eAppendix 1: References of included studies


42. Chen Qab, Garabrant DHcd, Hedgeman Ec, Little RJAb, Elliott MRbe, Gillespie Bb, Hong Bc, Lee S-Yb, Lepkowski JMe, Franzblau Acd, Adriaens Pf, Demond AHf, Patterson DGJg. Estimation of Background Serum 2,3,7,8-TCDD Concentrations By Using Quantile Regression in the UMDES and NHANES Populations. *Epidemiology* 2010;21(4)(Supplement):S51-S57.


60. Delfino RJa, Tjoa Ta, Gillen DLb, Staimer Na, Polidori Ac, Arhami Mc, Jamner Ld, Sioutas Cc, Longhurst Je. Traffic-related Air Pollution and Blood Pressure in Elderly Subjects With Coronary Artery Disease. *Epidemiology* 2010;21(3):396-404.


75. Freedman LS, Tasevska N, Kipnis V, Schatzkin A, Mares J, Tinker L, Potischman N. Gains in Statistical Power From Using a Dietary Biomarker in Combination With Self-reported Intake to Strengthen the Analysis of a Diet-Disease Association: An Example From CAREDS. *American Journal of Epidemiology* 2010;172(7):836-842.


131. Kiran Sa, Cocco Pb, Mannetje Ac, Satta Gb, D'Andrea Ib, Becker Nd, de Sanjose Sef, Foretova Lg, Staines Ah, Maynadie Mj, Nieters Ad, Brennan Pk, Boffetta Plm. Occupational Exposure to Ethylene Oxide and Risk of Lymphoma. *Epidemiology* 2010;21(6):905-910.


134. Krishnan S, Cozier YC, Rosenberg L, Palmer JR. Socioeconomic Status and Incidence of Type 2 Diabetes: Results From the Black Women's Health Study. *American Journal of Epidemiology* 2010;171(5):564-570.


143. Lenters Va, Uiterwaal CSb, Beelen Ra, Bots MLb, Fischer Pc, Brunekreef Bab, Hoek Ga. Long-Term Exposure to Air Pollution and Vascular Damage in Young Adults. *Epidemiology* 2010;21(4):512-520.


187. Ogundiran TO, Huo DZ, Adenipekun A, Campbell O, Oyesegun R, Akang E, Adebamowo C, Olopade OI. Case-Control Study of Body Size and Breast Cancer


269. Wikstrom A-Kab, Cnattingius Sa, Stephansson Oac. Maternal Use of Swedish Snuff (Snus) and Risk of Stillbirth. Epidemiology 2010;21(6):772-778.


273. Wu CF, Kuo IC, Su TC, Li YR, Lin LY, Chan CC, Hsu SC. Effects of Personal Exposure to Particulate Matter and Ozone on Arterial Stiffness and Heart Rate Variability in Healthy Adults. American Journal of Epidemiology 2010;171(12):1299-1309.


Study design

1. Type of design:
   - Treatment studies
     - RCT
     - Nonrandomized trial (quasi-experiment)
   - Observational studies
     - Cohort study
       - Prospective cohort
       - Retrospective cohort
     - Case-control study
     - Cross-sectional study

2. Researched population
   - Patients
   - Healthy individuals

3. If patients: type of Patients included

   _______________________________________________________________________

4. Number of participants

   _______________

5. Principal analysis used in the study:
   - Linear/logistic/Cox/Poisson regression
   - Mixed models
   - Generalized Estimating Equations (GEE)
   - T-tests
   - ANOVA / MANOVA / Repeated measures ANOVA
   - Item response theory (IRT)
   - Chi-square test
   - Nonparametric test
   - Other: ________________________________
6a. A Questionnaire was used for the assessment of:
- Covariates/Predictor
- Outcome
- Both
- None

6b. Did the questionnaire consist of different items resulting in a total score (or total scores per dimension/subscale)?
- Yes
- No
- Unclear
- No information
- Not applicable

6c. Is more information available on the questionnaire (e.g. response/answer categories, score calculation, etc.)?
- Yes
- No

**Missing data information:**

7a. Is the percentage/number of missing data described?
- Yes
- No
- Unclear

7b. Is the location of missing data described?
- Yes
- No
- Unclear
- Not applicable

7c. What is the location of missing data presented?
- Missing in variables
- Missing total scores (also attrition)
- Missing items
- Missing cases (unit nonresponse: did not show up/return questionnaire)
- Planned missingness (eg. missing by design)
- Other: ____________________________
- Unclear
- Not applicable

7d. What type of missing data is reported?
- Nonresponse (item/variable/unit)
- Dropout
- Attrition
- Lost to follow up
7e. What is the percentage of missings in the data reported?
   Total score: ______________
   Item: ______________
   Cases: ______________

