is that sites are used that are representative for the study population.
If you choose home addresses of participants beware of confidentiality issues (i.e. sites will advertise location of cohort members).
The continuous monitoring site will measure for a complete year (filter/sampler should be changed every 2 weeks), and should be a background site that best represents the 'background' concentrations of PM and NOx that the population is exposed to, i.e. the continuous site should describe temporal variations in time. Whether this is an 'urban', 'suburban' or 'rural' background site will depend on the study population and

needs to be discussed with the aforementioned working group. A continuous background site should not be influenced by local traffic.

1. Every effort should be made to co-locate one of the monitoring sites in each city / area with an existing background PM and NOx monitoring station operated by a local, regional or national agency for comparison and validation purposes. This will allow us to 'calibrate' our measurements against longer time trends in monitoring data available locally from such sites.
2. In PM/NOx areas, NOx sites should be selected in the same way for sites with and without PM measurements. For example, do not locate combined PM/NOx traffic measurements at a balcony and for the NOx only measurements on lamp posts. Otherwise there might be a systematic difference between NOx measurements with and without a combined PM measurement. Suggestion is to select 40 NOx locations. Then randomly select in a stratified way 20 locations for PM measurements. In this case, there will be no systematic differences between NOx locations with and without PM measurements
3. Fulfils micro-environmental criteria (below)

Micro-environmental criteria for site selection are described below:
Unrestricted air flow around the sampler: sampler inlet should be placed at least 20 centimetres from any vertical surface (such as a wall), and if possible the sampler inlet should be located at least twice the distance from an obstacle as the height of that obstacle.

- Inlet should be at least 2 meter from a High volume sampler inlet and 1 meter from a medium volume sampler (such as the Harvard impactor), and at least 50 centimetres from a NO_x measurement. Sampling inlet should be at least 3 meters from the outlet of a pump.
- Sampler inlet should not be near exhaust flues or vents from homes or other buildings (at least 5 meters away).
  Preferably the sampler inlet should also be at least 5 m away from air conditioning.
- Sampling height (inlet) at least 1.5 m (preferably 2 m) above the surface on which the sampler is placed and preferably between 1.5 and 3 m above the ground
- The sampler should not be placed near the drip line of trees
- Electricity available (PM sampling only)
- Safe from vandalism or accidental damage
- Accessible for field workers

Other practical criteria:

- Do not locate the monitoring site within 25m of a traffic light or traffic junction / intersection
- A site should be at least 2 meters from the roadside (when there is more than 2m distance between road and building)
- Do not locate the monitoring site within 100 m of locations where construction works are ongoing or where construction works are planned during the measurement periods (contact the local authorities for information about planned construction works) or at locations that are likely to be subject to maintenance (e.g. affected by street-cleaning operations).

- Do not locate the monitoring site in a street where large changes in traffic intensities are to be expected in the future (for example due to traffic regulations).
- When using a residential address a non-smoking family is preferred. Do not use a balcony where cigarettes are smoked as a location for a monitoring site.
- Do not locate the monitoring site within 25m of locations where smokers are allowed to smoke and/or gather (for example close to the entrance of restaurants, hospitals, schools or other public buildings, because smokers usually congregate close to the entrance of these buildings).
- Do not use a location within 100 m of road tunnels, small industries and car-parks as these locations may be affected by local emission sources
- Do not use a location where there are a lot of birds

Do not use a location in highly exposed conditions which may affect operation of the samples. Avoid wind-exposed or turbulent environments.

Ad 1. This criterion is used to exclude sites close to buildings. Problems may occur due to surface reactions on walls or other vertical objects and turbulence behind buildings resulting in atypical concentrations. In urban areas it may not always be possible to find a site that is at a distance of twice the height of an obstacle. Make sure however that the site is at least 20 centimeters away from a wall. Also, do not select a yard enclosed completely by buildings.

Ad 2. These criteria are used to ensure that the sampler does not sample filtered air from another sampler

Ad 3. If the sampler is placed on a roof the preference is to have a roof without a chimney. Otherwise, the sampler should be spaced as far away as possible but at least 5 meters from a chimney removing waste gases from burning natural gas. Do not use a roof with chimneys emitting waste gases from burning oil, coal or solid waste.

Ad 4. As the measurements are conducted to estimate human exposure, the measurements should preferably be conducted at breathing height, which is about 1.50 meter. Particularly at background sites, the vertical concentration gradient is small, however. Hence basically any sampling height is acceptable. In Northern Europe, background is typically operated as rooftop measurements. For street locations we ought to be more strict but for practical purposes measurements conducted at any site which is accessible to the public will need to be conducted at a height not reachable by folks walking by. A balcony on the first floor is a good location when it is maximum 4m above street levels; ground floor of office blocks is sometimes very high.

If the sampler is located on top of a roof the sampler should be at least 1.5 m (preferably 2 m) above that surface. This is to limit the impact of wind blown dust from the surface.

Ad 5. Trees act as obstacles and may adsorb pollutants. The sampler should not be placed so close to trees that rain from tree leaves may affect the sampler. As a rule of thumb, the sampler should be placed the height of the tree away from that tree. A site location should also not be overhung by other dense vegetation.

Ad 6. A nearby power source should be available, but avoid long, exposed power cables. The power supply should be water-tight. The power supply should have the appropriate voltage/ampere and should be uninterrupted.
In most cases permission to use the power source will need to be negotiated. This thus takes time to arrange!

Ad 7. This is a practical but important criterion. In our experience protection from vandalism is achieved by selecting sites that are not readily visible from public roads (so samplers should not attract attention), are on balconies or protected by closed fences or gates or have guards present so they cannot be accessed by the public, or are too high to climb. In addition, the samplers may be protected by placing a special gate (to be rented or loaned from municipalities or renting companies) around the equipment. Locations that we have selected in the past as measurement sites included terrains of a police station, fire brigade, graveyard, gas distribution station, roof of city hall, garden of a monastery, garden of a church and recreational garden of a hospital.
Schools may not be the optimal choice since children are generally curious and more likely than adults to experiment with interesting looking objects. Other locations which should be avoided are locations near bus stops, where low walls/fences are available (on which children can climb to reach to the samplers) or other locations where children gather together.
It is advisable to add identity tags and a contact number to samplers / supports to provide assurance about the legitimacy of the sampling.

Ad 8. Sites should where possible be selected to provide ease of access for deployment and collection.

Ad 9. To aid in obtaining permission from local authorities (if necessary) and property owners, we will prepare a glossy leaflet to describe the project emphasizing the international context. We will make sure that the leaflet is available mid September to all centres performing measurements in the first round.

If local authorities or property owners give permission, give them your contact information and ask them whether they can inform you when there is something wrong with the measurement (pump stopped, broken equipment etc).

3.1.2 Site characterization

From all final selected sites a detailed documentation should be prepared in order to be able to correctly interpret the measurements. From all selected sites a detailed documentation needs to be prepared:

1. A text explaining the rationale for site selection including cohort locations
2. A table that lists the number of sites in each category of site type (regional background, urban background, traffic site and potentially other predictor variables)
3. Provide a digital map with the location of the sites added, for example use Google maps. These maps should include an overview of the complete study area with the site locations as also more detailed maps for each site specifically
4. Provide also a normal paper map with the locations of sites added.
5. Coordinates measured by GPS (this can serve as a check on the geo-coding as well). The coordinates have to be measured accurately and measurements of the coordinates should be repeated at each visit.
6. Description about which coordinate system is used
7. Provide digital photographs such that the immediate surroundings of the site (with the site location also on the photograph) from all four directions can be seen, or make a 360° video of the surroundings of each site.
8. Complete the site characterization form (Appendix 2).

3.1.3 Temporal aspects

Samples of PM will be collected at 20 sites per area. Six units will be made available for each area so that 5 sites can be monitored simultaneously, in addition to the background site where measurements are continuous for a complete year. Samples of NOx will be collected at 40 sites per area.

To limit impact of temporal variation (e.g. meteorological or pollution conditions such as episodes or local traffic disruption), it is important to scatter the measurement locations in each sampling session as much as possible, not only based on logistics. So do not use 5 locations that are located close together within one 2-week measurement period, but make sure that these 5 locations are distributed over the study area in order to limit the impact of temporal variation.
Further, also vary the site types within a 2-week periods, so do not measure only traffic sites in one 2-week period, while in another period only regional background sites are measured.

Each site will be monitored three times, each time for two weeks, once in the cold season (December – February), once in the warm season (June – August) and once in the intermediate spring or fall season (Scandinavian countries will evaluate whether the intermediate season should be the spring (spring dust!) or the fall). The first 2-week measurement period should be finished before Christmas.
Two-week measurement periods in a season do not have to be consecutive weeks.

PM/NOx areas:

PM measurements will be performed in 4 x 2 weeks. NOx sites that are co-located with PM measurements should only measure NOx when also PM measurements are conducted at that location. So in each 2 week measurement period there will be 5 PM/NOx measurements and 5 NOx only measurements. Try as much as possible to have contrast in measurements in the same 2 weeks, so include regional, urban and traffic sites. And also distribute the sites in a 2-week period over the study area (when possible) and do not measure only in one part of the study area. Installation of the 5 PM measurements in a 2-week period should be done within as short time as possible, but at least within 3 days. In that way, for these measurements there is large overlap in time period.

One background site will be monitored during a complete year (starting in October/November) in order to adjust discontinuous measurements obtained at the other sites for temporal variation. As mentioned, each sampling period will be two weeks. Hence PM monitoring will be conducted during 3 (seasons) * 4 (periods needed to cover 20 sites with 5 pieces of equipment) * 2 (weeks of sampling) = 24 weeks distributed over the year that are simultaneous with other measurements. To obtain a true annual average, 14-day average measurements at the background site will also be made in weeks that no other site are measured, including holidays. As the main purpose of the measurements at the site, is adjustment for temporal variation of the measurements made at the 20 sites, sampling periods at the continuous site should exactly coincide with those at the other sites, at the expense of a few days without measurements. An example: The first measurements in a study area start on October 14 2008. Hence measurements start at five monitoring locations and the continuous monitoring site at October 14. October 28 all measurements are stopped. Due to practical reasons, measurements at the next five sites as well as the continuous site start on the 30th of October

NOx only areas:

All 40 measurements should start in the same 2-week period, with the start day of all measurements within a maximum of 3 days. Make sure that you collect your samplers in the same order as you installed them so all sites measure 14 days.

At the reference measurement site it is not necessary to do measurements year-round, provided that a routine monitor at a background location can provide annual average concentrations in your area.

If a PM and/or NOx measurement has failed, the measurement should be repeated in a period when there are no measurements. This is in order to have 3 two-week measurements for each location. This should preferably take place in the same season (for example after finishing a seasonal campaign). The data from the continuous site will be used to adjust for these temporal differences.

If a PM measurement fails, the NOx measurement for this site should also be conducted again (and visa versa). This in order to have for each site PM and NOx measurements for the same period of 2-weeks.

Monitoring periods should not include unusual events such as bonfires and major holidays (i.e. school holidays of a week or longer, this does not count if it is only one free day in a 2-week monitoring period). Each centre should therefore make a planning when to do the measurements. Because there are 24 weeks with measurements plus some time for cleaning and moving etc, i.e. in total ~ 32 weeks in a year; there are 20 weeks in a year without measurements to avoid holidays etc. In order not to complicate the logistics, we will not attempt to perform measurements during the same weeks in the various study areas.
For the areas with both PM and NO_x measurements, in each 2-week time period 5 PM + NO_x sites and 5 NO_x sites should be monitored, of which on 5 sites both NO_x and PM will be measured. Sites in a 2-week period should be distributed over the study area, and should not be spatially clustered.
For the NO_x only areas, it is preferable and possible to measure all 40 locations during the same three periods of two weeks each.

Every effort should be made to complete the entire monitoring campaign within 12 months. With careful planning, this should be feasible, as in TRAPCA measurements in the Netherlands, Munich and Stockholm took 13-14 months with four campaigns of two-week samples.

We will use data from urban / rural background routine monitoring sites to document differences between the two years of measurements. Previous studies have documented that the spatial contrasts will remain more or less constant but the absolute levels may differ related to weather conditions. For each study area, the calculated concentrations will be scaled to year 1, using the ratio of concentrations of PM10, PM2.5, NOx and Black Smoke (if available). If data are not available, the ratio of another related pollutant needs to be used.

Samplers will be operated for two weeks continuously. In order to prevent overloading of the filter, timers will be used to turn the pump on for 15 minutes during each two hours. In each study, the timer will be programmed such that samples are taken from 0-0^{15}, 2-2^{15}, 4-4^{15}, 6-6^{15}, 8-8^{15}, 10-10^{15}, 12-12^{15}, 14-14^{15}, 16-16^{15}, 18-18^{15}, 20-20^{15}, 22-22^{15}. In this way a representative sample can be taken. In 14 days (336 hours) effectively a 42-hour sample (336 / 8) is taken.

3.1.4 PM10 and PM2.5 measurements

Measurements will be made with Harvard impactors following procedures developed in the TRAPCA, Ultra and RUPIOH study. All centres will use the same impactors and pump units supplied by IRAS. Measurements will be done using common SOPs. Excel files will be provided by IRAS which contain the relevant calculations. Two µm pore size Teflon filters are used that are pre- and post weighed. Next reflectance of PM2.5 will be measured. Teflon PM2.5 filters will be analysed by XRF for elemental content. Filters will be stored for future analyses of chemical and in vitro toxicity.

Weighing will be performed at strictly controlled T and R.H. (20^0 C and 35% R.H. respectively) using a microbalance. Detailed QA/QC is included in our weighing SOP (available upon request).

Reflectance of all PM2.5 filters will be measured according to SOPs used before (RUPIOH, TRAPCA, ULTRA). This SOP broadly follows the OECD protocol for Black Smoke measurements. Reflectance measurements have been shown to correlate well with actual measurements of elemental carbon and can be considered as a marker for traffic emissions ('diesel soot') in many areas.
Each filter will be measured at five standard locations and the average will be used in the calculations. Absorbance as defined by ISO 9835 will be calculated from the reflectance.

