APPENDIX AVAILABLE ON REQUEST

Research Report 139

Effects of Long-Term Exposure to Traffic-Related Air Pollution on Respiratory and Cardiovascular Mortality in the Netherlands: The NLCS-AIR Study

Bert Brunekreef et al.

Appendix C. Data Management Air Pollution Concentrations of National Air Quality Monitoring Network

Note: Appendices Available on the Web appear in a different order than in the original Investigators' Report. HEI has not changed these documents. Appendices were relettered as follows:

Appendix C was originally Appendix 1
Appendix D was originally Appendix 2
Appendix E was originally Appendix 3
Appendix F was originally Appendix 4

Correspondence may be addressed to Dr. Bert Brunekreef, Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, Netherlands.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83234701 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

This document was reviewed by the HEI Health Review Committee
but did not undergo the HEI scientific editing and production process.

Appendix 1: Data management air pollution concentrations of National Air Quality monitoring network

The NAQMN data were obtained from the Laboratory of Air Research of the RIVM (Laboratorium voor Luchtonderzoek; LLO). Original measurement data (as 24-h averages) were downloaded from a validated database (RIL+ - database (Computer Information System for Air)). The downloaded data were taken to IRAS on a CD-ROM as txt-files. The obtained air pollution data were transferred to Excel-files. However, data were not placed in the desired format in these Excel-files, i.e. monitoring stations were placed under each other, with the measurement data of different years in different files, while the desired format would be that the monitoring stations would be placed in columns per component and per year. Therefore, the original Excel-files have been changed manually in Excel in the desired way. During this procedure, several checks have been applied to ensure that the procedure would be done correctly, e.g. check if data were pasted at the correct place (i.e. pasted at the correct time period and at the correct monitoring station), control if start- and end-date of monitoring stations were similar with start- and end-date as described on a list obtained from RIVM, with an overview of start- and end-dates of all monitoring stations and which components were measured by each station. Because in the new Excel-files the names of the measurement stations were removed, the monitoring station numbers were controlled. Annual average concentrations per component across all monitoring stations were calculated in the original files and in the new files and these were compared with each other (these annual average concentrations have again been calculated in SAS and compared with the annual averages from the original files, after the new files were imported in SAS). Furthermore, a person who has not conducted the above described procedure checked the new files by at random comparing data in the original and new files.

A few data from the original files were not transferred to the new Excel files:

- Data from station 19001-Czerniawa were not transferred to the new files (this station is a monitoring station in Poland).
- PM_{10}-measurement data from station 2-Bilthoven-van Leeuwenhoeklaan were not transferred to the new files, because this station only had PM_{10}-measurements on 2 days in 1996 (December 4 and 5) over the period 1976-1996.
- CO-data from station 934-Kollumerwaard-Hooge Zuidwal were not transferred since there was only one month of measurements in the total period 1976-1996 (December 1996).

During data management, a few possible mistakes and questionable measurement values were found in the original files. Furthermore, in SAS the 24-h averages were checked per station and per year using boxplots and histograms and a few questionable or 'out of bounds' measurement values were discovered in the original files. These questionable measurement values were checked with the Laboratory of Air Research and eventually corrected in SAS. The complete list of possible mistakes and questionable measurement values found in the original database is described below:

