Procedure for extrapolation back in time

Version January 25 2012

Preface and summary

This procedure includes a more extended description of the procedure for extrapolation back in time for the estimated concentrations at cohort addresses in order to clarify and harmonize the exposure assessment in all ESCAPE study areas. Use this procedure when extrapolating estimated concentrations back in time for cohort addresses in your study area. This text is a specification of chapter 7 of the exposure manual.

Based on analyses in the Ruhr area, UK, France, The Netherlands and Stockholm it was decided to update the previous procedure (version July 15 2011) with a ratio method in addition to the absolute difference method because spatial contrasts were larger in the past (1990s). The trend in spatial contrast over time however clearly differs per pollutant with more evidence for NO2 and NOx and more mixed evidence for PM10. The few available previous papers have used a ratio method. Trends from routine networks can be difficult to interpret because of changes in configuration and methods. The previous procedure has therefore been adapted by **applying both a ratio and absolute difference method to all pollutants, with a priori preference for effect estimates based upon a ratio method for local pollutants (NO2, NOx, PM2.5 absorbance, and PM coarse) and the absolute difference method for regional pollutants (PM10, PM2.5).** Because of difficulties in interpreting trends from network data, **we propose to consider the ESCAPE exposure data based on the 2008-2010 measurement data (i.e. the non-backextrapolated data) as the main exposure variable (except for the pregnancy outcome cohorts – see below)**. We also propose that back-extrapolation will only be conducted for the baseline addresses and not for follow-up addresses to limit the number of exposure and epidemiological analyses.

Our experience is that the baseline period is not always just one year, but also can have multiple years. The described procedure can be used for all the different baseline options. Because knowledge of the time period of the baseline year/period is needed, communication between the person conducting the back-extrapolation and the person having the cohort information is needed.

Section 1 describes which data should be collected in each study area. Section 2 describes the procedures for back-extrapolating annual averages. Section 3 describes the procedures for back-extrapolating for birth cohort studies for which trimester-specific averages are needed. Section 4 describes that the daily data may also be used in some epidemiological analyses as exposure variable.

If you feel that (parts of) this procedure do not apply to your study area/data please contact us.

1 To be collected air pollution data

A temporal trend will be applied to the modelled ESCAPE concentrations for each baseline address.

In the large majority of the study areas, monitoring of air pollution by routine continuous network sites has been in operation for many years so that trends over time can be addressed. Concentration data have to be collected for all types of sites (regional, urban, traffic) and as far back in time as possible.

Data from a European database based upon AIRBASE are available, in which an assessment has been made of changes in sites and monitoring methods by the European Topic Center on Air pollution (dr Frank de Leeuw). AIRBASE does however not have all sites and is only complete after 2000.

Local centers should therefore also collect reliable routine monitoring data. Check with the health WPs what the baseline period and the follow-up period of cohorts is. We would like to back extrapolate at least to the year of the baseline address of each study. Whether a trend can be estimated over a time period longer than 10 years and to the baseline year is area-specific, and depends whether there have been large changes in road network, emission sources, land use, etc. This should be evaluated at the local level by the exposure group.

Daily air pollution data should be collected, for example because the ESCAPE "annual" average may fall in two different years (e.g. the ESCAPE reference site has been running from November 2008 – November 2009). If available, collect data for all pollutants measured in ESCAPE. Please check that the daily data of one year consist of concentrations of at least 75% of the days in a year (i.e. 75% data capture for annual mean). If data capture is < 75% the site should not be used.

If there is more than one representative background site available, averages per day over these sites will be estimated, if these sites have more than 75% valid daily data in one year.

A careful evaluation of changes in the network needs to be performed, as monitoring techniques may have changed, sites may have changed or (traffic) conditions around specific sites may have changed. When this will not be taken into account a wrong temporal trend can be applied.

If the study area contains a sufficient number of sites (e.g. > 5), some indication of the spatial pattern stability can be obtained by comparing routine monitoring data over time and the ESCAPE predictions versus the routine monitoring data (see below about spatial trends). If a small number of sites exist, trends over time for all site types together and for the different site types separately will be estimated using mixed modelling. Only data from background site(s) will however be used for back extrapolation. If data are available, please assess the urban-rural gradient over time if network design and quality allow this. A further analysis

could be the traffic-background gradient, though this less certain if only few sites are available. These data can be helpful in deciding whether a ratio or absolute difference method works better in your area. Please document this and send that to WP2 coordinator center (Rob Beelen – r.m.j.beelen@uu.nl).