$a = ((A/2) / V) * \ln (R_0/R_f) * 10^5$

with:

a = absorption coefficient ($m^{-1} * 10^{-5}$)
A = loaded filter area (0.00078 m2)
V = sampled volume (m3)
R0 = average reflectance field blank (%)
Rf = reflectance sampled filter (%)

Reflectance measurements will be conducted after weighing. Consistent with the fact that most elemental carbon is associated with sub micrometer particles, reflectance of PM2.5 and PM10 filters has been shown to be nearly equal and highly correlated. Therefore reflectance measurements will be limited to PM2.5 filters.

Elemental composition

Each exposed Teflon filters for $PM_{2.5}$ or PM_{10}, after weighing and reflectance measurements at Utrecht, will be analyzed with **XRF** for a suite of 35 trace elements (Na through Pb). The XRF analysis will be organized by UniBas, where the filters will be compiled, organized, and shipped to the Chester Lab in the U.S. or Cooper Lab in Canada (bids will be compared). The XRF analysis at the Chester Lab is U.S. EPA certified and has been validated with inter-lab comparisons.

UniBas plans to ship filters to the Chester lab in batches, 3 times a year (20 filters/area*11 areas=220 filters, plus duplicates and blanks), after each monitoring season, to minimize potential filter misplacement and to enforce regular quality control.

Duplicates and blanks

A field blank and field duplicate should be collected. Blanks and duplicates will only be obtained for PM10; we will assume that they are valid for PM2.5 as well. The duplicates will be obtained by attaching an extra

PM10 inlet to the pump unit (each unit can accommodate four sampling inlets). While it is not a 'true' duplicate (because taken with the same pump units), the critical orifices are different, the impactors are different and handling and weighing are different which we feel is good enough. Duplicates measurements should be conducted at the 'continuous measurement' site, so that the duplicate can be considered as a 'backup measurement' as well. One sample should be assigned as the main sample (we will not average main and duplicate sample).

3.1.5 NO_x measurements

NO_x will be measured using Ogawa passive diffusion badges. The analysis is spectrophotometrically based upon the Saltzman method. We will use SOP OGAWA protocol for NO-NOx Sampler. These samplers provide a separate concentration for NO_x and NO_2, and hence NO by subtraction. NO_x will be measured at exactly the same sites as PM10 and PM2.5 during exactly the same time periods. In addition, 20 additional sites will be selected for additional spatial resolution. As these samplers are small and do not require electricity, they can be used at locations which are difficult to monitor for PM. The spatial correlation between NO_x and PM and soot is expected to be high but area-specific (37), so that in the epidemiological analyses, measurement error correction techniques can be used to estimate area-specific exposure response relationships for PM and soot starting from exposure response relationships for NOx.

Duplicates and blanks

A field blank and field duplicate should be collected. Duplicates (and blanks) will be made at the continuous monitoring site. In this way the duplicate also serves as a backup for the primary measurement.

3.1.6 Additional ambient data

To characterize the sampling days meteorological data and data on gaseous air pollutants need to be collected for the entire study period (October 2008 – October 2010). Air pollution data should be obtained from routine monitoring sites in the study city or area (urban and/or rural background). Meteorological data can be obtained from routine stations in or near the study area. Routine data can be used to document differences between the two sampling years. Data on relative humidity and temperature are needed for calculating the sampling rate of the Ogawa badge. Hourly data for the following variables need to be collected for the entire study period (if available):

- Temperature
- Relative humidity
- Barometric pressure
- Wind speed
- Wind direction
- Precipitation
- SO_2
- NO_2

- NO
- O_3
- PM10
- PM2.5
- Black Smoke

3.1.7 Calculation of annual average concentrations

After adjustment for differences in weather conditions using data from the continuous monitoring location(s), the average of the three two-week measurements is calculated and used in further land use regression model development (see section 3.3). Procedures have been described before (Janssen et al. 2001; Hoek et al. 2002). For pregnancy outcome cohorts for which time varying pollution estimates are needed (e.g., for month or trimester of pregnancy), daily data from routine monitoring networks will be used to refine the exposure estimates (see section 3.4).

3.2 Geographical data on air pollution predictor variables needed for land use regression modelling

Air pollution modelling will be done by land use regression and (where possible) dispersion modelling (see section 3.5).
Development of land use regression models will be conducted by collecting potentially important predictor variables for air pollution for the geographical coordinates of each of the monitoring sites using GIS analyses. First, the locations of monitoring sites will be geocoded (section 3.2.1). For each coordinate, values for potential predictor variables will be collected (section 3.2.2), and geographic information system (GIS) data will be collected (section 3.2.3). These variables will then be linked to measurements in a statistical model to develop land use regression models (section 3.3), which can be used to estimate air pollution concentrations at cohort addresses (section 3.4).
Regression modelling will start when air pollution measurements have been conducted. Before the air pollution measurements have been finished, GIS information should already be collected (e.g. information on land use, address density / number of inhabitants) as these variables will be used as predictor variables in the land use regression models.
Evaluation of which GIS-data are available and collection of GIS data should thus already start during the air pollution measurements!

3.2.1 Geocoding

Each of the monitoring site locations will be geocoded using a GPS, i.e. a geographical coordinate will be determined. Use local/national coordinates for GPS-reading because local/national GIS data will be in the same coordinate system. With these coordinates GIS analyses will be conducted and the values for potential

predictor variables will be determined for each of the monitoring locations. Also mark the site locations on a standard paper map.
The coordinates have to be measured accurately and measurements of the coordinates should be repeated at each start of a measurement, i.e. 3 GPS reading are then available for each site. Because large spatial variability of air pollution concentrations occurs within tens of meters from major roads geographical accuracy is important. Geocoding may result in imprecision, depending on the software and the detail of the address information. Always measure the coordinates of a measurement location yourself and do for example not use the coordinates of an existing monitoring monitor with which the ESCAPE measurement is co-located.
Plot geocoded site locations on GIS maps (for example with a road network) and compare these maps with the paper maps with the locations marked.
Not only geographical coordinates of the locations of the monitoring sites are needed, but also the geographical coordinates of the addresses of study participants (home, where possible day care, school, work). If addresses of study participants have not been geocoded yet, this should be done locally (no EU-wide address-coordinate database is available). The developed land use regression model will be used to estimate exposure for these cohort coordinates.
The coordinates should preferably be from the same year as the air pollution concentrations and the potential predictor variables.

3.2.2 Potential predictor variables for land use regression modelling

Potential predictor variables have to be available in GIS datasets in order to perform GIS analyses.
Air pollution regression modelling is often applied in settings in which motorized traffic is an important determinant of small-scale variations in air pollution. Therefore most predictor variables attempt to characterize the impact of road traffic at different spatial scales. There is however no fundamental problem to include other predictors such as shipping emissions or industrial sources. Table 5 lists some of the potential predictor variables that can be applied in modelling spatial variations within cities. Significant predictor variables include various traffic representations, population density, land use, physical geography (e.g. altitude) and climate, but there might be other relevant area-specific predictor variables.
Traffic intensity data, especially those for municipal roads, are often problematic as they are only available for a small number of streets, and mainly on major roads, in many cities.
In the absence of traffic data, or to avoid the problems of acquiring them, several LUR applications have successfully explored the use of the length of specific road types without traffic intensity data.

Table 5. Potential predictor variables that can be applied in land use regression modelling

Variable	Specification	Spatial scale / buffer size
Traffic intensity nearest street	Motor vehicles per day, separated into light, medium-heavy and heavy vehicles	NA
Distance to nearest street	Typically distance of object to centre of the road	NA

Variable	Specification	Spatial scale / buffer size
Traffic intensity buffers	Motor vehicles per day, separated into light, medium-heavy and heavy vehicles in buffers around the sampling point	Circles with radii of 100m, 300m, and 500m around the sampling point
Height	Height above ground	NA
Distance to nearby major road	Typically distance of object to centre of the nearest major road	Within 500 meter
Traffic intensity nearest on major road	Motor vehicles per day, separated into light, medium-heavy and heavy vehicles in buffers around the sampling point	NA
Population density	Population density in buffers around the sampling point	Circles with radii of 300m, 1000m and 3000m around the sampling point
Household density	Household density in buffers around the sampling point	Circles with radii of 300m, 1000m and 3000m around the sampling point
Land use	Land use in buffers around the sampling point (e.g. residential land, industry, urban green)	Circles with radii of 300m, 1000m and 3000m around the sampling point
Altitude	If important altitude differences exist	NA
Meteorology	If important meteorological differences exist (e.g. wind speed, wind direction, temperature)	NA
Region variable	If the study area is large there might be regional variation in air pollution concentrations. These regional variations cannot be taken into account by the other variables.	NA

NA = not applicable

Predictor variables in land use regression models are usually computed for circular zones around each monitoring site, using buffer functions available in GIS. The selection of buffer size is crucial in determining the performance of the model, and the spatial resolution of the estimates. Ideally, buffer sizes should be selected to take account of known dispersion patterns.

Various monitoring studies have shown that the impact of a major road on concentrations of traffic-related air pollutants declines exponentially with distance to the road. Beyond about 100 m from a major urban road, or 500 m from a major freeway, variability is limited. In inner-city areas, however, buildings may cause marked departures from this simple distance-decay pattern. In street canyons, for example, marked accumulation of air pollution may occur, especially against the windward side of buildings, with the result that concentrations may differ substantially from one side of the road to another. At the same time, NO_2 formation in street canyons may be limited by the availability of free oxygen, so that more of the pollution remains in NO form. There is also evidence to suggest that air pollution concentrations fall virtually to

background levels behind a row of uninterrupted buildings. Especially in the compact European urban areas, much of the variation in traffic-related air pollution is therefore extremely local. Use of buffer sizes with radii of more than 500 meters for traffic intensity may thus be misleading, because they would incorporate sources too far removed to have a significant effect. In urban areas the buffer size should however be even smaller (for example a radius of 100 meters).
For variables such as population or address density, and land use the buffer size may however be larger (see Table 5).

In summary, for land use regression development the following data are needed:

- Monitored air pollution data (section 3.1);
- Geographical coordinates of the monitoring sites;

And as potential predictor variables (with varying buffer sizes):

- Digital road networks (by road type);
- Traffic flows and composition (fleet mix);
- Meteorology (wind speed, wind direction, temperature etc);
- Topography (altitude, slope angle);
- Address density / population density in different buffers;
- Land cover and land use in different buffers;
- Variable(s) that indicate(s) region;
- Or other relevant predictor variables that are locally available in GIS.

In order to ensure consistency, facilitate licensing and avoid duplication of effort, core data will be centrally pre-processed and provided where possible. Four data sets will be priorities in this respect. Extractions of all these data will be made available to partners on request.

1. Routine air pollution data. Data from Airbase/EMEP will be extracted for relevant years and locations from the Airbase database, screened for consistency and data quality. These data will not be mixed with purpose-designed measurements, but used for assessment of historic time trends.
2. Land cover data. These data are key inputs to the land use regression models. The data will also be used as the geographic base for all GIS data sets. Data will be derived from the EU CORINE land cover data for 1990 (CORINE90) and 2000 (CORINE 2000). These data will be obtained in vector form, re-projected to an appropriate projection, and cleaned. Analysis will also be undertaken of temporal-correlations in order to interpret changing relationships between different pollutants (e.g. due to changes in emission characteristics).
3. Meteorological data. These data will be used in dispersion modelling and, in many cases, to analyze and interpret temporal variations in monitored concentrations (e.g. from the short-term monitoring campaigns) and to help develop time-specific land use regression models. Data will be obtained from the ECMWF (European Centre for Medium-Range Weather Forecasts) for all study areas, cleaned and matched to the basal, land cover data.

4. Digital road data. These data are essential inputs to land use regression and dispersion models. A high-resolution, Europe-wide database will be constructed, based on the 1:10,000 scale Eurostreets data set supplied by Geodan. The database will be centrally purchased, cleaned, re-projected and matched to the land cover data.

Partners will be free to enhance and replace these data where better data sets are available locally (e.g. to update recent changes in road networks), but in these cases sensitivity analyses will be carried out to assess the effects of using alternative data sources. Other data sets will mainly be sourced and delivered locally – e.g. traffic flows/counts, emissions inventories/factors, topography. Linkage of these data (including attaching traffic counts to the road network) will also be done locally.

Because of the large spatial variation in air pollution concentrations close to roads and in urban areas, geographic precision of GIS databases of potential predictor variables is important and should be evaluated and documented. For example, digital road networks are often idealized representations of the true position of the roads.

The coordinates, air pollution data and predictor data are preferably from the same year or time period.

Retrospective information on changes in land use, road networks and traffic flows will be collected to reconstruct historical trends. Spatial pollution patterns tend to be fairly stable across large urban areas as land use often does not change quickly or abruptly over time (26).

A list / table of proposed potential relevant GIS-predictor variables for each city/area (including the above mentioned centrally provided datasets) should be sent to the exposure assessment working group (e-mail to Rob Beelen, r.m.j.beelen@uu.nl), including for each predictor variable:

- Description of variables
- Year(s) of data availability
- Type of data (vector / raster)
- Resolution
- Proposed buffer sizes (radii per variable)
- Whether a variable is a regional background, urban background or traffic predictor variable
- Accuracy of the datasets

3.2.3 GIS analyses to collect values for potential predictor variables

When all relevant potential GIS-datasets have been collected, buffer analyses will be conducted in order to have predictor values for the different buffer sizes. Values for potential predictor variables will then be collected for each of the coordinates of monitoring sites. GIS analyses will be conducted using ArcInfo / ArcGIS.

3.3 Development of a land use regression model

After the GIS analyses have been conducted, for each monitoring location the following information is available:

- Geographical coordinate
- Average air pollution concentration
- Values for potential predictor variables

The average concentrations and values for potential predictor variables will be used to develop prediction models using stochastic modelling techniques. Stochastic modelling techniques involve developing statistical associations between potential 'predictor variables' and measured pollutant concentrations as a basis for predicting concentrations at unsampled sites. Regression techniques are often used for this purpose and the technique is often called land use regression (LUR) modelling. The developed regression equations are then used to predict concentrations at unsampled sites. This technique was successfully developed in a number of European cities for nitrogen dioxide (Briggs, Collins et al. 1997; Briggs, de Hoogh et al. 2000) and has more recently been applied to other pollutants as part of EU studies (Carr, von Ehrenstein et al. 2002; Brauer, Hoek et al. 2003). These studies have shown that regression models can explain a large part of the spatial variations in air pollution concentrations.
Model development needs to balance the need for local optimization (to reflect source characteristics and environmental conditions in each study area) and for inter-comparability between study areas (to aid collective analysis and interpretation of the results).