- Small negative concentrations are present, because of the sensitivity of the measurement equipment.
- In the original file, monitoring station 540-Leiduin had remarkable data (>1200 $\mu g/m^3$) for SO_2 during the period January 1986 – March 1986. Furthermore, according to the overview list, SO_2 measurements at this station should not have begun before December 3 1986. It has been confirmed by RIVM that these data are no valid data and these data were removed in SAS.
- According to the overview list, monitoring station 618-Bilthoven-s should only have measured SO_2. However, starting in 1980, also NO_2 and NO measurement data were found in the original files for this station. It has been confirmed by RIVM that these NO_2 and NO measurement values are not measured by station 618-Bilthoven-s, and these NO_2 and NO data were removed in SAS.
- A lot of stations have '-99' as measurement value on some days for Black Smoke (BS) measurements. Furthermore, station 435-Naaldwijk has '-77' as measurement value for BS on a few days, and station 540-Leiduin has '-88' as a measurement value for BS on a few days. These measurement values are indicator values for missing data (confirmed by RIVM) and were removed in SAS.
- Stations 621-Cabauw (100m) and 622-Cabauw (200m) have the same x,y-coordinates as station 620-Cabauw-Zijdeweg, but measurements on stations 621 and 622 were conducted on a height of 100m and 200m respectively. The stations 621 and 622 were not be used in for imputing of values for missing values and were removed in SAS.
- At February 28 1992, the station number of monitoring station 435-Naaldwijk changed into 443, but the name and x,y-coordinate of the station remained the same (the station was moved over a small distance, but because the x,y-coordinate defines a square of 100m x 100m, the x,y-coordinate remained the same). Reason for this small change in location was that station 435 was located too close to an air pollution source and the station was therefore

not a spatially representative regional station. Results of station 435 were therefore not used in the imputing of values for missing values and were removed in SAS.

- The name and station number of station 517-Haarlem changed in 1989 in station 537-Haarlem-Amsterdamsevaart, but the x,y-coordinates of these stations are the same (the station was moved over a small distance, but because the x,y-coordinate defines a square of 100m x 100m, the x,y-coordinate remained the same). Reason for this small change in location was that station 517 was located too close to an air pollution source and the station was therefore not a spatially representative regional station. Results of station 517 were therefore not used in the imputing of values for missing values and were removed in SAS.

Per component, two new excel files were made for the periods 1976-1986 and 1987-1996 respectively. In each of these new files, the monitoring stations were placed in columns and the days in rows. A schematic overview of the management of NAQMN data is shown in Figure 1.

Imputing values for missing values and calculation of average concentrations

The new excel files were imported in SAS. Because of the modernization of the monitoring network in 1985-1986 and the start of the modernized network on April 1 1986, data of the year 1986 were transferred from the file for the period 1976-1986 to the file for the period 1987-1996 in SAS. The period 1976-1985 is called the old network and the period 1986-1996 is called the new network. The mistakes described in chapter 3 were corrected in these new SAS-files in SAS. With SAS, values for missing values were imputed. Missing values were estimated to prevent bias in the comparisons across sites. In figure 2, the procedure for imputing of missing values and calculation of annual average concentrations for each station is described. In this section, the procedure for the imputing of values for missing values is more extensively described together with the choices for the different criteria to impute values for missing values.

First, the total number of measurement years for each monitoring station in the total period and in the periods 1976-1985 and 1986-1996 was calculated. This was done in order to evaluate which stations only measured a limited period of time over the total period. In the main text tables 1 and 2 the number of stations per component that measured 50%, 67%, and 80% of the total measurement time (1976-1996), of the time in the old network (1976-1985), and of the time in the new network (1986-1996), respectively, are shown. In table 1 is shown that only a small number of measurement stations measured more than 50% of the total measurement time (1976-1996). This is largely due to the revision of the monitoring network in 1985-1986. Because of the small number of stations that measured more than 50% of the total measurement time and because the

monitoring stations before and after the revision of 1985-1986 largely differed, it was more suitable to split up the total measurement period in 2 time periods, i.e. before the revision of 1985-1986: old network (1976-1985) and after the 1985-1986 revision: new network (1986-1996). In table 2 is shown that a large part of the stations in the old network measured at least 67% of the total measurement time (1976-1985). However, in the new network, compared with the old network, a much smaller percentage of the total stations measured at least 67% of the total measurement time (1986-1996). Therefore, as a first criterion, the minimum number of measurement years per station is defined as 50% of the maximum number of measurement years per period. Stations with less measurement years than 50% of the maximum number of measurement years per period were removed. This percentage was used for both the old and new network. Removing of stations that measured less than 50% of the measurement time in a period is done to keep a consistent set of stations in each period that all measured a considerable amount of time.