2. Procedure for back extrapolation of annual averages

The procedure is that both <u>an absolute difference and ratio between different time periods will be estimated </u>using data from site(s) that have data for both the baseline period as well as the ESCAPE measurement period, however this does not imply that the time trend needs to be linear. If solid data from multiple routine monitoring sites exist, we may use multiple differences and ratios within the study area, e.g. differentiating between rural and urban background sites (e.g. in large study areas). Typically, the number of sites will not allow that, so the assumption is that the pattern over space is similar in an area.

In summary, the procedures for the absolute difference and ratio procedures are (<u>example calculations are shown in Appendix I</u>):

Absolute difference method:

1. Check what the baseline period is for the cohort(s) in your study area
2. Collect daily air pollution data for routine monitoring site(s) that both cover the year before and after the baseline period as well as the period of the ESCAPE measurements. Data from all <u>background</u> site(s) will be used for extrapolation.
3. Calculate the "annual" average concentration for the routine monitoring site(s) covering the measurement period, so based on exact dates, of the ESCAPE reference site: $C_{routine\text{-}ESCAPE}$
4. Calculate for each study subject the average concentration based on the year before and the year after the recruitment date for each study subject for the routine monitoring site(s): $C_{routine\text{-}baseline}$. In case of trends during the recruitment period of all subjects, this results in different, but gradual changes in the correction.
5. Calculate for each study subject for the routine monitoring site(s) data the absolute difference between the average one year before and one year after the recruitment date and the annual average covering the ESCAPE measurement period: $Diff_{routine} = C_{routine\text{-}baseline} - C_{routine\text{-}ESCAPE}$
6. Calculate for each study subject the back extrapolated concentration ($C_{extrapolate}$) by adding $Diff_{routine}$ to the modelled ESCAPE annual mean concentration (C_{ESCAPE}): $C_{extrapolated} = C_{ESCAPE} + Diff_{routine}$

Ratio method:

The first 4 steps are similar as for the absolute difference method.

1. Check what the baseline period is for the cohort(s) in your study area

2. Collect daily air pollution data for routine monitoring site(s) that both cover the year before and after the baseline period as well as the period of the ESCAPE measurements. Data from all <u>background</u> site(s) will be used for extrapolation.

3. Calculate the "annual" average concentration for the routine monitoring site(s) covering the measurement period, so based on exact dates, of the ESCAPE reference site: $C_{routine-ESCAPE}$

4. Calculate for each study subject the average concentration based on the year before and the year after the recruitment date for each study subject for the routine monitoring site(s): $C_{routine-baseline}$. In case of trends during the recruitment period of all subjects, this results in different, but gradual changes in the correction.

5. Calculate for each study subject for the routine monitoring site(s) data the ratio between the average one year before and one year after the recruitment date and the annual average covering the ESCAPE measurement period: $Ratio_{routine} = C_{routine-baseline}$ / $C_{routine-ESCAPE}$

6. Calculate for each study subject the back extrapolated concentration ($C_{extrapolate}$) by multiplying the modelled ESCAPE annual mean concentration (C_{ESCAPE}) with the ratio: $C_{extrapolated} = C_{ESCAPE} * Ratio_{routine}$

<u>We characterize the baseline based upon exact recruitment date and a year before and after to avoid back-extrapolation to be influenced too much by specific weather circumstances in a specific year. This may become important when a cohort was recruited in e.g. 2 consequent years. The procedure does result in gradual differences when a longer recruitment period occurs, but that is likely real. We used the exact date instead of calendar year averages to avoid jumps in back-extrapolation when recruitment occurs in multiple years.</u>

<u>This procedure will be applied to all pollutants for which ESCAPE models have been developed: NO2, NOx, PM2.5, PM10, PM2.5 absorbance and PM coarse, provided that routine monitoring data are available.</u>
<u>We suggest not to use another pollutant (e.g. PM10 trend from routine monitoring station to assess a PM2.5 trend), because this introduces too much uncertainty due to different trends for different pollutants e.g. NO2 versus NOx and Black Smoke. We furthermore propose to use the back-extrapolated concentration only in sensitivity analysis.</u>

Correlations between the back extrapolated concentrations by both methods will be calculated and evaluated.