Models can be developed for each study area separately, but data from different areas may also be pooled to develop a model based on data from a larger number of monitoring sites.

Some studies use the untransformed concentrations whereas other studies use the logarithm of the concentration (Gilbert et al 2005; Henderson et al. 2007; Ryan et al. 2007; Moore et al. 2007; Madsen et al. 2007; Jerrett et al. 2007) in an attempt to better approximate a normal distribution of the residuals. When a log transformation is used, the interpretation of the model changes from an absolute contribution of variables in the model to a relative change, as illustrated by Gonzales et al. (2005). Within the ESCAPE project untransformed concentrations will be used as these are more readily interpretable. We will carefully address the question of how this will affect normality of residuals and model fit.

Models will be trained and calibrated against local data, but they will also be governed by common protocols. Issues to be covered by these protocols are:

- Rules for developing land use regression models (section 3.3.1) – variable selection, inclusion criteria, model structure;
- Rules for validation (section 3.3.2) - method for model validation, choice of performance measures. Further, standard diagnostic tests for ordinary least squares regression will be applied, and after that a decision will be made which model / scenario will be used to estimate concentrations for the addresses of study participants.

In addition, central supervision and training will be provided for the study teams and calibration of models will be done by comparing the performance of the various models for selected reference areas and data sets (using TRAPCA and UK data).

3.3.1 Land use regression model development

Standard linear regression will be used to develop a LUR model that best predicts the measured concentrations, i.e. a model that maximizes the percentage explained variability (R^2) and minimizes the error (RMSE – Root Mean Square Error).

A supervised forward stepwise procedure will be used to develop a regression model. Predictor variables for the regional, urban and local scale have to be specifically defined. The regression models will then be developed using these a priori defined predictor variables. There is no restriction regarding the number of predictor variables that is used in the final models. We recommend to develop the model while stating in advance the sign of the regression slope of a specific predictor in the model. For example we know that traffic emissions will increase the concentration. Regression models that include a negative slope for traffic intensity variables will therefore not be accepted as the final model. We therefore do not recommend the use of automatic selection procedures, implemented in statistical packages. Both SAS and SPSS can be used to develop regression models using a supervised forward stepwise approach.

Three scenarios will be used:

- Full dataset:
  All possible predictor variables will be evaluated and offered to the model at the same time
- Thematic dataset:
  The predictor variables will be evaluated per spatial scale, i.e. regional variables will be offered first, followed by urban scale variables and then local scale / traffic variables.
- Interpolation and Regression:
  First, concentrations measured at regional background sites will be interpolated using interpolation and the residual concentration at urban and traffic sites will be calculated. Regression models will then be developed to explain these residual concentrations using the urban and traffic predictor variables using the Full dataset-scenario.

These three scenarios will be explained in detail below.

Predictor variables for the regional background, urban background and local/traffic scale have to be specifically defined before model development starts. Monitoring sites are also defined as a regional background, urban background or traffic location.

Before developing the model, make a table with the a priori selected predictor variables per spatial scale, the size of the buffers and the direction of the effect. See example Table 6.

Table 6. Example table with predictor variables per spatial scale, buffer sizes and direction of effect.

Predictor variable	Buffer size (radius)	Direction of effect
Regional spatial scale		
Total built up land	20 km	+
Coordinate variables	NA	NA
Region variable indicating the region of the study area	NA	NA
Urban spatial scale		
High density residential land	100m, 200m, 300m, 500m, 1000m, 3000m	+
Low density residential land	100m, 200m, 300m, 500m, 1000m, 3000m	+
Industry	100m, 200m, 300m, 500m, 1000m, 3000m	+
Port	100m, 200m, 300m, 500m, 1000m, 3000m	+
Urban green	100m, 200m, 300m, 500m, 1000m, 3000m	-
Number of inhabitants	100m, 200m, 300m, 500m, 1000m, 3000m	+
Local spatial scale		
Traffic intensity on nearest road	50m, 100m, 200m, 300m	+

Full dataset scenario

Initially, univariate regression analyses will be conducted for all possible predictor variables so that each predictor variable is regressed against monitored concentrations. The concentrations of all sites will be used. The model with the highest adjusted explained variance (R^2) (and lowest RMSE) is regarded as the 'start model'. To this 'start model' the remaining variables will be added separately, and the effect on the adjusted R^2 and RMSE recorded. The predictor variable with the highest additional increase in adjusted R^2 will be maintained in the model if three criteria are satisfied: (1) the increase in adjusted R^2 is greater than 1%, (2) the coefficient conforms to the pre-specified direction, and (3) the direction of effect for predictors already included in the model does not change. This ensures that models involving counter-intuitive associations be avoided, even if they give a stronger basis for prediction (as indicated by adjusted R^2 value and RMSE). When a variable is included, 'outer rings' of buffers will be calculated for this variable and these variables will also be offered to the model in the next steps. For example, urban green with 500m buffer is included in the model (see Table 6), then outer rings will be calculated by calculating urban green in 1000m buffer minus urban green in 500m buffer, and urban green in 3000m buffer minus urban green in 500m buffer. This will result in 2 new variables which will then also be offered to the model in the next steps.

The addition of variables in a supervised stepwise process will be repeated until there are no remaining predictor variables that add more than 1% to the adjusted R^2 of the previous regression model.

At this stage, the residuals of this 'intermediary model' will be computed and correlations with all remaining suitable predictor variables (including subsequent 'outer rings' of buffers) will be explored. Any variables significantly ($p < 0.05$) correlated with the residuals will then be offered, starting with the most significantly correlated variable, in follow-up regression analyses using the same supervised stepwise approach and inclusion criteria (i.e. adjusted R^2 has to increase by more than 1%, coefficient conforms to pre-specified

direction, and the direction of effect for predictors already included in the model does not change). If these inclusion criteria are not fulfilled, the predictor variable will not be included in the model.
In using adjusted R^2 as an inclusion criterion, some variables may become highly non-significant as additional variables are included in the model. As a final step, therefore, variables with p value >0.2 will be sequentially removed from the model, starting with the least significant, until all predictor variables in the 'final model' have a $p \leq 0.2$.

Thematic dataset scenario

Regression models will be separately developed per spatial scale. The model for the regional background scale will be developed first. The final regional background model will then be the start model for the urban background model, and the final model of both scales will be the start model for the traffic scale. The model development procedure within each spatial scale is the same as described above for the 'Full dataset' scenario, but is described in more detail below.

Regional background scale

Univariate regression analyses will be conducted using only regional background predictor variables. Monitored concentrations of all sites will be used. The model with the highest adjusted explained variance (R^2) (and lowest RMSE) is regarded as the 'start model'. To this 'start model' the remaining regional background variables will be added separately, and the effect on the adjusted R^2 and RMSE recorded. The predictor variable with the highest additional increase in adjusted R^2 will be maintained in the model if three criteria are satisfied: (1) the increase in adjusted R^2 is greater than 1%, (2) the coefficient conforms to the pre-specified direction, and (3) the direction of effect for predictors already included in the model does not change.
When a variable is included, 'outer rings' of buffers will be calculated for this variable and these variables will also be offered to the model in the next steps.
The addition of variables in a supervised stepwise process will be repeated until there are no remaining regional background predictor variables that add more than 1% to the adjusted R^2 of the previous regression model.

Urban background scale

The final regional background model will then be used as the start model for the urban background scale. The model will be further developed using only urban background predictor variables. Monitoring concentrations of all sites will be used. To this start model the urban background variables will be added separately, and the effect on the adjusted R^2 and RMSE recorded. The predictor variable with the highest additional increase in adjusted R^2 will be maintained in the model if three criteria are satisfied: (1) the increase in adjusted R^2 is greater than 1%, (2) the coefficient conforms to the pre-specified direction, and (3) the direction of effect for predictors already included in the model does not change.

When a variable is included, 'outer rings' of buffers will be calculated for this variable and these variables will also be offered to the model in the next steps.
The addition of variables in a supervised stepwise process will be repeated until there are no remaining urban background predictor variables that add more than 1% to the adjusted R^2 of the previous regression model.

Traffic scale

The final combined regional and urban background model will then be used as the start model for the traffic scale. The model will be further developed using only traffic predictor variables. Monitoring concentrations of all sites will be used. To this start model the traffic predictor variables will be added separately, and the effect on the adjusted R^2 and RMSE recorded. The predictor variable with the highest additional increase in adjusted R^2 will be maintained in the model if three criteria are satisfied: (1) the increase in adjusted R^2 is greater than 1%, (2) the coefficient conforms to the pre-specified direction, and (3) the direction of effect for predictors already included in the model does not change.
When a variable is included, 'outer rings' of buffers will be calculated for this variable and these variables will also be offered to the model in the next steps.
The addition of variables in a supervised stepwise process will be repeated until there are no remaining traffic predictor variables that add more than 1% to the adjusted R^2 of the previous regression model.
The resulting model is a model in which regional background, urban background and traffic predictor variables have been used. At this stage, the residuals of this 'intermediary model' will be computed and correlations with all remaining suitable predictor variables (including subsequent 'outer rings' of buffers) will be explored. Any variables significantly ($P < 0.05$) correlated with the residuals will then be offered, starting with the most significantly correlated variable, in follow-up regression analyses using the same supervised stepwise approach and inclusion criteria (i.e. adjusted R^2 has to increase by more than 1%, coefficient conforms to pre-specified direction, and the direction of effect for predictors already included in the model does not change). If these inclusion criteria are not fulfilled, the predictor variable will not be included in the model.
In using adjusted R^2 as an inclusion criterion, some variables may become highly non-significant as additional variables are included in the model. As a final step, therefore, variables with p value >0.2 will be sequentially removed from the model, starting with the least significant, until all predictor variables in the 'final model' have a $p \leq 0.2$.

Interpolation and Regression scenario

Depending on the number of sites in an area, the size of the area and the number of sites per site type, the 'Interpolation and Regression'-scenario can be applied. First, concentrations measured at regional background sites will be interpolated using inverse distance squared weighted interpolation and the residual concentrations at urban background and traffic sites will be calculated (= monitored concentration minus interpolated regional background concentration). Regression models will then be developed to explain these

residual concentrations using the urban background and traffic predictor variables according to the 'Full Dataset' scenario.

Initially, univariate regression analyses will be conducted for all possible urban background and traffic predictor variables so that each predictor variable is regressed against the residual concentrations. The residual concentrations of urban background and traffic sites will be used (so data of regional background sites will not be used for LUR model development). The model with the highest adjusted explained variance (R^2) (and lowest RMSE) is regarded as the 'start model'. To this 'start model' the remaining urban background and traffic variables will be added separately, and the effect on the adjusted R^2 and RMSE recorded. The predictor variable with the highest additional increase in adjusted R^2 will be maintained in the model if three criteria are satisfied: (1) the increase in adjusted R^2 is greater than 1%, (2) the coefficient conforms to the pre-specified direction, and (3) the direction of effect for predictors already included in the model does not change. This ensures that models involving counter-intuitive associations be avoided, even if they give a stronger basis for prediction (as indicated by adjusted R^2 value and RMSE).

When a variable is included, 'outer rings' of buffers will be calculated for this variable and these variables will also be offered to the model in the next steps.

The addition of variables in a supervised stepwise process will be repeated until there are no remaining urban background and traffic predictor variables that add more than 1% to the adjusted R^2 of the previous regression model.

At this stage, the residuals of this 'intermediary model' will be computed and correlations with all remaining suitable predictor variables (including subsequent 'outer rings' of buffers) will be explored. Any variables significantly ($p < 0.05$) correlated with the residuals will then be offered, starting with the most significantly correlated variable, in follow-up regression analyses using the same supervised stepwise approach and inclusion criteria (i.e. adjusted R^2 has to increase by more than 1%, coefficient conforms to pre-specified direction, and the direction of effect for predictors already included in the model does not change). If these inclusion criteria are not fulfilled, the predictor variable will not be included in the model.

In using adjusted R^2 as an inclusion criterion, some variables may become highly non-significant as additional variables are included in the model. As a final step, therefore, variables with p value >0.2 will be sequentially removed from the model, starting with the least significant, until all predictor variables in the 'final model' have a $p \leq 0.2$.

3.3.2 Validation and tests for regression analysis

Model validation is a crucial part of applying land use regression methods. Various approaches have been taken with respect to validation.

One approach is leave-one-out cross-validation, in which a model is developed for n-1 sites and the predicted concentrations are compared with the actually measured concentrations at the left-out site. This procedure is repeated n times and the overall level of fit between the predicted and observed concentrations, across all sites, then computed as a measure of model performance. Usually the structure of the model remains constant for each estimate (Brauer et al. 2003; Hochadel et al. 2006). Performance measures can be

the correlation between measured and estimated concentration, mean difference (and range/SD) between measured and estimated concentration, R^2 value and RMSE-value of regression analysis between measured and estimated concentration, average R^2-value and RMSE-value for the n models.
Another approach is to sub-divide the monitoring sites into a training dataset for model development and a smaller group of sites for model validation (Briggs et al. 1997). This approach requires less intensive computer processing, but may be disadvantaged by the a priori division of sites (e.g. concentrations measured at the training and validation sites may differ, especially when the total number of sites is small). A combination of these approaches may provide a reasonable compromise. For example, a form of grouped jacknife analysis may be used, in which the monitoring sites could be divided into several (n) equal groups, and the analysis repeated n times using a different group for validation such that all sites are used in model development and validation (Briggs et al. in press).