For imputing values for missing values, the averages of daily measurements across all the monitoring stations was calculated for each day. Before this was done, measurement data of street stations were transferred to a new file, and these data were not used in the calculation of averages of daily measurements, because the measurement data of street stations could have a relatively high impact on the calculated averages of daily measurements. However, missing values for street stations were imputed based on the calculated averages of daily measurements without data of street stations. Averages of daily measurements for each day were be calculated based on data of regional and city stations.

If an average of daily measurements was based on a number of stations that is lower than a predefined number of stations, the average of daily measurements was considered missing. This number of stations is based on the maximum number of stations with measurements on a day for each year. Criterion 2 is that the average of daily measurements has to be composed of at least 50% of this maximum number of stations with measurements on a day for that specific year; Otherwise, the average daily concentration was set to missing.

After this, ratios were calculated by dividing the measurement data by their corresponding averages of daily measurements. Per monitoring station and year, average ratios were calculated for each season, because there is significance variance between seasonal ratios (season 1: April – September; season 2: October – March) within some monitoring stations. Seasonal average ratios have to be composed of a minimum of 92 ratios (50% of 6 months), if lower than 92 ratios, the seasonal average ratio was set to missing (criterion 3).

The next step was the imputing of missing values. Values for missing values were calculated per monitoring station by multiplying the seasonal average ratios with the corresponding averages of daily measurements.
The correlation between the unadjusted and adjusted (i.e. with imputed missing values) annual average concentrations per monitoring stations was calculated.
Annual average concentrations per station and the number of values (i.e. days) over which the average concentrations have been calculated, were calculated. With these data, average concentrations over more than one year were calculated. The annual average concentrations for each station have to be composed of at least 292 values (80% of 365 days); otherwise, annual average concentration was be set to missing.

The spatial representativeness of a regional monitoring station may be influenced by nearby local sources, for example nearby industries. For example, station 234-Putte and station 315-Sas van Gent monitored the influence of the industries of Antwerpen and Gent, respectively, on the air pollution measurements. Station 415-Maassluis was used to characterize the (higher) air pollution background levels in the Rijnmond-area.(2) The spatial representativeness of some regional stations has been questioned in a report by Buijsman. The questioned monitoring stations were the stations 107-Posterholt (possibly influence of nearby roads), 412-Hoek van Holland (possibly other local sources), 903-Delfzijl (possibly local sources in industry-area of Delfzijl), 419-Rockanje (station is located near entrance of a campsite), and 627-Bilthoven (changes in surrounding of location).(2) If possible, it has to be evaluated whether such local sources influence the spatial representativeness of regional monitoring stations.