The used procedure and assumptions will likely induce exposure misclassification. However, this is believed to be minor compared with that which would exist when temporal variations in air pollution will be ignored

Table 1 lists the pros and cons of the three considered exposure variables: ESCAPE 2008-2010 LUR. ESCAPE back-extrapolated using absolute difference and ESCAPE back-extrapolated using ratio method.

Spatial trend

One of the assumptions is that the temporal variations are similar across a study area. This assumption will be evaluated in some ESCAPE study areas where previous study-specific spatially dense monitoring has taken place. We re-sampled many of the sites used in 1999 in the TRAPCA study (Munich, Stockholm and the Netherlands), so that we will have direct evidence of the agreement between spatially distributed measurements of PM and NO_2 in these three European cities/areas obtained several years apart. A paper about these comparisons has been planned. Further information might come from the SAPALDIA study and ECRHS Spain.

In addition, in areas where multiple routine monitoring sites are available, correlations between the annual means of different years can be calculated to evaluate spatial patterns over time. 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).

Table 1   Overview of methods

Back-extrapolation method	Status	Pro	Con
None, ESCAPE measurements	Main analysis	1. Harmonized measurements 2. For all components of interest, including e.g. PM2.5 absorbance, coarse PM, elements.	1. Absolute levels may be too low for relevant exposure window, affecting location of the exposure response curve 2. Contrast may have been larger in the past if pollution levels were higher in the past, resulting in too high RR estimates
Absolute difference method	Currently in manual, stated as sensitivity analysis	1. Corrects absolute level if there is a trend *	1. Makes estimates dependent on monitoring networks that differ between centers, especially in the past 2. Trends may be misspecified due to (frequent) changes in the network with respect to site, characteristics of the site , see table 2 3. Not possible for components that are not measured in the network 4. Contrast constant (if one trend specified)
Ratio method	Proposed at Barcelona plenary meeting.	1. Corrects absolute level if there is a trend* 2. Corrects for higher contrast in the past. 3. Possible for components not measured in network (PM2.5, PM2.5 absorbance) if we assume that the trend is the same for measured components 4. Used in previous studies (Slama, 2007; Chen, 2010)	1. Point 1 and 2 above 2. May artificially increase contrast. May occur in case of a trend in time that is the same across the area e.g. related to decreases in large-scale pollution. May thus occur for more regional pollutants (PM2.5, PM10). 3. May be imprecise for low concentrations (e.g. close to zero)

*   if one trend is used for the cohort, no effect on RR estimates for linear analysis. Concentration-response function is affected. If different trends are used for subareas of the cohort or cohorts recruited during a long time period, also RR estimates could be affected.

3. Exposure estimation for birth cohort studies

While modelled annual average concentrations are sufficient for most study areas within the ESCAPE project, <u>pregnancy outcome studies in WP3 require more detailed temporal resolution</u>. In pregnancy outcome 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.

A similar procedure as for the back extrapolation of annual averages will be used, with the difference that <u>for each day during a pregnancy period a temporal adjustment will be made</u>. This will be done because for the exposure groups it is unknown for each subject-ID what the corresponding pregnancy period is. By calculating daily temporal adjusted concentrations it is possible for the person conducting the epidemiological analyses to calculate the correct average concentrations per month, trimester or complete pregnancy period (or other time period if needed).

<u>The exposure groups should therefore check with the healths WPs what the complete time period is in which all pregnancies for a study fall. For all the days in that complete time period a temporal adjustment will then be made</u>. This approach makes the assumption that the spatial pattern is constant in time.

Daily air pollution data from one or more <u>background</u> routine monitoring sites are therefore needed. Only sites with more than 75% data capture will be used. As equipment malfunctioning may occur for prolonged periods, it is still possible that trimester averages contain a large fraction of missing values. Therefore an additional criterion is that each trimester specific average is made up of 75% of the trimester. This can only be calculated by the epidemiological analysts having access to birth dates of the cohort. In case of a too large amount of missing values, imputation of daily vales could be used. For procedures contact Rob Beelen (IRAS).

Both absolute difference method and ratio method will be used (<u>example calculations can be found in Appendix II</u>):

Absolute difference method:

1. Check what the time period is which covers the pregnancy periods of all study subjects in your study area
2. Collect daily air pollution data for routine monitoring site(s) covering both the period that subjects were pregnant and the period that ESCAPE measurements were conducted. Data from all available <u>background</u> site(s) will be used for extrapolation.
3. If there are too many missing data in the daily data from the routine monitoring site(s) (i.e. less than 75% data capture in a year), impute the missing data in order to get a complete dataset. Contact us if this applies.