Further, standard diagnostic tests for ordinary least squares regression will be applied to all models:

- Influential observations
  The influence of each observation on the estimates will be measured. Influential observations are those that, according to various criteria, appear to have a large influence on the parameter estimates. Cook's D, which is a measure of influence, will be used to evaluate whether there are influential observations. Cook's D measures the change to the estimates that results from deleting each observation. If there are influential observations, it will be evaluated whether the parameter estimates in the regression model change when this influential observation is excluded from the analyses.
- Heteroscedasticity of the residuals
  A plot will be made of the monitored concentrations and the residuals to evaluate whether there is heteroscedasticity.
- Normality of the residuals
  A test for normality of residuals will be conducted (this test is however not that important).

The most important test of these tests is the test for influential observations.

In addition, ordinary kriging will be conducted on the residuals of the final regression models to evaluate spatial autocorrelation in the residuals. If there is spatial autocorrelation of the residuals the assumption of independence for the residuals is violated. In most land use regression studies, it was observed that residuals of regression models did not exhibit spatial autocorrelation anymore, suggesting that ordinary linear regression is appropriate. Semi-variograms will be made for the residuals which will be used to evaluate whether there is a pattern of spatial autocorrelation in the residuals.

If the tests for ordinary least squares regression are good, and if there is no spatial autocorrelation of the residuals, then the model with the best validation results will be chosen as the model which will be used to predict concentrations at the cohort addresses.

If there is significant spatial autocorrelation of the residuals, then we will use universal kriging methods instead of ordinary least squares regression modelling.
However, whether universal kriging is possible depends on the number of sites in an area, the size of the area and the number of sites per site type. This is because universal kriging will first be conducted for the regional background sites, to estimate the regional background concentrations for the urban background and traffic sites. If there only are only a small number of regional background sites or if the study area is large, universal kriging might not be possible. As predictor variables in the universal kriging methods, the predictor variables of the land use regression model with the best validation results will be used.
First, regional background concentrations will be estimated for the urban background and traffic sites using universal kriging with only the regional background predictor variables of the LUR model with the best validation results. For the urban background and traffic sites the residual concentrations will be calculated (= monitored concentration minus predicted regional background concentration). These residual concentrations will be used as dependent variables in universal kriging methods with urban background and traffic predictor variables as independent variables of the LUR model with the best validation results.
The resulting universal kriging model will then be validated, and compared with the land use regression models, and can be used to estimate concentrations at cohort addresses.
R will be used to conduct universal kriging.

3.4 Exposure assignment using land use regression models

When land use regression models (or universal kriging models) have been developed, it is straightforward to calculate exposure estimates for the cohort addresses. After geo-coding the addresses, the values for the predictor variables in the model are obtained using GIS and entered in the prediction model.
The distribution of the estimated concentrations at the addresses will be explored to assess the predicted values and to evaluate whether there are any extreme predictions. If there are extreme predictions it will be checked why there are extreme predictions (e.g. extreme values for predictor variables). If necessary, such extreme values for predictor variables could be truncated and given the highest value which occurs at one of the monitoring sites for that specific predictor variable.

If residential history is available, this can be done for multiple addresses. In some cohorts, information may be available for home and work / school / day care address, and also for these addresses predictions can be made.

3.5 Dispersion models

In some areas dispersion models have been developed to describe spatial variation. This is especially attractive if historical exposures need to be estimated. We first need an inventory of study areas where models have been developed. During the air pollution measurement periods, it should already be evaluated whether dispersion models are available in the study area and which input data are necessary for the dispersion models.

To use dispersion models, we will also need data on point sources, line and area emissions or emission factors. Data will likely not be available in all study areas to use dispersion modelling. Given the complexity to obtain the required detailed spatially resolved input data, it is unlikely that dispersion models will be developed for ESCAPE specifically.
A disadvantage of using dispersion models is that in all areas different models will be used, which will likely have different validity. Predictions by dispersion models will first be validated with the monitoring data collected in ESCAPE. Based upon the validation we can decide whether dispersion or land use regression models will be used to assign exposures. An attractive option is to use the output of a dispersion model as one of the inputs of a land use regression model (in addition to the other potential predictors mentioned before).

In some countries, maps of background air pollution may exist (e.g. Netherlands, Sweden) that may also be useful. These maps have been developed using a combination of network data and dispersion modelling and are updated each year. We may evaluate the usefulness of these data after comparing with our monitoring data.

3.6 Birth cohorts

While modelled annual average concentrations are sufficient for most applications to epidemiological studies, pregnancy outcome studies require more detailed temporal resolution. In these studies, it is common to express exposure as the average concentration per month or trimester of a specific pregnancy (Slama et al. 2007). The required exposure thus needs to contain a spatial and temporal component. To date this has not been systematically evaluated. One simple option is to develop LUR models using annual average concentrations as is commonly done and then use continuous routine monitoring data to produce a temporally varying component. This approach makes the assumption that the spatial pattern is constant in time.
The spatial exposure estimates are yearly averages that do not allow testing for a higher susceptibility to atmospheric pollutants during a given trimester of pregnancy. To seasonalize the exposure model (i.e. include a temporal component depending on the conception and delivery dates, we applied the temporal observations observed in one background station in Munich operated by the Bavarian Environmental Protection Agency to the exposure estimate . Of the two background stations operating during the study period, one is located 60m away from a busy road, and one is in a location distant from an important source of traffic, which is the one we used to build the temporal component of our model. For NO_2, this was done by averaging the NO_2 daily mean levels over the pregnancy of each woman (of continuous sampling site), by dividing this average by the average NO_2 level during the TRAPCA measurement campaign from 1999-2000, and multiplying the corresponding coefficient by the NO_2 estimate from the TRAPCA II spatial model. Using the same approach, we also estimated trimester-specific exposure variables. Assumption: we assumed that temporal variations in the considered atmospheric pollutants were similar across the metropolitan area. Although reasonable, this assumption is likely to have induced exposure

misclassification, which we believe to be minor compared with that which would exist had temporal variations in air pollution been ignored.

3.7 Historical exposures

A challenge is that for several of the included studies, concentrations need to be estimated for the past. Spatially dense networks of air pollution monitoring stations have not existed in most if not all of our study areas. Historical emission databases of sufficient resolution and quality exist in only a few locations, so that the usability of dispersion modelling is limited as well. Nevertheless, tools exist that allow backward extrapolation:

- In the large majority of the study areas, central site monitoring of air pollution has been in operation for many years so that trends over time can be addressed.
  Correlations between concentrations of different years will be calculated. Previous studies have shown that the correlation between air pollution concentrations of different years is high, even over a period of more than 10 years (Beelen et al, 2007). Trends over time will be estimated using mixed modelling. This trend will then be applied to the estimated concentrations using LUR models, and historical concentrations will be estimated.
- Models to estimate background concentrations of PM and NOx in previous years exist, typically on fairly large spatial scales of no less than 5 x 5 km, these will be used as well.
  These models will be used to evaluate trends over time for the background concentration.
  Correlations between concentrations of different years will be calculated, and trends in concentrations over time will be determined using mixed modelling procedures.
- Historic data on land use, road networks etc. exist as well which allow us to judge whether the spatial ranking of current ambient concentrations has been stable.
  For example, correlations between values between different years will be evaluated and trends over time will be evaluated. For data like road networks, it will be evaluated whether the road network changed over the years, and which changes occurred. A recent study (Beelen et al, 2007, 2008) showed that major roads remained likely in place, and the major roads are the roads that are most important for air pollution exposure.
- Data on residential histories of study participants allow us to back calculate exposures as well.
  If residential history is available exposure will be estimated for the other addresses as well using the methods described above.

Furthermore, many epidemiologic studies selected into this project still have follow-up planned in future years, so that assessment of current exposures will be used in truly prospective analyses in years to come. Finally, we will re-sample the sites used in 1999 in the TRAPCA study, so that we will have direct evidence of the agreement between spatially distributed measurements of PM and NO_2 in three European cities/areas (Munich, Stockholm and the Netherlands) obtained several years apart.

3.8 Residential history and other addresses

We will make an inventory for which cohorts an assessment of residential history can be made, and whether other addresses for cohort members are also available (for example work address, school address day care address). An estimate of air pollution concentrations can be made using the methods described above.

3.9 Noise assessment

The main objectives of the noise exposure activities are:

- Assessment of the current road traffic noise exposure for each individual participating in selected cohort studies, and if relevant as well for rail and/or aircraft noise, using existing noise maps;
- Estimation of the change in noise exposure due to infrastructural changes or moving behaviour of the study subjects after their enrolment in the studies;
- Development of a protocol to ensure the comparability of noise exposure levels across cohorts;
- Implementation of quality assurance and control procedures for the assessment of the individual noise exposure.
- An option might be to measure noise at air pollution monitoring sites. Noise assessment should be coordinated together with air pollution exposure working group.

The current individual exposure to noise of transport sources will be assessed by linking local existing noise maps to the home address of participants in selected cohorts using GIS. The availability of noise maps for a substantial part of the study participants is a requisite for a cohort to be included in this part of the study. To increase power, current noise levels for those participants with missing noise exposure indicators in these cohorts will be estimated if relevant input data is available.

Study participants might have moved or infrastructural works affecting their noise exposure (noise barriers, new roads, change of traffic circulation) might have occurred after their enrolment in the cohort study. An inventory of a selection of possible changes will be made locally. Relevant changes could lead to exclusion of study participants from the statistical analysis or to assess quantitatively the historical exposure if appropriate historical input data is present.

Due to the implementation of EC Directive 2002/49/EC (Assessment and Management of Environmental Noise) member states are obliged to produce maps for transport noise for their larger agglomerations (>250,000 inhabitants), major civil airports (>= 50,000 movements per year), major roads (>= 6 millions vehicles per year) and major railways (>= 60,000 trains per year) before the end of 2007.

The primary indicators for the noise assessment will be the L_{den} and L_{night}; they are widely available since they are defined in the EC Directive. In addition, the $L_{Aeq,24h}$ and L_{day} will be estimated from the L_{den} and L_{night} using local available information.

Under the influence of this EC Directive member states are in the process of replacing their national standards and models for noise to internationally accepted and standardized procedures. The existing noise maps in the study areas are likely to be based on interim methods. After an inventory of available noise maps, the noise models used locally and the availability of noise (input) data, a protocol for the current and

historical noise exposure assessment will be developed to ensure cross-centre comparability in the pooled data-analysis making use of the experiences with the noise assessment in EC sponsored multi-centre studies on adverse health effects of long-term noise exposure (RANCH and HYENA). Appropriate quality assurance and control procedures will be implemented in this protocol. The noise assessment has links with the modelling of the air pollution exposure since the input data (road network, traffic flows, etc.) overlaps. A close collaboration will be established between the team responsible for the noise and for the air pollution assessment.

A data inventory will take place in months 1-6; data acquisition for 5-10 selected cohorts will take place in months 7-18; modelling of noise exposure at the home address will take place in months 19-30 so that the modelled exposure data will be available for epidemiologic analysis simultaneously with the air pollution data.

4 Health endpoints

We will study four 'effect' categories:

- Adverse pregnancy outcome studies, and birth cohort studies of children for outcomes such as asthma and allergy (WP3 Goran Pershagen, Karolinska, Sweden) (section 4.1)
- Cohort studies of respiratory biomarker and morbidity endpoints (WP4 Nino Kuenzli, IMIM, Spain) (section 4.2)
- Cohort studies of cardiovascular biomarker and morbidity endpoints (WP5 Annette Peters, HMGU, Germany) (section 4.3)
- Cohort studies of non-accidental and cause specific mortality, and cancer incidence (WP6 Paolo Vineis, Imperial College, UK) (section 4.4)

In this chapter, a brief description of each work package will be given. .

4.1 Pregnancy outcomes and birth cohorts

Background

In this work package, studies on adverse pregnancy outcomes and birth cohort studies of development of childhood disease such as asthma and allergy will be included. This work package will concentrate on studies that have recruited subjects from the mid 1990s onwards. As the health events to be considered occur at the beginning of life or early in life, relevant exposure time windows are by definition current or recent, which is different from the work packages dealing with middle aged or elderly adults. In addition to assessment of long term exposure to air pollution, studies on pregnancy outcome will also aim to estimate short term exposure during specific periods of foetal life.

Aims of Work Package

We will address the following aims:

1. To quantify the impact of ambient air pollution on adverse pregnancy outcomes.
2. To quantify the impact of ambient air pollution on development of asthma, allergy, lung function and infections in children during the first ten years of life.
3. To quantify the impact of ambient air pollution on cognitive function development in children during the first five years of life.

In several of the study cohorts, genetic information is available and in those we will address gene environment interactions.

Cohort studies in Work Package

Table 7 lists period of recruitment, period of follow-up, N of subjects and types of outcomes under study. The included cohort studies are listed here for each of the 3 Work Package aims.

Aim 1: The relevant studies all provide data on length of gestation, birth weight and length. In addition, detailed information on intrauterine growth and head circumference is provided in some of the studies. The collaboration will address pregnancy outcomes in three birth cohort studies which were started in the 1990s (PIAMA NL, BAMSE Stockholm, LISA/GINI Munich), and in the ABCD cohort in Amsterdam, the Oslo part of the Norwegian Mother and child cohort (MOBA) and the Copenhagen part of the Danish Birth Cohort. Additional studies which will supply information on pregnancy outcomes are INMA (Sabadell, San Sebastian and Aviles), RHEA (Crete), GASPII (Rome), Born in Bradford (Bradford), EDEN (Poitiers and Nancy, APREG (Györ, HU) and KANC (Kaunas, LI).

Aim 2: There are four main birth cohorts which will be included: PIAMA NL, BAMSE SE as well as the LISA/GINI cohorts in Munich and the MAAS study (Manchester) These studies all started in the 1990s and have follow-up data until ages 6-10. Follow-up includes objective markers of sensitisation (IgE, skin prick tests), lung function and bronchial hyper responsiveness. Genetic markers are available on a significant fraction of the studied children.

Aim 3: A cohort study from Amsterdam (ABCD) focuses on developmental issues including cognitive outcomes. Similar variables are available also in the INMA, RHEA and EDEN studies.
WP3 will collaborate with the Taiwan Asthma Birth Cohort (TABC). The study has ~ 1,700 newborns with a large fraction being followed up. Analytic approaches will be harmonized with THSS to replicate and compare findings in the Taiwan environment with partly different air pollution composition and sources.