Figure 1: A schematic overview of the management of NAQMN data

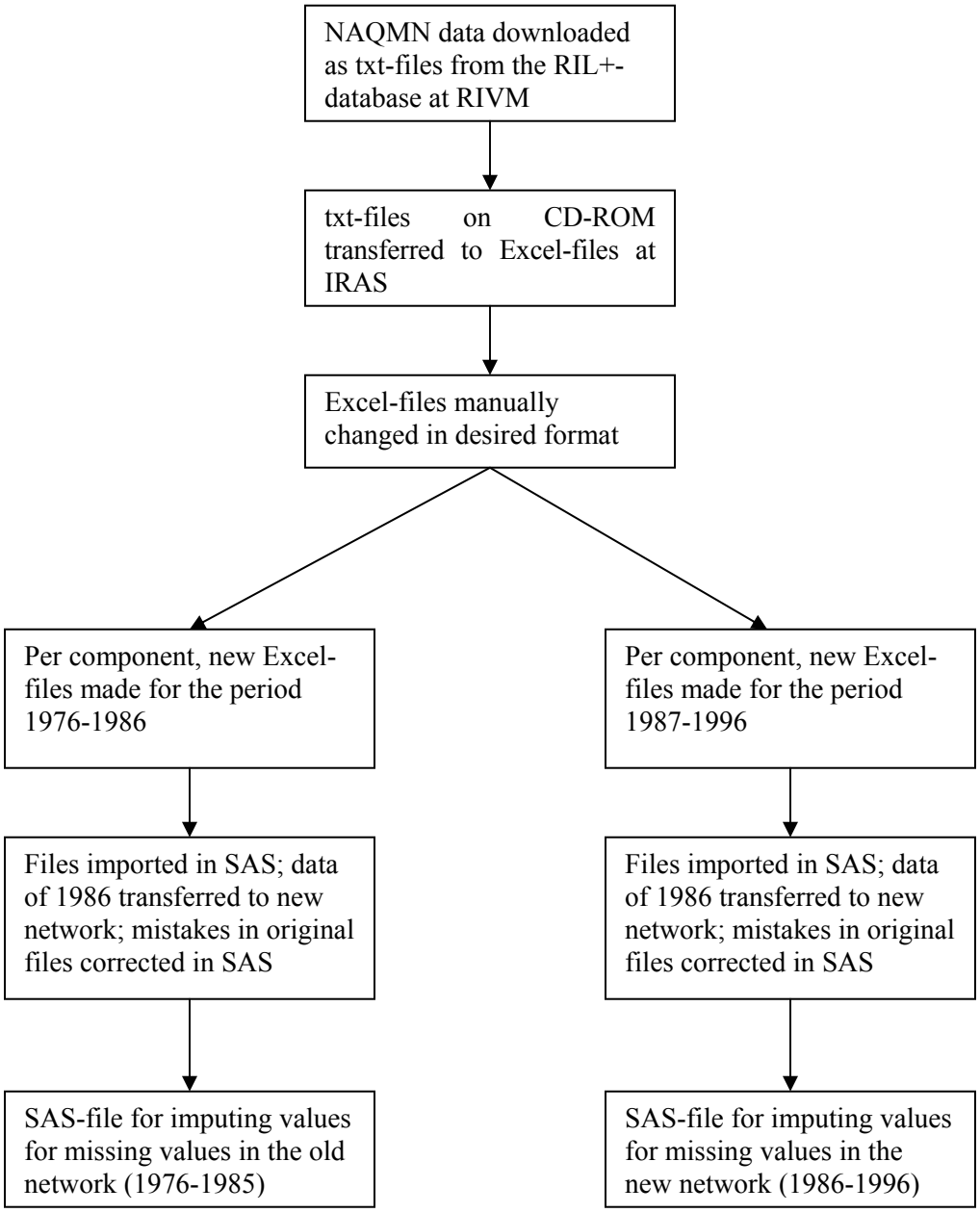

Figure 2: Procedure for imputing values for missing values and calculation of annual average concentrations for each station

1. *Criterion 1:* Monitoring stations that do not have measurement data for at least 50% of the maximum measurement time per period (old network and new network separately), were deleted. This was done to keep a consistent set of stations that all measured a considerable amount of time.

2. Classification of monitoring stations in regional, city, and street stations. Measurement data of street stations were transferred to a new file. These street stations were not used in step 3, but missing values for street stations were also imputed using the averages of daily measurements across all monitoring stations which are calculated without measurement data of street stations.

3. Calculation of averages of daily measurements across all monitoring stations (excluding street stations). *Criterion 2:* For each year, the maximum number of monitoring stations with measurements on a day, was calculated. An average of daily measurements has to be composed of data of at least 50% of this maximum number of stations with measurements on a day for that specific year. Otherwise, the average daily concentration was set to missing.