4. Calculate the "annual" average concentration for the routine monitoring site(s) covering the measurement period, so based on exact dates, of the ESCAPE reference site: $C_{\text{routine-ESCAPE}}$

5. Calculate for the routine monitoring site(s) <u>for each day</u> the absolute difference between the daily concentration (C_{daily}) and the annual average covering the ESCAPE measurement period: $\text{Diff}_{\text{routine}} = C_{\text{daily}} - C_{\text{routine-ESCAPE}}$

6. Calculate <u>for each day</u> the back extrapolated concentration ($C_{\text{extrapolated}}$) by adding $\text{Diff}_{\text{routine}}$ to the modelled ESCAPE annual mean concentration (C_{ESCAPE}): $C_{\text{extrapolated}} = C_{\text{ESCAPE}} + \text{Diff}_{\text{routine}}$

7. For each subject the back-extrapolated concentration for each trimester (or other time-period) can then be calculated using the daily back-extrapolated concentrations

Ratio method:

The first 4 steps are similar as for the absolute difference method.

1. Check what the time period is which covers the pregnancy periods of all study subjects in your study area

2. Collect daily air pollution data for routine monitoring site(s) covering both the period that subjects were pregnant and the period that ESCAPE measurements were conducted. Data from all available <u>background</u> site(s) will be used for extrapolation.

3. If there are too many missing data in the daily data from the routine monitoring site(s) (i.e. less than 75% data capture in a year), impute the missing data in order to get a complete dataset. Contact us if this applies.

4. Calculate the "annual" average concentration for the routine monitoring site(s) covering the measurement period, so based on exact dates, of the ESCAPE reference site: $C_{\text{routine-ESCAPE}}$

5. Calculate for the routine monitoring site(s) <u>for each day</u> the ratio between the daily concentration (C_{daily}) and the annual average covering the ESCAPE measurement period: $\text{Ratio}_{\text{routine}} = C_{\text{daily}} / C_{\text{routine-ESCAPE}}$

6. Calculate <u>for each day</u> the back extrapolated concentration by multiplying the ratio with the modelled ESCAPE annual mean concentration for each subject (C_{ESCAPE}): $C_{\text{extrapolated}} = C_{\text{ESCAPE}} * \text{Ratio}_{\text{routine}}$

7. For each subject the back-extrapolated concentration for each trimester (or other time-period) can then be calculated using the daily back-extrapolated concentrations

<u>This procedure will be applied to all pollutants for which ESCAPE models have been developed: NO2, NOx, PM2.5, PM10, PM2.5 absorbance and PM coarse. Here we need to make assumptions for components that are not routinely measured, probably especially PM2.5 and PM2.5 absorbance.</u>

Absolute difference method

If there are no PM2.5 data, we use the difference for PM10 multiplied by the median PM2.5/PM10 ratio for all ESCAPE measurement sites in the study area. So if in step 5 a Diff$_{routine}$ of 20 ug/m3 is obtained for a certain day and the median PM2.5/PM10 ratio is 0.7, we would use 20*0.7 = 14 ug/m3 for that day for PM2.5.

Back-extrapolated PM coarse concentrations can then be calculated from the extrapolated PM10 and PM2.5 concentrations.

For absorbance we will use Black Smoke if this available, using the formula: absorbance (10^{-5} m^{-1}) = 0.37 + 0.10 BS (ug/m3) (Roorda-Knape et al, 1998). If Black smoke has not been measured, we will use the NOx difference from step 5 multiplied by the median ratio of PM2.5 absorbance to NOx from ESCAPE sites.

Ratio method

If there are no PM2.5 data, we use the ratio for PM10 multiplied by the median PM2.5/PM10 ratio for all ESCAPE measurement sites in the study area during ESCAPE. So if in step 5 a Ratio$_{routine}$ of 30/20 is obtained for PM10 for a certain day and the median PM2.5/PM10 ratio is 0.7, we would use a ratio of 30/20*0.7 = 1.05 for that day for PM2.5.

Back-extrapolated PM coarse concentrations can then be calculated from the extrapolated PM10 and PM2.5 concentrations.

For absorbance we will use Black Smoke if this available, using the formula: absorbance (10^{-5} m^{-1}) = 0.37 + 0.10 BS (ug/m3) (Roorda-Knape et al, 1998). If Black smoke has not been measured, we will use the NOx ratio from step 5 multiplied by the median ratio of PM2.5 absorbance to NOx from ESCAPE. sites

Correlations between the back extrapolated concentrations by both methods will be calculated and evaluated.