Table 7. Summary table of studies included in WP3

Study	Recruitment	Follow-up	N of subjects	Outcomes*
ABCD, Amsterdam, NL	2003-2004	2008-2009	8,000	PO,CF
PIAMA, nationwide, NL	1996-1997	1997-2005	3,000	AAI,PO
BAMSE, Stockholm, SE	1994-1996	1994-2006	4,000	AAI,PO
LISA/GINI, Munich, DE	1995-1997	1995-2006	6,500	AAI,PO
INMA, Sabadell, San Sebastian, Aviles, ES	2004-2006	2004-2006	2,000	AAI,PO,CF
APREG, Gyor, HU	2000-2006	2000-2006	1 800	PO
MOBA, Oslo, NO	1999-2008	1999-2008	8,000	PO
MAAS, Manchester, UK	1995-1997	1996-2005	1,000	AAI
RHEA, Heraklion, GR	2007-2008	2008-2009	3,500	AAI,PO,CF
BiB, Bradford, UK	2007-2008	2007-2008	10,000	PO
EDEN, Poitiers, Nancy, FR	2003-2006	2008-2011	2,000	AAI,PO,CF
KANC, Kaunas, LI	2007-2008	2008-2009	8,000	AAI,PO
GASPII, Rome, IT	2003-2004	2003-2008	700	AAI,CF
NBC Copenhagen, DK	1996-2002	1996-2010	20,000	AAI,PO
Total			**78,500**	

* Outcomes under investigation: AAI=asthma, Allergy and Infections, PO=pregnancy outcome, CF=cognitive function

Biomarkers, gene-environment interactions and other environmental exposures

This WP will rely heavily on the use of biomarkers to assess phenotypes, particularly regarding allergy and sensitization. Genetic analyses have already been performed in several of the birth cohorts, including genes involved in inflammation and anti oxidative pathways. These data will be used in combined analyses for assessment of interactions with air pollution exposure, for example in relation to development of asthma and allergy in children. Interactions with other exposures are also of paramount interest, such as with allergens (pollen etc), diet (antioxidants etc) and indoor air pollution (passive smoking etc).

4.2 Chronic obstructive lung diseases in adults

Background

Ambient air pollution contributes to the exacerbation of Chronic Obstructive Pulmonary Disease (COPD) and asthma while effects of pollution on the underlying chronic disease processes that ultimately lead to COPD or onset of asthma in adults is poorly investigated. Smoking is the most important cause of COPD but it is now recognized that this sever and chronic lung disease has other causes as well. Studies such as those participating in WP4 suggest cross-sectional associations of chronic bronchitis symptoms with ambient air pollution. However the role of pollution in the onset of COPD or asthma - the most important respiratory diseases in adults - needs to be elucidated.

This work package describes analyses among five European respiratory health studies that provide at least one baseline and one follow-up assessment of respiratory health, including asthma and lung function data.

WP4 hypothesizes that long-term exposure to traffic-related air pollution is associated with the prevalence of chronic bronchitis symptoms; the prevalence and incidence of COPD, objectively defined with lung function; the level and change of forced expiratory lung volumes, namely FVC and FEV1; the incidence of adult asthma. Moreover, based on current knowledge of the most relevant underlying mechanisms we hypothesize that these associations are modified by endogenous and/or exogenous factors relevant in the defence and/or amplification of oxidative stress and systemic inflammation.

Aims of Work Package

To investigate these hypotheses we will address the following five aims for symptoms, COPD (based on lung function measurements) and doctors' diagnosed asthma:

1. To investigate the effect of ambient air pollution on the prevalence of chronic bronchitis symptoms
2. To investigate the effect of ambient air pollution on the prevalence and incidence of COPD
3. To investigate the effect of ambient air pollution on the level and change of FVC and FEV1
4. To investigate the effect of ambient air pollution on the incidence of asthma

Cohort studies in Work Package

Five geographically diverse European lung health cohort studies have been identified for inclusion. The project combines only studies with comparable health assessment protocols, known residential address, and with the most recent lung function assessments taken not more then five years ago. Moreover, all studies have extensive questionnaire, blood specimen and DNA data to investigate endogenous and exogenous factors that may determine susceptibility to adverse chronic effects of air pollution. Table 8 lists period of recruitment and follow-up, and the number of subjects expected in each outcome category relevant to the main hypotheses. Extensive information on respiratory disease phenotype and potential confounding factors is available from all cohorts. All studies have information on air quality that will be complementary to the state-of-the art assessment of exposure to traffic related pollutants as planned in ESCAPE.

ECRHS and SAPALDIA (www.ecrhs.org/ and www.sapaldia.ch/): In 1990, the European Community Respiratory Health Survey (ECRHS) and the Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) study teams developed, in close collaboration, standardized questionnaires and study protocols to conduct cross-sectional lung health studies across Europe (ECRHS I, age 20-45) and Switzerland (SAPALDIA 1, age 18-60). Both studies followed their subjects up within ~10 years. The two studies provide an unprecedented resource of population-based multi-centre respiratory health research including 27 European cities from the far North to the South. In view of the large number of cities involved in these two studies, ESCAPE is unable to conduct measurements in all; for some cities, we will collect GIS information only and use this to model exposure.

UK 1946 Cohort (http://www.nshd.mrc.ac.uk/): This is one of the oldest birth cohorts worldwide, having followed since birth individuals born in the same week in the UK in 1946. Spirometry data is available at

age 43, age 53 and age 62 (2008) years. Recent respiratory disease research has focused on genetic influences on lung function and relationship with cognitive ability. Key parts of the questionnaire are the same as those used in the other WP4 studies.

SALIA: The Study on the influence of Air pollution on Lung function, Inflammation and Aging (SALIA) started with a cross-sectional study in 1985 with a follow-up going on in 2007/08. The study included women from seven areas in and around the Ruhr region. Air pollution measured within 5 km as well as proximity to traffic correlated with the prevalence of COPD (based on lung function). The finding of an association between air pollution and COPD prevalence in women is the first of its kind, corroborated by similar SALIA findings on mortality.

Table 8. Summary table of studies included in WP4

Study	Recruit-ment	Follow-up	N of subjects	Prevalent COPD a)	Incident COPD a)	Prevalent asthma	Incident asthma
ECRHS	1991-93	2000-02	4,900	490	120	630	300
SAPALDIA	1991	2001-02	6,500	800	660	620	250
UK 1946 Birth Cohort	(1946) 1999	2007/08	3,700	360	90	330	120
SALIA	1985-90	2007-08	2,600/800 b)	230 b)	120 b)	160 b)	100 b)
EGEA	1991-95	2003-06	1,400	15 d)	12 d)	540	190
Total	--	--	**~19,000**	**1,900**	**1,000**	**2,270**	**960**

a) Defined as GOLD≥I. Expected numbers at end of follow-up (known in case of ECRHS, SAPALDIA, EGEA; estimated for the others. Chronic respiratory symptoms prevalence (for aim WP4.1) is ~30-40% higher.

b) N of follow-up with lung function = 800, with questionnaire (asthma data): ~2'600 in 2008

c) Only random control sample qualifies

EGEA (http://ifr69.vjf.inserm.fr/~egeanet/): The Epidemiological Study on the Genetics and Environment of Asthma, bronchial hyper responsiveness and atopy – EGEA – has been initiated in 5 French cities shortly after ECRHS. EGEA used largely the ECRHS questionnaire and other standard procedures employed in ECRHS and SAPALDIA. Asthma phenotypes are a uniquely well defined in EGEA cases while controls were based on a population sample. Promising research on the genetics of asthma has been developed facilitating the investigation of gene-environment interactions. A submitted paper reports significant associations between air pollution and asthma severity in this cohort.
A more detailed description of these studies is shown in Appendix 1.

WP4 will collaborate with the Taiwan Health Screening Study (THSS). The study has >80'000 subjects with a large fraction being followed up. Analytic approaches will be harmonized with THSS to replicate and compare findings in the Taiwan environment with partly different air pollution composition and sources.

Biomarkers, gene environment interactions and other environmental exposures

WP4 studies have biomarker data, in particular genetic information, IgE, C-reactive protein and other markers of inflammation. Three studies are partners in European initiatives for genotyping such as GABRIEL. We anticipate the availability of genotyping data on a large number of subjects to investigate interactions with effects of air pollution, in particular genes involved in oxidative stress defence and pathways of inflammation to explore gene-environment interactions.

4.3 Cardiovascular morbidity in adults

Background

A recent study from the US reported the association between long-term exposure to PM2.5 and incidence of cardiovascular event in nearly 66,000 women from the Women Health Initiative conducted in 36 metropolitan areas of the United States (WHI). The effects reported by the WHI study were larger than those reported earlier for the American Cancer Society Study and suggest that a considerable portion of the burden associated with long-term exposure to PM2.5 is attributable to the development of cardiovascular diseases. However, there is very little direct information on the effects of particulate air pollution on intermediate preclinical phenotypes of cardiovascular diseases.

For the European cohort studies on health effects of long-term air pollution exposures, we will fill the gaps in knowledge concerning the development of susceptibility to the incidence of cardiovascular disease events within European populations. In particular, the extent to which long-term exposure to air pollution induces inflammation and atherosclerosis, the main underlying condition determining the risk of acute coronary artery disease events, is of importance for determining the overall public health impact and for developing prevention strategies for at risk populations if measures for reducing ambient air pollution might be lagging behind.

Aims of Work Package

The overall aim of this work package is to delineate the role of long-term exposures ambient particles in affecting intermediate phenotypes of cardiovascular disease and in increasing the incidence of coronary artery disease and cerebrovascular disease in highly exposed populations. In addition, we aim to integrate the emerging evidence on genetic susceptibility as potential effect modifier of this association. Specifically, we will address the following four aims:

1. To quantify the impact of ambient air pollution on markers of inflammation, and gene-environment interactions.
2. To quantify the impact of ambient air pollution on blood pressure and prevalence of hypertension.
3. To quantify the impact of ambient air pollution on the risk for build-up of preclinical atherosclerosis measured by intima-media thickness (IMT) of the carotid artery.
4. To quantify the impact of ambient air pollution on incident coronary events.

Cohort studies in Work Package

We will build on studies assessing the distribution of cardiovascular disease risk factors with detailed information on cardiovascular risk factors including demographic and life-style characteristics, medical history, medications, and genetics to define susceptibility to long-term exposures by ambient air pollution. Table 9 provides an overview on recruitment and follow-up as well as on how the studies contribute to the four specific aims to be addressed within this work package.

The inflammatory marker C-reactive protein (CRP), nowadays considered an important cardiovascular risk factor, is the main outcome for specific aim WP5.1. It has been assessed with a high-sensitivity assay in Krakow, Poland, Ruhr area and Augsburg, Germany, 8 cities in Switzerland, in Stockholm, Sweden and in the Greater Helsinki Area and Turku, Finland providing in total around 30,000 samples. Additional measurements of inflammatory markers such as interleukin-6 are available in some but not all locations. Studies providing data on blood pressure as well as data on physician-diagnosed and treatment for hypertension will be the basis to address specific aim WP5.2. We will consider for at least 5 locations, which have characterised noise within the cities, the exposure to ambient pollution and noise jointly for disentangling their contributions. Data on standardized blood pressure measurements are available for Krakow, Poland, Ruhr area and Augsburg, Germany, 8 cities in Switzerland, Vorarlberg region, Austria, Copenhagen, Denmark, Stockholm, Sweden, and Helsinki/Turku, Finland, providing a total sample of around 150,000.

Table 9. Summary table of studies included in WP5

Study	Aims [a]	Recruitment	Follow-up	N of subjects	N of coronary events [b]
SIDRIA	4	1994-1995	2008	17,000	390
REGICOR	3	2000	2007	1,500	-
HAPIEE (only CRACOW)	1,2,4	2002-2004	2006-2007	9,200	360
RECALL	1,2,3,4	2000-2003	2006-2008	4,800	90
KORA	1,2,3,4	1994-1995; 1999-2001	2004-2005 2006-2008	7,000	230
SAPALDIA	1,2,4	1991	2002	9,600	350
VHM&PP	2	1995-2000	2005	60,600	-
DHC (Copenhagen)	2,4	1993-1997	2005	40,000	4000
Twin Gene (Stockholm) 60-year olds (Stockholm)	1,2,4 1,2,3,4	2001-2002 1997-1998	2004-2008 2007-2008	2,000 4,300	 600
FINRISK (Greater Helsinki Area and Turku)	1,2,4	1993, 1997, 2002, 2007	2008	10,500	330
			Total N	**166,500**	**6,350**

a) The specific aims that the study mainly contributes to

b) Expected incidence of coronary events

IMT measurements, specific aim WP5.3, are available in Girona, Spain, Ruhr area and Augsburg, Germany, and Stockholm, Sweden resulting in a total sample size of more than 10,000 measurements providing thereby a large database with a North-South-gradient across Europe.
Incidence of coronary and cerebrovascular disease will be assessed based on hospital discharge records and death certificates to address specific aim WP5.4. These data are available for Turin and Rome, Italy, Krakow, Poland, Ruhr area and Augsburg, Germany, 8 cities in Switzerland, Stockholm, Sweden and Helsinki, Finland providing in total over 100,000 subjects. In WP5.4 we will consider noise as well.
Existing phenotype and genotype data will be the basis to address gene-environment interactions in the various sub-WPs. In the studies included under specific aim WP5.1, genetic analyses of polymorphisms within the CRP gene are either performed or planned and will become available by year 4 of project in Krakow, Poland, Ruhr area and Augsburg, Germany, 8 cities in Switzerland, Stockholm, Sweden and Helsinki, Finland providing a total sample of 21,300. Additionally, polymorphisms enhancing oxidative stress such as indicated by Glutathion-S-Transferase M1-Polymorphisms are foreseen to become available in a subset of studies.
A more detailed description of these studies is shown in Appendix 1.