4. Calculation of ratios: measurement values divided by their corresponding averages of daily measurements (also done for street stations by using the averages of daily measurements calculated without measurement data of street stations).

5. Calculation of average ratios. Because there is significance variance between seasonal ratios (season 1: April – September; season 2: October – March) within some monitoring stations, seasonal average ratios were calculated per year. *Criterion 3:* Seasonal average ratios per year have to be composed of a minimum of 92 ratios (50% of 6 months). If the number of ratios is lower than 92 ratios, the seasonal average ratio was set to missing.

6. Imputing of missing values. Values for missing values were calculated per monitoring station by multiplying the seasonal average ratios with the corresponding averages of daily measurements across all monitoring stations.

7. Calculation of annual average concentrations per station and calculation of the number of values (i.e. days) over which the average concentrations have been calculated (with these data, average concentrations over more than one year can be calculated). *Criterion 4:* annual average concentrations for each station have to be composed of at least 292 values (80% of 365 days), otherwise the annual average concentration was set to missing.

Figure 3: Configuration of NAQMN on January 1, 1994

● and ○: Regional background stations

▽: City stations

References appendici

(1981). Noise and environment (Geluid en omgeving). Alphen aan de Rijn, Samson.

Abbey, D. E., N. Nishino, et al. (1999). "Long-term inhalable particles and other air pollutants related
to mortality in nonsmokers." Am.J.Respir.Crit Care Med. 159(2): 373-382.

Backlund, E., P. D. Sorlie, et al. (1999). "A comparison of the relationships of education and income with mortality: the National Longitudinal Mortality Study." Soc Sci Med 49(10): 1373-84.

Bethlehem, M., J. Nelissen-Braeke, et al. (2000). Handleiding voor het invoerprogramma voor de vragenlijsten op de PC. Report NLCS 90.02 (Manual data-entry questionnaires). Maastricht, Maastricht University.

Bloemen, H., L. v. Bree, et al. (1998). Fine particles in the Netherlands. Report 650010006. (in Dutch with English summary). Biltoven, National Institute of Public Health and the Environment.

Bosma, H., H. D. van de Mheen, et al. (2001). "Neighborhood socioeconomic status and all-cause mortality." Am.J Epidemiol. 153(4): 363-371.

Brauer, M., G. Hoek, et al. (2003). "Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems." Epidemiology 14(2): 228-39.

Breeijen, H. d., R. Bausch-Goldbohm, et al. (1990). Procedures voor de verwerking, cleaning en nutrientberekening van de voedingsgegevens uit de vragenlijst (Procedures for the handling, cleaning and nutrient calculation of dietary data from the questionnaire). Zeist, Instituut CIVO - Toxicologie en Voeding TNO Zeist.

Briggs, D., S. Collins, et al. (1997). "Mapping urban air pollution using GIS: a regression-based approach." Int J Geographical Information Science 11(7): 699-718.

Buijsman, E. (1994). De herziening van het Landelijk Meetnet Luchtkwaliteit in 1993/94 (The revision of the National Air Quality Monitoring Network). Bilthoven, RIVM.

Chow, J. (1995). "Measurement methods to determine compliance with ambient air quality standards for suspended particles." J Air Waste Manag Assoc 45: 320-382.

Collins, S. (1998). Modelling spatial variations in air quality using GIS. GIS and Health. A. Gatrell and M. Loytonen. London, Taylor & Francis.

Cyrys, J., J. Heinrich, et al. (2003). "Comparison between different traffic-related particle indicators: elemental carbon (EC), PM2.5 mass, and absorbance." J Expo Anal Environ Epidemiol 13(2): 134-43.

Dassen, A., J. Dolmans, et al. (2000). Noise in the 5th Environmental assessment (Geluid in de vijfde Milieuverkenning). Biltoven, RIVM.

Dassen, A., J. Jabben, et al. (2001). Development and use of EMPARA: a model for analyzing the extent and the effects of local environmental problems in the Netherlands. Den Haag.