8. Is the fraction of missing information presented?
   □ Yes
   □ No
   □ Unclear

9. Are the potential reasons for missing data discussed (eg. exhaustion, deceased, lack of motivation, lost in mail, etc.)?
   □ Yes
   □ No
   □ Unclear

10a. Was the missing data mechanism evaluated?
   □ Yes
   □ No
   □ Unclear

10b. Which method was used to test the mechanism?
   □ Differences in characteristics between missing and non-missing group described
   □ Analysis between cases with complete and missing data:
     □ Descriptive statistics (e.g. comparing means / percentages)
     □ Chi-square tests
     □ T-tests
     □ Univariate t-Test comparisons
     □ Little’s MCAR test
     □ Logistic regression analysis with missing data as outcome
   □ Other: __________________________
   □ Unclear
   □ Not applicable

10c. What category of missing data mechanism is reported?
   □ MCAR
   □ MAR
   □ MNAR
   □ Other: __________________________
☐ Unclear
☐ No information
☐ Not applicable
Methods used to handle the missing data:

11. Handling method

- Missing total score/unit score methods
  - Complete-case analysis
  - Pairwise deletion
  - Mean substitution/arithmetic mean imputation/unconditional mean imputation/median imputation
  - Single regression imputation (e.g. Stochastic)
  - Hot-deck imputation – matching nonrespondents to resembling respondent
  - Last value carried forward
  - Multiple imputation
  - Full Information Maximum Likelihood Estimation

- Missing item score methods
  - Unconditional random imputation
  - Item mean substitution/person mean substitution
  - Corrected item mean substitution
  - Two-way imputation
  - Response-function imputation
  - Multivariate normal imputation
  - Fully conditional specification
  - Similar Response pattern imputation
  - Item correlation substitution
  - Multiple response-function imputation
  - Including a missing category
  - Unclear
  - No information
  - Other ________________________________
  - Not Applicable

If multiple imputation is used (only 12a, 12b and 12c):

12a. Are the number of variables used in the imputation model clearly described?
  - No
  - No but normality discussed
  - Yes
  - Not applicable

12b. Is the number of multiple imputations presented?
  - Yes
  - No
  - Unclear
  - Not applicable

12c. Was the imputation process evaluated (i.e. convergence studied, imputed values compared with observed values, etc)?
  - Yes
  - No
13. Is it described how non-normal / categorical variables were dealt with in the missing data method?
   - □ Yes
   - □ No
   - □ Unclear
   - □ Not applicable

14a. Was a sensitivity analysis performed to investigate the influence of how the missing data were handled on the study results?
   - □ Yes
   - □ No
   - □ Unclear
   - □ Not applicable

14b. What kind of sensitivity analysis was performed?
   - □ Complete case analysis versus imputation method
   - □ Different imputation techniques were compared
   - □ Other: __________________________
   - □ Not applicable

14c. Are the results of the sensitivity analysis clearly described?
   - □ Yes
   - □ No
   - □ Unclear
   - □ Not applicable

15. What Software package was used for the missing data method?
   - □ SPSS
     - □ AMOS – Structural equation modeling tool in SPSS
   - □ EQS – Structural equation modeling software
   - □ HLM – Hierarchical data modeling
   - □ LISREL – Structural equation modeling software
   - □ Mplus – Statistical modeling program
   - □ SAS
   - □ SOLAS for missing data analysis
   - □ Stata
   - □ EMCOV
   - □ S-Plus
   - □ R
   - □ S-Plus en R packages
     - □ Amelia - Amelia II: A Program for Missing Data
     - □ NORM - analysis of multivariate normal datasets with missing values
     - □ CAT - Analysis of categorical-variable datasets with missing values
☐ MICE - Multivariate imputation by chained equations
☐ MI - Missing Data Imputation and Model Checking
☐ MIX - Multiple imputation for Mixed Categorical and Continuous Data
☐ missMDA - Handling missing values with/in multivariate data analysis (principal component methods)
☐ mitools - Tools for multiple imputation of missing data
☐ mlmmm - ML estimation under multivariate linear mixed models with missing values
☐ mvnml - ML estimation for multivariate normal data with missing values
☐ PAN - Multiple imputation for multivariate panel or clustered data
☐ MIXED
☐ No information
☐ Unclear
☐ Not applicable

16. What software was used for the primary/general analyses?
☐ SPSS Version___________
☐ SAS Version___________
☐ Stata Version___________
☐ Statistica Version___________
☐ R Version___________
☐ Mplus Version___________
☐ S-plus Version___________
☐ EQS Version___________
☐ LISREL Version___________
☐ SUDAAN Version___________
☐ Other: _____________________
☐ No Information