We will collaborate with the Taiwan Health Screening Study (THSS). The study has >80,000 subjects with a large fraction being followed up. Analytic approaches will be harmonized with THSS to compare findings in the Taiwan environment with partly different air pollution composition and sources.

Biomarkers, gene environment interactions and other environmental exposures
As part of the work package, biomarkers will serve as intermediate phenotypes specifically in addressing specific aim WP5.1. Gene-environment interactions will be addressed as part of all WPs. A strategy for enlarging the database will be developed and a common list of genetic markers in candidate genes will be established to inform local decisions on ongoing genotyping efforts within year 1 to 3 of this project. For this work package, the role of the exposure to noise is very important as it potentially confounds the association between long-term exposure to air pollution and cardiovascular disease. We will address this in WP5.2 and WP5.4 because research on noise has consistently shown an association between traffic noise and coronary events and hypertension.

4.4 Mortality and cancer incidence

Background
Estimated effects of particulate air pollution on life expectancy form the largest part of the estimated health impact of air pollution in Europe. This work package will estimate the effect of ambient air pollution on life expectancy using data from several cohorts in Europe. In addition, most of the cohorts have been recruited in areas where Cancer Registries are operating, so that cancer incidence will be also investigated in most of them. At least cancers of the lung and upper aero-digestive tract have been associated with exposure to air

pollution. Cohort studies from different areas of Europe are extremely useful for the assessment of the health impact of air pollution and have been pointed out as of highest priority by various review bodies, including WHO and the European Science Foundation. The ESCAPE project will contribute significantly to resolve open questions concerning the long-term effects of air pollution in Europe. Strengths of the work package studies are, among others, the large number of deaths observed in the cohorts and availability of DNA and haemoglobin adduct data measured in relation to air pollution in a previous study (GENAIR). On the basis of such positive features we will be able to estimate more accurately than in previous studies the burden of mortality, and of incident cancers (particularly of the lung) attributable to air pollution in different European countries.

Aims of Work Package

The aims of this work package are to:

1. To quantify the impact of ambient air pollution on cancer incidence.
2. To quantify the impact of ambient air pollution on cardiovascular disease mortality
3. To quantify the impact of ambient air pollution on respiratory mortality
4. To estimate the risk of death attributable to air pollution under a range of European circumstances.

Approximately 37,000 deaths will occur before the end of 2007 in the 645,800 subjects overall recruited. From all the subjects detailed questionnaires on dietary habits and other potential confounders or effect modifiers have been collected.

Cohort studies in Work Package

Included cohorts are briefly described in Table 10. Selected cohorts within the European Prospective Investigation into Cancer and Nutrition (EPIC) will contribute a total of 312,000 participants from cohorts in the Netherlands (n=38,000), Turin, Varese and Florence, Italy (n=36,000), Umea, Sweden (n=25,000), San Sebastian and Murcia, Spain (n= 14,000), Tromso, Norway (n=37,000), the Athens part of the nationwide Greek EPIC cohort (n=6,000) and the participants in the nationwide cohorts in France (n=72,000) and the UK (n=84,000) who live in areas covered by our intensive exposure assessment. EPIC started in the early or mid-1990s and 8-year follow-up is complete. There are already about 12,000 deaths available for analysis from these cohorts.

Table 10. Summary table of studies included in WP6

Study	Recruitment	Follow-up	N of subjects	N of events
EPIC (selected sub-cohorts)	1991-1998	8 years	312,000	12,000 deaths, 1,500 incident lung cancers
SIDRIA	1994-5	13 years	17,000	560 deaths, 72 lung cancers
KORA	1985-1993	21 years	18,000	2,000 deaths
SAPALDIA	1991	15 years	10,000	600 deaths
VHM&PP	1985-2006	9.6 yrs	185,000	15,500 deaths, 2,000

Study	Recruitment	Follow-up	N of subjects	N of events
				incident lung cancers
HUBRO	2000	7 years	18,000	1,310 deaths
GAZEL	1989		25,000	795 deaths
DCH Copenhagen	1993-1997	8 years	40,000	3,500 deaths, 600 incident lung cancers
SALIA	1985-1993	13-21 years	4,800 women	586 deaths
FINRISK	1992-2007	Max 15 yrs	16,000	300 deaths
Total			**645,800**	**~37,000 deaths, ~4,000 incident lung cancers**

Another 333,800 participants will be included from cohorts established in Rome and Turin, Italy, Augsburg, Germany, Switzerland, Vorarberg, Austria, Oslo, Norway, France, Copenhagen, Denmark, Ruhr area, Germany and Helsinki, Finland. Additional cohort characteristics are shown in table 5, including period of recruitment, period of follow-up and number of events. In total, the work package will include 645,800 subjects with an estimated number of ~37,000 deaths and ~ 4,000 incident lung cancers that occurred in the respective follow-up periods. All study areas will be covered by our intensive exposure assessment.

5 Epidemiological analysis

This chapter provides a draft – still skeleton-like - strategy for epidemiological analyses in the ESCAPE project. Both cohort-specific analyses adjusted for potential confounders and meta-analyses will be conducted. More specific analyses can be developed and used in the different work packages and studies (section 5.1). Both individual-level and area-level confounders will be used (section 5.2), and effect-modification will be evaluated (section 5.3). Spatial autocorrelation will also be evaluated (section 5.4). Linearity of effects will be assessed (section 5.5). To develop the epidemiological analysis strategy in more detail, a working group will be set up (section 5.6).

5.1 Work package specific analyses

Specific analyses will be used for each endpoint and study. The following section describes briefly the general principles of data-analyses.

City-specific analyses with subsequent pooling of the regression coefficients applying methods for meta-analyses are foreseen. This will allow using optimised city-specific confounder models and assessing regional heterogeneity across Europe. Meta-analyses will be conducted by endpoint. Depending on feasibility (e.g. identical protocols or questions) data will be pooled. Regression models will consider repeated measurements.

The specific statistical model will differ between endpoints. For mortality and cancer incidence, we will use Cox proportional hazard models to estimate relative risks in the cohort analyses, after adjustment for relevant confounders (smoking, dietary variables, BMI, alcohol intake, social class indicators, noise) and stratifying by study. For more continuous endpoints, linear regression will be used.

5.2 Confounders

Both individual-level and area-level confounders will be used and will be selected and defined a priori. Confounders are described for each study and can vary for each work package and for each study.
Studies have shown that both individual and area-level (socioeconomic) characteristics affect morbidity and mortality. Studies have therefore used area-level socioeconomic characteristics (such as socioeconomic status at the neighbourhood, or even at larger spatial scales) in addition to individual-level confounders (e.g., Hoek 2002; Beelen 2008, Jerrett, 2005). These area-level variables can for example be assessed using GIS-data: if area-level characteristics are available in GIS an overlay with the coordinates of the home addresses of subjects can be made. For each study area the relevant spatial scale(s) and available area-level characteristics will be evaluated. In the Dutch cohort study (Beelen et al. 2008), the correlation between air pollution exposure and SES differed by spatial scale (neighbourhood versus COROP, a much larger scale).

5.3 Effect-modification

The reanalysis (Krewski et al., 2000) of two US cohort studies (Dockery et al., 1993; Pope et al. 1995) showed that educational attainment significantly modified the air pollution-mortality associations. A recent Dutch study (Beelen et al. 2008) found also suggestive evidence for higher air pollution effects on mortality in subjects with low education, and also in subjects with low fruit consumption. An interaction with smoking status has also been reported (Beelen 2008).
Evaluation will therefore take place whether education, nutrition, smoking status (etc.) modifies the effect of air pollutants. Further possible effect modifiers that will be evaluated include gender. In most of the Work packages gene-environment interactions will also be assessed.
Possible effect modifiers are health outcome specific, and evaluation of other effect modifiers may also be conducted.

5.4 Spatial autocorrelation

We will also evaluate spatial autocorrelation. The Health Effects Institute (HEI)-sponsored re-analysis (Krewski et al. 2000) of the Six Cities (Dockery et al. 1993) and American Cancer Society (ACS) (Pope et al. 1995) cohort studies has raised the question to what extent spatial autocorrelation in health and/or exposure data needs to be taken into account in the analysis of these types of studies. One important difference between the ESCAPE studies and the ACS study is that we will be estimating exposure to air pollution at the address level by using GIS data and stochastic modelling. In comparison, the ACS study was using exposure data which were identical for each subject living in one of the communities that were included. Further, the individual ESCAPE studies have a much smaller spatial scale compared to the HEI re-

analysis of the ACS study which focused on distances of about 600 km to remove spatial auto-correlation. Possibly, spatial autocorrelation over a large area will play a lesser role in the ESCAPE study areas compared with the United States. Spatial autocorrelation at a sub-city level (within-city variation) may be more important in the ESCAPE study areas.
We will evaluate spatial autocorrelation aspects using spatial techniques which have been developed further after the re-analysis of the ACS study by Jerrett and co-workers.
The possible presence of spatial autocorrelation in the original mortality variables does not necessarily imply that complicated statistical models accounting for this need to be applied. It would be preferable that if there is spatial autocorrelation, this could be explained using explicit variables such as area level confounders.

5.5 Linearity of effects

Potential non-linear effects will be evaluated. Potential non-linear relationships between air pollution exposure and confounders with health outcomes will be tested using non-parametric smoothing methods. We may also consider using simple indicator variables, with one of the categories being define by the air quality guideline.

5.6 Statistical analysis working group

A working group that will further discuss and develop the epidemiological analysis strategy in more detail will be set up. This working group will be composed of representatives of the coordinating centre and appropriate members to be recruited from participating centres representing each of the epidemiological work packages.

The task of the group is to harmonize data analysis where necessary. Further specification of the data analysis has to be done in the respective Work Packages. The product is the data analysis manual.

6 Quality assurance / quality control

Quality assurance will include preparation of this study manual, standard operating procedures, training workshops, comparison of equipment and procedures, a site visit by the coordinator's team and a mid-term review.

The **standard operating procedures** will include detailed instructions on the measurement methods that will be applied. We will make use of the SOP's that have been successfully developed in the framework of the Ultra, TRAPCA and RUPIOH studies. The structure of these SOP's will be followed in the current study as well. The SOPs that we will use are shown in Table 11.

Table 11. Standard Operating Procedures (SOP's) to be used in Escape

SOP	Status
NO_x sampling	Manufacturer
PM10 and PM2.5 sampling	To be updated
Weighing of filters	IRAS SOP
Reflectance measurement	IRAS SOP
XRF	Lab SOP
Exposure modelling	To be developed
Epidemiological data analysis	To be developed

Site visits to the various components of the study will be conducted by the coordinator's study team. The site visit will consist of a systematic review of the implementation of the various SOP's. The SOP's will all include instructions for detailed documentation of all information needed to be able to conduct the audits (field forms, laboratory journals, field and laboratory blanks, reproducibility checks etc.). The site visits will commence immediately after fieldwork has started within the separate work packages to detect any problem that may need correction in an early stage of implementation.

A **mid-term review** by independent researchers to evaluate the progress of the study will be planned after 18 months; at this point, the first round of exposure monitoring and modelling will be nearly completed, and the second round will be starting. We plan a further scientific review at month 36, at which point in time we can and should make updated strategic decisions about the final deliverables.

7 Time frame

The overall strategy is to efficiently utilize health & confounder data from European cohort studies. To these studies, air pollution exposure assessment will be applied at the individual home address level of participants in each of these studies. Air pollution exposure assessments will be conducted in the first two years of the study. Epidemiological data will be managed, pooled and made accessible for pertinent analyses simultaneously. Exposure models will be developed concurrently and then subsequently until month 30. Epidemiologic data analyses will be prepared in the first two years of the study. Analyses will start after month 24 until month 42. Reporting of exposure response relationships will be conducted in months 37-48. Health impacts will be assessed in months 37-48. Table 12 shows the time frame for all WPs.
Section 7.1 describes the list of milestones and planning of reviews. Section 7.2 describes the time frame, milestones and deliverables specified for each work package.

Table 12. Overview of the time frame for all Escape work packages

Months		**0**	**6**	**12**	**18**	**24**	**30**	**36**	**42**	**48**
WP1 Coordination										
WP2 Exposure Assessment										
	Tier 1	19 cities/areas								
	Tier 2				19 cities/areas					
	Modelling				Tier 1		Tier 2			
WP3 Birth Cohorts										
	Preparation									
	Analysis									
	Reporting									
WP4 Respiratory										
	Preparation									
	Analysis									
	Reporting									
WP5 Cardiovascular										
	Preparation									
	Analysis									
	Reporting									
WP6 Mortality & Cancer										
	Preparation									
	Analysis									
	Reporting									
WP7 Impact & Dissemination										

Months		0	6	12	18	24	30	36	42	48
Milestones		1.1			2.1	3.1-6 4.1-4 5.1-4 6.1-4	2.2	3.1	3.7-12 4.5-8 5.5-8 6.5-8	4.1
Deliverables		1.1 2.1		2.2				2.3 2.4		3.1-6 4.1-4 5.1-4 6.1-4 7.1-2

7.1 List of milestones and planning of reviews

Milestones are control points where decisions are needed with regard to the next stage of the project. For example, a milestone may occur when a major result has been achieved, if its successful attainment is required for the next phase of work. Another example would be a point when the consortium must decide which of several technologies to adopt for further development. Table 13 lists the milestones of the project.