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

Edwards, J., J. Ogren, et al. (1983). "Particulate air pollutants: a comparison of British "Smoke" with optical absorption coefficient and elemental carbon concentration." Atmos Environ 17: 2337-2341.

Elskamp, H. (1989). National Air Quality Monitoring Network - Technical description. Bilthoven, RIVM.

Filleul, L., V. Rondeau, et al. (2005). "Twenty five year mortality and air pollution: results from the French PAARC survey." Occup Environ Med 62(7): 453-60.

Finkelstein, M. M., M. Jerrett, et al. (2004). "Traffic air pollution and mortality rate advancement periods." Am J Epidemiol 160(2): 173-7.

Fiscella, K. and P. Franks (1997). "Poverty or income inequality as predictor of mortality: longitudinal cohort study." BMJ 314(7096): 1724-7.

Fischer, P., G. Hoek, et al. (2000). "Traffic-related differences in outdoor and indoor concentrations of particles and volatile organic compounds in Amsterdam." Atmos Environ 34: 3713-3722.

Geurs, K. (1995). Environmental-effects of traffic and transport policy at the front door (Milieu-effecten van verkeers- en vervoerbeleid aan de voordeur). Bilthoven, RIVM.

Harms, L. (2000). Traffic divided (Verkeer verdeeld). Bilthoven, RIVM.

Hoek, G., B. Brunekreef, et al. (2002). "Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study." Lancet 360(9341): 1203-9.

Hoek, G., P. Fischer, et al. (2001). "Estimation of long-term average exposure to outdoor air pollution for a cohort study on mortality." J Expo Anal Environ Epidemiol 11: 459-469.

Hofstra, P. and D. Bergsma (1986). Assessment of traffic data for the Dutch Noise Law (Bepaling van verkeersgegevens ten behoeve van de Wet Geluidhinder). Groningen, Hofstra Verkeersadviseurs BV.

Janssen, N., P. v. Vliet, et al. (2001). "Assessment of exposure to traffic related air pollution of children attending schools near motorways." Atmos Environ 35: 3875-3884.

Jerrett, M., A. Arain, et al. (2005). "A review and evaluation of intraurban air pollution exposure models." J Expo Anal Environ Epidemiol 15(2): 185-204.

Jerrett, M., R. T. Burnett, et al. (2005). "Spatial analysis of air pollution and mortality in Los Angeles." Epidemiology 16(6): 727-36.

Katsouyanni, K., G. Touloumi, et al. (2001). "Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project." Epidemiology 12(5): 521-31.

Krewski, D., R. Burnett, et al. (2003). "Rejoinder: reanalysis of the harvard six cities study and american cancer society study of particulate air pollution and mortality." J Toxicol Environ Health A 66(16-19): 1715-22.

Krewski, D., R. Burnett, et al. (2000). Reanalysis of the Harvard Six Cities study and the American Cancer Society study of particulate air pollution and mortality. Cambridge, Health Effects Institute.

Kromhout, H. and R. Vermeulen (2001). "Application of job-exposure matrices in studies of the general population: some clues to their performance." Eur Respir Rev 11: 80-90.

Kromhout, H., R. Vermeulen, et al. (2004). "Minisymposium 5.2: Experiences with the ALOHA JEM in longitudinal analyses within ECRHS II." Occup Environ Med 61(11): e31.

Laan, H. (1992). Environmental Traffic Maps: An analysis on a national level (Verkeersmilieukaarten: Een analyse op nationaal niveau). Bilthoven, RIVM.

Laden, F., J. Schwartz, et al. (2006). "Reduction in Fine Particulate Air Pollution and Mortality: Extended Follow-up of the Harvard Six Cities Study." Am J Respir Crit Care Med 173(6): 667-72.

Lewne, M., J. Cyrys, et al. (2004). "Spatial variation in nitrogen dioxide in three European areas." Sci Total Environ 332(1-3): 217-30.