4. Daily air pollution data

In these procedures we will use daily air pollution data collected for the period between baseline and end of follow-up. In some of the epidemiological analyses in addition to our modelled (long-term) estimates, also daily air pollution data will be used as exposure variable (e.g. for blood pressure analyses). Please provide therefore also the daily air pollution data together with your modelled concentrations to the person who will merge the exposure and health data and /or who will conduct the epidemiological analyses.
This could be the daily data in which the missing values have been imputed.

Appendix I: Example calculations for procedures for back extrapolation of annual averages

Step 1: Check what the baseline period is for the cohort(s) in your study area

Step 2: Collect daily air pollution data for routine monitoring site(s) that both cover the year before and after the baseline period as well as the period of the ESCAPE measurements. Data from all <u>background</u> site(s) will be used for extrapolation. In case of trends during the recruitment period of all subjects, this results in different, but gradual changes in the correction.

Absolute difference method

Subject ID	Baseline date	Step 3: Calculate $C_{routine-ESCAPE}$	Step 4: Calculate $C_{routine-baseline}$ (baseline date +/- 1 year)	Step 5: Calculate $Diff_{routine} = C_{routine-baseline} - C_{routine-ESCAPE}$	Modelled ESCAPE concentration from LUR model (C_{ESCAPE})	Step 6: Calculate $C_{extrapolated} = C_{ESCAPE} + Diff_{routine}$
1	1-2-1998	20	35	35 – 20 = 15	40	40 + 15 = 55
2	5-2-1998	20	36	36 – 20 = 16	35	35 + 16 = 51
3	8-2-1998	20	37	37 – 20 = 17	25	25 + 17 = 42

Ratio method

Subject ID	Baseline date	Step 3: Calculate $C_{routine-ESCAPE}$	Step 4: Calculate $C_{routine-baseline}$ (baseline date +/- 1 year)	Step 5: Calculate $Ratio_{routine} = C_{routine-baseline} / C_{routine-ESCAPE}$	Modelled ESCAPE concentration from LUR model (C_{ESCAPE})	Step 6: Calculate $C_{extrapolated} = C_{ESCAPE} + Diff_{routine}$
1	1-2-1998	20	35	35 / 20	40	40 * 35/20 = 70
2	5-2-1998	20	36	36 / 20	35	35 * 36/20 = 63
3	8-2-1998	20	37	37 / 20	25	25 * 37/20 = 46.3

Appendix II: Example calculations for procedures for back extrapolation for birth cohort studies

Step 1: Check what the time period is which covers the pregnancy periods of all study subjects in your study area

Step 2: Collect daily air pollution data for routine monitoring site(s) covering both the period that subjects were pregnant and the period that ESCAPE measurements were conducted. Data from all available <u>background</u> site(s) will be used for extrapolation.

Step 3: If there are too many missing data in the daily data from the routine monitoring site(s) (i.e. less than 75% data capture in a year), impute the missing data in order to get a complete dataset

Absolute difference method

Day	C_{daily} at routine site(s)	Step 4: Calculate $C_{routine-ESCAPE}$	Step 5: Calculate $Diff_{routine} = C_{daily} - C_{routine-ESCAPE}$	Modelled ESCAPE concentration LUR model (C_{ESCAPE}) for a subject	Step 6: Calculate $C_{extrapolated} = C_{ESCAPE} + Diff_{routine}$
1-1-1998	30	20	$30 - 20 = 10$	45	$45 + 10 = 55$
2-1-1998	32	20	$32 - 20 = 12$	45	$45 + 12 = 57$
3-1-1998	38	20	$38 - 20 = 18$	45	$45 + 18 = 63$

Ratio method

Day	C_{daily} at routine site(s)	Step 4: Calculate $C_{routine-ESCAPE}$	Step 5: Calculate $Ratio_{routine} = C_{daily} / C_{routine-ESCAPE}$	Modelled ESCAPE concentration LUR model (C_{ESCAPE}) for a subject	Step 6: Calculate $C_{extrapolated} = C_{ESCAPE} * Ratio_{routine}$
1-1-1998	30	20	30 / 20	45	$45 * 30/20 = 67.5$
2-1-1998	32	20	32 / 20	45	$45 * 32/20 = 72$
3-1-1998	38	20	38 / 20	45	$45 * 38/20 = 85.5$