Table 13. Overview of the milestones of the Escape project

Milestone number	Milestone name	Work package(s) involved	Expected date	Means of verification
M1.1	Kick-off meeting	1,2-8	3	List of action points
M1.2	2nd annual meeting	1,2-8	18	List of action points
M1.3	3d annual meeting	1,2-8	33	List of action points
M1.4	4th annual meeting	1,2-8	47	List of action points
M2.1	First tier monitoring completed	2	15	Samples received
M2.2	Second tier monitoring completed	2	27	Samples received
M3.1	Progress report Intrauterine Growth Paper	3	18	S.C. approval
M3.2	Progress report for Prematurity Paper	3	18	S.C. approval
M3.3	Progress report for Asthma and L.F. Paper	3	18	S.C. approval
M3.4	Progress report for Sensitisation paper	3	18	S.C. approval
M3.5	Progress report for Infection paper	3	18	S.C. approval
M3.6	Progress report for Cognition paper	3	18	S.C. approval
M3.7-12	Analyses for D3.1-6 ready	1,2,3	42	Lead author approval
M4.1	Progress report Bronchitis Prevalence Paper	4	18	S.C. approval
M4.2	Progress report COPD paper	4	18	S.C. approval

Milestone number	Milestone name	Work package(s) involved	Expected date	Means of verification
M4.3	Progress report Lung function paper	4	18	S.C. approval
M4.4	Progress report Asthma paper	4	18	S.C. approval
M4.5-8	Analyses for D4.1-4 ready	1,2,4	42	Lead author approval
M5.1	Progress report inflammation and gene-environment paper	5	18	S.C. approval
M5.2	Progress report hypertension paper	5	18	S.C. approval
M5.3	Progress report Intima Media Thickness paper	5	18	S.C. approval
M5.4	Progress report CVD incidence paper	5	18	S.C. approval
M5.5-8	Analyses for D5.1-4 ready	1,2,5	42	Lead author approval
M6.1	Progress report cancer incidence paper	6	18	S.C. approval
M6.2	Progress report cardiovascular mortality paper	6	18	S.C. approval
M6.3	Progress report respiratory mortality paper	6	18	S.C. approval
M6.4	Progress report attributable risk paper	6	18	S.C. approval
M6.5-8	Analyses for D6.1-4 ready	1,2,6	42	Lead author approval
M7.1	ESCAPE website on air	7	6	Internet access
M7.2	Dissemination strategy developed	7	6	S.C. approval

S.C. = Steering Committee

A mid-term review will be planned after 18 months; at this point, the first round of exposure monitoring and modelling will be nearly completed, and the second round will be starting. Progress reports will be made available detailing the progress towards each deliverable, so that a clear picture will be provided about which deliverables are on schedule, and which ones require further action. Data management for the participating cohorts will have started, and a clear view will be available at that point in time in the project about the contents and dates of the respective deliverables for the full project. We plan a further scientific review at month 36, at which point in time we can and should make updated strategic decisions about the final deliverables.

7.2 Time frame, milestones and deliverables specified for each work package

WP2 Exposure Assessment

Milestones:

M2.1-M2.2: Monitoring completed in tier 1 (month 15) and tier 2 (month 27)

Deliverables:

D2.1: Exposure measurement protocol (month 2)

D2.2: Exposure modelling protocol (month 12)

D2.3-4: Papers: Within and between city variation of ambient air pollution in Europe // Land use regression modelling in 39 cities in Europe (month 40)

WP 3 Pregnancy outcomes and birth cohorts

Milestones:

M3.1-M3.6: Development of the common protocol for local data analyses for each specific aim WP7.1-WP3.6. Month 18

M3.7-M3.12: Results from the local analyses for the specific aim WP3.1-WP3.6 ready. Month 42

Deliverables:

D3.1-D3.6: Reporting of the meta-analyses for specific aims WP3.1-WP3.6. Month 48

WP 4 Chronic obstructive lung diseases in adults

Milestones:

M4.1-M4.4: Development of the common protocol for local data analyses for each specific aim WP4.1-WP4.4. Month 18

M4.5-M4.8: Results from the local analyses for the specific aim WP4.1-WP4.4 ready. Month 42

Deliverables:

D4.1-D4.4: Reporting of the meta-analyses for specific aims WP4.1-WP4.4. Month 48

WP5 Cardiovascular morbidity in adults

Milestones:

M5.1-M5.4: Development of the common protocol for local data analyses for each specific aim WP5.1-WP5.4. Month 18

M5.5-M5.8: Results from the local analyses for the specific aim WP5.1-WP5.4 ready. Month 42

Deliverables:

D5.1-D5.4: Reporting of the meta-analyses for specific aims WP5.1-WP5.4. Month 48

WP6 Mortality and lung cancer incidence

Milestones:

M6.1-M6.4: Development of the common protocol for local data analyses for each specific aim WP6.1-WP6.4. Month 18

M6.5-M6.8: Results from the local analyses for the specific aim WP6.1-WP6.4 ready. Month 42

Deliverables:

D6.1-D6.4: Reporting of the meta-analyses for specific aims WP6.1-WP6.4. Month 48

References

1. Burgess JA, Walters EH, Byrnes GB, Giles GG, Jenkins MA, Abramson MJ, et al. Childhood adiposity predicts adult-onset current asthma in females: a 25-yr prospective study. Eur Respir J 2007;29(4):668-75.
2. Nel A. Atmosphere. Air pollution-related illness: effects of particles. Science 2005;308(5723):804-6.
3. Gauderman WJ, Avol E, Gilliland F, Vora H, Thomas D, Berhane K, et al. The effect of air pollution on lung development from 10 to 18 years of age. N Engl J Med 2004;351(11):1057-67.
4. McConnell R, Berhane K, Yao L, Jerrett M, Lurmann F, Gilliland F, et al. Traffic, susceptibility, and childhood asthma. Environ Health Perspect 2006;114(5):766-72.
5. McConnell R, Berhane K, Gilliland F, Molitor J, Thomas D, Lurmann F, et al. Prospective Study of Air Pollution and Bronchitic Symptoms in Children with Asthma. Am J Respir Crit Care Med 2003;168(7):790-797.
6. Gauderman WJ, Vora H, McConnell R, Berhane K, Gilliland F, Thomas D, et al. Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. Lancet 2007;369(9561):571-7.
7. Sram RJ, Binkova B, Dejmek J, Bobak M. Ambient air pollution and pregnancy outcomes: a review of the literature. Environ Health Perspect 2005;113(4):375-82.
8. Perera FP, Rauh V, Whyatt RM, Tsai WY, Tang D, Diaz D, et al. Effect of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among inner-city children. Environ Health Perspect 2006;114(8):1287-92.
9. Brauer M, Gehring U, Brunekreef B, de Jongste J, Gerritsen J, Rovers M, et al. Traffic-related air pollution and otitis media. Environ Health Perspect 2006;114(9):1414-8.
10. Brauer M, Hoek G, Smit HA, de Jongste JC, Gerritsen J, Postma DS, et al. Air pollution and the development of asthma, allergy and infections in a birth cohort. Eur Respir J 2007.
11. Brauer M, Hoek G, Van Vliet P, Meliefste K, Fischer PH, Wijga A, et al. Air pollution from traffic and the development of respiratory infections and asthmatic and allergic symptoms in children. Am J Respir Crit Care Med 2002;166(8):1092-8.
12. Gehring U, Cyrys J, Sedlmeir G, Brunekreef B, Bellander T, Fischer P, et al. Traffic-related air pollution and respiratory health during the first 2 yrs of life. Eur Respir J 2002;19(4):690-8.
13. Morgenstern V, Zutavern A, Cyrys J, Brockow I, Gehring U, Koletzko S, et al. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. Occup Environ Med 2007;64(1):8-16.
14. de Marco R, Accordini S, Cerveri I, Corsico A, Anto JM, Kunzli N, et al. Incidence of chronic obstructive pulmonary disease in a cohort of young adults according to the presence of chronic cough and phlegm. Am J Respir Crit Care Med 2007;175(1):32-9.

15. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, Stutz EZ, et al. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. Soz Praventivmed 2005;50(4):245-63.
16. Schikowski T, Sugiri D, Ranft U, Gehring U, Heinrich J, Wichmann HE, et al. Long-term air pollution exposure and living close to busy roads are associated with COPD in women. Respir Res 2005;6:152.
17. Sunyer J, Jarvis D, Gotschi T, Garcia-Esteban R, Jacquemin B, Aguilera I, et al. Chronic bronchitis and urban air pollution in an international study. Occup Environ Med 2006;63(12):836-43.
18. Kunzli N, Jerrett M, Mack WJ, Beckerman B, LaBree L, Gilliland F, et al. Ambient air pollution and atherosclerosis in Los Angeles. Environ Health Perspect 2005;113(2):201-6.
19. Dockery DW, Pope CA, 3rd, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med 1993;329(24):1753-9.
20. Pope CA, 3rd, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. Am J Respir Crit Care Med 1995;151(3 Pt 1):669-74.
21. Pope CA, 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama 2002;287(9):1132-41.
22. Laden F, Schwartz J, Speizer FE, Dockery DW. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. Am J Respir Crit Care Med 2006;173(6):667-72.
23. Lipfert FW, Perry HM, Jr., Miller JP, Baty JD, Wyzga RE, Carmody SE. The Washington University-EPRI Veterans' Cohort Mortality Study: preliminary results. Inhal Toxicol 2000;12 Suppl 4:41-73.
24. Enstrom JE. Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. Inhal Toxicol 2005;17(14):803-16.
25. Abbey DE, Nishino N, McDonnell WF, Burchette RJ, Knutsen SF, Lawrence Beeson W, et al. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. Am J Respir Crit Care Med 1999;159(2):373-82.
26. Jerrett M, Burnett RT, Ma R, Pope CA, 3rd, Krewski D, Newbold KB, et al. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 2005;16(6):727-36.
27. Miller KA, Siscovick DS, Sheppard L, Shepherd K, Sullivan JH, Anderson GL, et al. Long-term exposure to air pollution and incidence of cardiovascular events in women. N Engl J Med 2007;356(5):447-58.
28. Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA. Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. Lancet 2002;360(9341):1203-9.

29. Nafstad P, Haheim LL, Wisloff T, Gram F, Oftedal B, Holme I, et al. Urban air pollution and mortality in a cohort of Norwegian men. Environ Health Perspect 2004;112(5):610-5.
30. Filleul L, Rondeau V, Vandentorren S, Le Moual N, Cantagrel A, Annesi-Maesano I, et al. Twenty five year mortality and air pollution: results from the French PAARC survey. Occup Environ Med 2005;62(7):453-60.
31. Gehring U, Heinrich J, Kramer U, Grote V, Hochadel M, Sugiri D, et al. Long-term exposure to ambient air pollution and cardiopulmonary mortality in women. Epidemiology 2006;17(5):545-51.
32. Naess O, Nafstad P, Aamodt G, Claussen B, Rosland P. Relation between concentration of air pollution and cause-specific mortality: four-year exposures to nitrogen dioxide and particulate matter pollutants in 470 neighborhoods in Oslo, Norway. Am J Epidemiol 2007;165(4):435-43.
33. Nyberg F, Gustavsson P, Jarup L, Bellander T, Berglind N, Jakobsson R, et al. Urban air pollution and lung cancer in Stockholm. Epidemiology 2000;11(5):487-95.
34. Nafstad P, Haheim LL, Oftedal B, Gram F, Holme I, Hjermann I, et al. Lung cancer and air pollution: a 27 year follow up of 16 209 Norwegian men. Thorax 2003;58(12):1071-6.
35. Vineis P, Hoek G, Krzyzanowski M, Vigna-Taglianti F, Veglia F, Airoldi L, et al. Air pollution and risk of lung cancer in a prospective study in Europe. Int J Cancer 2006;119(1):169-74.
36. Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer P, Gehring U, et al. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. Epidemiology 2003;14(2):228-39.
37. Lewne M, Cyrys J, Meliefste K, Hoek G, Brauer M, Fischer P, et al. Spatial variation in nitrogen dioxide in three European areas. Sci Total Environ 2004;332(1-3):217-30.
38. Rijnders E, Janssen NA, van Vliet PH, Brunekreef B. Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. Environ Health Perspect 2001;109 Suppl 3:411-7.
39. Hazenkamp-von Arx ME, Gotschi Fellmann T, Oglesby L, Ackermann-Liebrich U, Gislason T, Heinrich J, et al. PM2.5 assessment in 21 European study centers of ECRHS II: Method and first winter results. J Air Waste Manag Assoc 2003;53(5):617-28.

Appendix 1: List of references to cohort studies included in the ESCAPE work packages

60 YEAR OLDS STOCKHOLM: Wandell PE, Wajngot A, deFaire U, Hellenius ML. Increased prevalence of diabetes among immigrants from non-European countries in 60-year-old men and women in Sweden. Diatetes Metab 2007;33:30-36.

ABCD AMSTERDAM NL: van Eijsden M, van der Wal MF, Bonsel GJ. Folic acid knowledge and use in a multi-ethnic pregnancy cohort: the role of language proficiency. Bjog 2006;113(12):1446-51.

BAMSE Stockholm SE: Wickman M, Kull I, Pershagen G, Nordvall SL. The BAMSE project: presentation of a prospective longitudinal birth cohort study. Pediatr Allergy Immunol 2002;13 Suppl 15:11-3.

BORN IN BRADFORD UK: http://www.birthcohorts.net/

COPENHAGEN ADULT COHORT (DIET, CANCER, HEALTH DCS): Tjønneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G, Overvad K. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: A population-based prospective cohort study of 57,053 men and women in Denmark. Scandinavian Journal of Public Health 2007 (in press).

DANISH NATIONAL BIRTH COHORT DK: Olsen J. The Danish National Birth Cohort - a data source for studying preterm birth. Acta Obstet Gynecol Scand 2005;84(6):539-40.

ECRHS multinational: Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. Eur Respir J 1994;7(5):954-60.

Sunyer J, Jarvis D, Gotschi T, Garcia-Esteban R, Jacquemin B, Aguilera I, et al. Chronic bronchitis and urban air pollution in an international study. Occup Environ Med 2006;63(12):836-43.

EDEN birth cohorts FR: Slama R, Sinno-Tellier S, et al. Relation Between Atmospheric Pollutants and Head Circumference in Utero and at Birth: a Cohort Study Relying on Ultrasound Imaging During Pregnancy. Epidemiology 2006;17(6 Suppl.):S129-S130.

EGEA FR http://ifr69.vjf.inserm.fr/~egeanet/ : Dizier MH, Bouzigon E, Guilloud-Bataille M, Betard C, Bousquet J, Charpin D, et al. Genome screen in the French EGEA study: detection of linked regions shared or not shared by allergic rhinitis and asthma. Genes Immun 2005;6(2):95-102.

Kauffmann F, Dizier MH. EGEA (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy)--design issues. EGEA Co-operative Group. Clin Exp Allergy 1995;25 Suppl 2:19-22.