Lucht, F. v. d. and H. Verkleij (2001). Gezondheid in de grote steden - Achterstanden en kansen. Houten, Bohn Stafleu Van Loghum.

Ministry of Transport, P. w., and Water Management (2001). Manual National Road Database - NWB (Handleiding Nationaal Wegenbestand - NWB). The Hague, Ministry of Transport, Public works, and Water Management.

Nafstad, P., L. L. Haheim, et al. (2004). "Urban air pollution and mortality in a cohort of Norwegian men." Environ Health Perspect 112(5): 610-5.

Osler, M., E. Prescott, et al. (2002). "Income inequality, individual income, and mortality in Danish adults: analysis of pooled data from two cohort studies." BMJ 324(7328): 13-6.

Pebesma, E. (2004). "Multivariable geostatistics in S: the gstat package." Computers & Geosciences 30: 683-691.

Pickett, K. E. and M. Pearl (2001). "Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review." J Epidemiol.Community Health 55(2): 111-122.

Pope, C. A., 3rd, R. T. Burnett, et al. (2002). "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution." Jama 287(9): 1132-41.

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

Reijneveld, S. A. (2002). "Neighbourhood socioeconomic context and self reported health and smoking: a secondary analysis of data on seven cities." J Epidemiol Community Health 56(12): 935-42.

Reijneveld, S. A., R. A. Verheij, et al. (2000). "The impact of area deprivation on differences in health: does the choice of the geographical classification matter?" J Epidemiol.Community Health 54(4): 306-313.

Roorda-Knape, M., N. Janssen, et al. (1998). "Air pollution from traffic in city districts near major motorways." Atmos Environ 32(11): 1921-1930.

Samet, J. M., F. Dominici, et al. (2000). "Fine particulate air pollution and mortality in 20 U.S. cities, 1987- 1994." N.Engl.J.Med. 343(24): 1742-1749.

Steenland, K., J. Henley, et al. (2004). "Individual- and area-level socioeconomic status variables as predictors of mortality in a cohort of 179,383 persons." Am J Epidemiol 159(11): 1047-56.

't Mannetje, A. and H. Kromhout (2003). "The use of occupation and industry classifications in general population studies." Int J Epidemiol 2003(32): 419-428.

van de Wiel HJ, Uiterwijk JW, et al. (1977). Instrument performance. Bilthoven, RIVM.

van den Brandt, P. A., R. A. Goldbohm, et al. (1990). "A large-scale prospective cohort study on diet and cancer in The Netherlands." J Clin Epidemiol 43(3): 285-95.

van Elzakker, H. (2001). Monitoring activities in the Dutch National Air Quality Monitoring Network in 2000 and 2001. Bilthoven, RIVM.

Vermeulen, R. and I. Wouters (1993). The ratio between black smoke, PM10, NO2, CO in traffic situations (de verhouding tussen zwarte rook, PM10, NO2, CO in afhankelijkheid van verkeerssituaties). Wageningen, Landbouwuniversiteit Wageningen.

Veugelers, P. J., A. M. Yip, et al. (2001). "Proximate and contextual socioeconomic determinants of mortality: multilevel approaches in a setting with universal health care coverage." Am J Epidemiol 154(8): 725-32.

VIG/Kadata (2001). Address Coordinates Netherlands (ACN) - Quality survey 2000 (Adres Coordinaten Nederland (ACN) - Kwaliteitsonderzoek 2000). Apeldoorn, Kadaster.

Volovics, A. and P. v. d. Brandt (1997). "Methods for the analyses of case-cohort studies." Biom J 39(2): 195-214.

Voorrips, L. E., R. A. Goldbohm, et al. (2000). "Vegetable and fruit consumption and lung cancer risk in the Netherlands Cohort Study on diet and cancer." Cancer Causes Control 11(2): 101-15.