EPIC multinational: Vineis P, Hoek G, Krzyzanowski M, Vigna-Taglianti F, Veglia F, Airoldi L, et al. Air pollution and risk of lung cancer in a prospective study in Europe. Int J Cancer 2006;119(1):169-74.

FINRISK FI: Raatikka VP, Rytkonen M, Nayha S, Hassi J. Prevalence of cold-related complaints, symptoms and injuries in the general population: the FINRISK 2002 cold substudy. Int J Biometeorol 2007 (e-pub ahead of print)

GASPII Rome IT: Porta D, Fantini MP. Prospective cohort studies of newborns in Italy to evaluate the role of environmental and genetic characteristics on common childhood disorders. The Italian J of Pediatrics. 2007 (in press).

GAZEL FR: Goldberg M, Leclerc A, Bonenfant S, Chastang JF, Schmaus A, Kaniewski N, et al. Cohort profile: the GAZEL Cohort Study. Int J Epidemiol 2006 (e-pub ahead of print)

HAPIEE Cracow PL: Peasey A, Bobak M, Kubinova R, Malyutina S, Pajak A, Tamosiunas A, et al. Determinants of cardiovascular disease and other non-communicable diseases in Central and Eastern Europe: rationale and design of the HAPIEE study. BMC Public Health 2006;6:255.

HEINZ NIXDORF RECALL Ruhr Area GE: Hoffmann B, Moebus S, Stang A, Beck EM, Dragano N, Mohlenkamp S, et al. Residence close to high traffic and prevalence of coronary heart disease. Eur Heart J 2006;27(22):2696-702.

HUBRO Oslo NO: Nafstad P, Haheim LL, Wisloff T, Gram F, Oftedal B, Holme I, et al. Urban air pollution and mortality in a cohort of Norwegian men. Environ Health Perspect 2004;112(5):610-5.

INMA Sabadell ES: http://www.birthcohorts.net/

KANC birth cohort Kaunas Lithuania:

KORA Augsburg GE: Scheurig AC, Thorand B, Fischer B, Heier M, Koenig W. Association between the intake of vitamins and trace elements from supplements and C-reactive protein: results of the MONICA/KORA Augsburg study. Eur J Clin Nutr 2007.

LISA/GINI Munich GE: Morgenstern V, Zutavern A, Cyrys J, Brockow I, Gehring U, Koletzko S, et al. Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. Occup Environ Med 2007;64(1):8-16.

MAAS Manchester UK: Custovic A, Simpson A, Woodcock A. Manchester cohort. Pediatr Pulmonol Suppl 2004;26:12-3.

MOBA Oslo NO: Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). Int J Epidemiol 2006;35(5):1146-50.

National Survey of Health and Development (the UK 1946 birth cohort): Wadsworth M, Kuh D, Richards M, Hardy R. Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). International Journal of Epidemiology 2006;35:49-54; Website: http://www.nshd.mrc.ac.uk/

PIAMA 3 regions NL: Brauer M, Hoek G, Smit HA, de Jongste JC, Gerritsen J, Postma DS, et al. Air pollution and the development of asthma, allergy and infections in a birth cohort. Eur Respir J 2007.

Brunekreef B, Smit J, de Jongste J, Neijens H, Gerritsen J, Postma D, et al. The prevention and incidence of asthma and mite allergy (PIAMA) birth cohort study: design and first results. Pediatr Allergy Immunol 2002;13 Suppl 15:55-60.

REGICOR Girona ES: Masia R, Sala J, Rohlfs I, Piulats R, Manresa JM, Marrugat J. [Prevalence of diabetes mellitus in the province of Girona, Spain: the REGICOR study]. Rev Esp Cardiol 2004;57(3):261-4.

RHEA Heraklion GR: http://www.birthcohorts.net/

SAPALDIA 8 communities SU: Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, Stutz EZ, et al. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. Soz Praventivmed 2005;50(4):245-63.

Bayer-Oglesby L, Schindler C, Hazenkamp-von Arx ME, Braun-Fahrlander C, Keidel D, Rapp R, et al. Living near main streets and respiratory symptoms in adults: the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults. Am J Epidemiol 2006;164(12):1190-8.

SIDRIA Rome & Turin IT: Ciccone G, Forastiere F, Agabiti N, Biggeri A, Bisanti L, Chellini E, et al. Road traffic and adverse respiratory effects in children. SIDRIA Collaborative Group. Occup Environ Med 1998;55(11):771-8.

Galassi C, Biggeri A, Ciccone G, Forastiere F. Environment and respiratory diseases in childhood: the Italian experience. Int J Occup Environ Health 2005;11(1):103-6.

TABC Taiwan: Lin YC, Wen HJ, Lee YL, Guo YL. Are maternal psychosocial factors associated with cord immunoglobulin E in addition to family atopic history and mother immunoglobulin E? Clin Exp Allergy, 2004;34;548-554.

THSS Taiwan: Chen T HH, Chiu YH, Luh DL, Yen MF, Wu HM, Chen LS, Tung TH, Huang CC, Chan CC, et al. Community-Based Multiple Screening Model: Design, Implementation, and Analysis of 42,387 Participants Cancer 2004; 1734-1743.

TWIN GENE: Lichtenstein P, deFaire U, Floderus B, Svartengren P, Svedberg P, Pedersen NL. The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. J Int Med 2002;252:184-205.

VHM&pp Vorarlberg region AU: Ulmer H, Kelleher C, Diem G, Concin H. Long-term tracking of cardiovascular risk factors among men and women in a large population-based health system: the Vorarlberg Health Monitoring & Promotion Programme. Eur Heart J 2003;24(11):1004-13.

Appendix 2: Outdoor site characterization form

Identification :			
Site code			
City			
Site address (street + street number)			
Post code			
Contact person			
Other Characteristics of site			
Location :			
Site type			
GPS Coordinates			
Precision of reading			
Type of GPS unit			
Traffic impacts :			
Date			
Starting time of count			
Nearest Street	Light vehicles		
	Heavy vehicles		
	Mopeds		
	Buses		
Nearby “influential” street from which the emissions might be relevant	Light vehicles		
	Heavy vehicles		
	Mopeds		
	Buses		
Street Characteristics :			
Distance to nearest street (m)			
Distance to nearby “influential” street (m)			
Distance to nearest intersection (m)			
Distance to nearest traffic light (m)			
Total width of the street (m)			
Buildings :			
Average height of buildings on sampling side of street (m)			
Average height of buildings on opposite side of street (m)			
Height of applied sampler above street level (m)			
Distance of applied sampler to building façade (m)			
Mainly uninterrupted homes on sampling side of the street		YES or NO	
If no, interruptions more than 25m?		YES or NO	
Mainly uninterrupted homes on opposite side of the street		YES or NO	
If no, interruptions more than 25 m?		YES or NO	
Other impacts within 100 meter :			
Is there a large parking lot (50 or more spaces)		YES or NO	
Is there a small industrial plant e.g. garage, fuel station		YES or NO	
Describe			

Outdoor site characterization plus traffic counts form: extra explanation

IDENTIFICATION

Site codes
All sampling locations have a unique code (see list Appendix 3, sample coding):

- 3 letters for the country and study area (see list Appendix 4, coding list study areas)
- 2 numbers for the site number (01-40)

Site address
The address from which the site is accessible.

Other Characteristics
Possible other characteristics of the site could be any description of where the sampler is attached (e.g. drain pipe, balcony, rooftop.

LOCATION

Site type:
The site type is either an Urban Street, Urban Background or a Regional Background

GPS Coordinates
These coordinates should be measured using the GPS device. Preferably use your National Grid Reference System. Do not use the GPS data from Google Earth! Take a GPS reading every time you install the sampler at the location. Several readings improve precision!

Precision of reading
All GPS devices can tell you the precision of their readings, depending on the reception. This should be within 15m. If it is not, walk

Type of GPS unit
Record the brand, model and type names of your GPS unit.

TRAFFIC IMPACTS

Nearest street
The nearest street to the sampler is not necessarily the address street from which the site is accessible. In case of a side road separate from, but parallel to a major road, this side road is the nearest street. The major road it is parallel to, is the nearest major street. Please give us meters, not yards, feet or inches.

Nearby influential street
The definition of a major street is arbitrary. If there is a road nearby (besides the nearest road) which you would expect to contribute to the pollution levels at the measuring location, also count this street. Mostly these two counts can be done simultaneously. This could be the case when you are sampling at a busy road which has a parallel road right beside it. In this case the parallel road is the nearest, but the busy road next to that is probably influencing the concentration also. Another example is sampling close to an intersection. If you are close to an intersection, emissions from both roads can contribute to the concentration. However, sampling at intersections (especially with major streets) is not recommended for this study.

Traffic counts
Count the traffic, driving in both directions of the street, for 15 minutes, using a stopwatch.
Count only between 10.00 and 15:00 hours (to exclude rush hour) and not on Saturdays or Sundays. Pay attention and document specific situations, for example school children being collected at lunchtime.

Count light vehicles, heavy vehicles and mopeds separately. Do not count buses. The definitions:

- Light vehicles: all cars and motor bikes
- Heavy vehicles: all vehicles with at least 4 rear wheels, which are not line-buses
- Mopeds: all mopeds, scooters and engined bicycles, but not electrically charged vehicles
- Buses: all line-buses

STREET CHARACTERISTICS

Distance to nearest street
This is the distance from the sampler to the side of the street.

Distance to nearby influential street
This is the shortest distance, measured from a bird's perspective, to a major street. This can be measured using Google Earth.

Distance to nearest intersection
This is the shortest distance, measured from a bird's perspective, to the nearest intersection. This can be measured using Google Earth. If there is no intersection within 100 meter's distance of the sampling point, do not measure it.

Distance to nearest traffic light
This is the shortest distance, measured from a bird's perspective, to the nearest traffic light. This can be measured using Google Earth, and should be recorded only if there is a traffic light within 100 meters distance of the measuring site.

Total width of the street
This is the distance from the building façade on one side of the street to the building façade on the opposite side of that street. If there are no building on one or both sides of the street, do not put down any value for the width of the street.

OTHER IMPACTS WITHIN 100 M

Describe
Sites have been selected carefully and are expected to be free of any influence from other sources. Occasionally, there might still be sources that ARE of influence. Add a short description of the neighbourhood source that might influence your measurement, and give an indication of size and distance to this source.

Appendix 3: Sample coding "E S C A P E" project

Coding samples :	**CCGLLCCddmmyyA**	
Country + group		3 digits (3 letters) (see coding list study manual, Appendix 4) [a]
Location nr	01 - 40	2 digits (2 numbers) [a]
Component	P1 ór P2 ór NO ór OX ór PH	2 digits (2 letters ór 1 letter + 1 number)
Date	ddmmyy	6 digits (6 numbers) start date
Additional	S or B or D	1 digit (1 letter) S=sample, B=blank, D=duplicate

a) Also used as location coding on site characterization form

Appendix 4: Coding list Study areas

Study Area	Code	Epidemiologic study acronym	Measurements
Györ, Hungary	HUG	APREG	PM+NOx
Florence, Italy	IFL	EPIC	NOx
Turin, Italy	ITU	EPIC, SIDRIA, ECRHS	PM+NOx
Varese, Italy	IVA	EPIC	NOx
Verona, Italy	IVE	ECRHS	NOx
Pavia, Italy	IPA	ECRHS	NOx
Rome, Italy	IRO	GASPII, SIDRIA	PM+NOx
Barcelona, Spain	SPB	ECRHS, INMA	PM+NOx
San Sebastian, Galdakao, Spain	SPS	EPIC, INMA, ECRHS	NOx
Huelva, Spain	SPH	ECRHS	NOx
Oviedo, Spain	SPO	ECRHS, INMA	NOx
Girona, Spain	SPG	REGICOR	PM+NOx
Athens, Greece	GRA	EPIC	PM+NOx
Heraklion, Greece	GRH	RHEA	PM+NOx
Oxford, Norfolk, Norwich, Ipswich, UK	UKO	EPIC, ECRHS, UK 1946 cohort	PM+NOx
Bradford, UK	UKB	BIB	NOx
Manchester, UK	UKM	MAAS, UK 1946 cohort	PM+NOx
Utrecht, Netherlands	NLU	EPIC	PM+NOx
Amsterdam, Netherlands	NLA	EPIC, ABCD	
Doetinchem, Netherlands	NLD	EPIC	
Maastricht, Netherlands	NLM	EPIC	PM+NOx
Rotterdam, Netherlands	NLR	PIAMA	
Antwerp, Belgium	BAN	ECRHS	
Heidelberg, Germany	GHE	EPIC	NOx
Erfurt, Germany	GER	ECRHS	NOx
Ruhr Area, Germany	GRU	SALIA, RECALL	PM+NOx
Munich, Germany	GMU	LISA + GINI	PM+NOx
Augsburg, Germany	GAU	KORA	
Lugano, Switzerland	SWL	SAPALDIA	PM+NOx
Basel, Switzerland	SWB	SAPALDIA, ECRHS	NOx
Geneva, Switzerland	SWG	SAPALDIA	NOx
Vorarlberg, Austria	AUV	VHM&PP	NOx
Copenhagen, Denmark	DCO	DCH, National Birth Cohort	PM+NOx
Oslo, Norway	NOS	HUBRO, MOBA	PM+NOx
Stockholm, Sweden	SST	BAMSE, TWINGENE, 60 YEAR OLDS	PM+NOx
Umea, Sweden	SUM	ECRHS, EPIC	NOx
Paris, France	FPA	ECRHS, EPIC, GAZEL, EGEA	PM+NOx
Grenoble, France	FGR	ECRHS, EGEA, GAZEL	NOx
Marseille, France	FMA	EPIC, EGEA, GAZEL	NOx
Lyon, France	FLY	EPIC, EGEA, GAZEL	NOx
Nancy, Poitiers, France	FNA	EDEN	NOx *
Helsinki, Turku, Finland	FIH	FINRISK	PM+NOx
Cracow, Poland	POC	HAPIEE	PM+NOx
Kaunas, Lithuania	LIK	KANC	PM+NOx
Taipei, Taiwan	TWP		PM+NOx