Environmental Research 109 (2009) 193–199

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Examination of the relationships between environmental exposures to volatile organic compounds and biochemical liver tests: Application of canonical correlation analysis ☆

Jing Liu [a,b], Wanzer Drane [a], Xuefeng Liu [a], Tiejian Wu [a,c,*]

[a] Department of Biostatistics and Epidemiology, East Tennessee State University, Johnson City, TN 37614, USA
[b] Institute of Epidemiology and Biostatistics, Shandong University, Shandong Province, China
[c] Department of Family Medicine, East Tennessee State University, USA

ARTICLE INFO

Article history:
Received 23 July 2008
Received in revised form
28 October 2008
Accepted 10 November 2008
Available online 30 December 2008

Keywords:
Canonical correlation
Environmental exposure
Liver function
Volatile organic compound

ABSTRACT

This study was to explore the relationships between personal exposure to 10 volatile organic compounds (VOCs) and biochemical liver tests with the application of canonical correlation analysis. Data from a subsample of the 1999–2000 National Health and Nutrition Examination Survey were used. Serum albumin, total bilirubin (TB), alanine aminotransferase (ALT), aspartate aminotransferase (AST), lactate dehydrogenase (LDH), alkaline phosphatase (ALP), and γ-glutamyl transferase (GGT) served as the outcome variables. Personal exposures to benzene, chloroform, ethylbenzene, tetrachloroethene, toluene, trichloroethene, *o*-xylene, *m*-,*p*-xylene, 1,4-dichlorobenzene, and methyl *tert*-butyl ether (MTBE) were assessed through the use of passive exposure monitors worn by study participants. The first two canonical correlations were 0.3218 and 0.2575, suggesting a positive correlation mainly between the six VOCs (benzene, ethylbenzene, toluene, *o*-xylene, m-,p-xylene, and MTBE) and the three biochemical liver tests (albumin, ALP, and GGT) and a positive correlation mainly between the two VOCs (1,4-dichlorobenzene and tetrachloroethene) and the two biochemical liver tests (LDH and TB). Subsequent multiple linear regressions show that exposure to benzene, toluene, or MTBE was associated with serum albumin, while exposure to tetrachloroethene was associated with LDH and total bilirubin. In conclusion, exposure to certain VOCs as a group or individually may influence certain biochemical liver test results in the general population.

1. Introduction

Volatile organic compounds (VOCs) are contained in a wide variety of commercial, industrial, and residential products including fuel oils, gasoline, solvents, cleaners and degreasers, paints, inks, dyes, refrigerants, and pesticides. Because of their ubiquitous presence in the environment, accurate assessment of the risk to public health posed by VOCs requires both the quantification of these exposures on a population-wide basis and the evaluation of potential health effects associated with varying exposure levels.

The liver is the major site for processing chemicals and drugs which enter the blood stream. The liver helps by removing these chemicals from the blood stream and changing them into products that can be readily removed through the bile or urine. In this process, unstable toxic products are sometimes produced, which can attack and injure the liver. Many VOCs could cause toxic chemical injury to the liver through this type of mechanism, in addition to direct toxicity (Brautbar and Williams, 2002; Xiao and Levin, 2000).

Biochemical liver tests include tests that are routinely measured in all clinical laboratories (Cahill, 1999). Several tests including serum aspartate aminotransferase, γ-glutamyltransferase, and alkaline phosphatase can serve as sensitive indicators of liver injury (Giannini et al., 2005). A recent cohort study (Kim et al., 2004) found that there was a positive association between aminotransferase concentration, even within normal range, and mortality from liver disease, suggesting that moderately increased aminotransferase activity is significant in predicting liver disease.

In exploring the health effects of environmental exposures, observational epidemiologic studies often deal with data that include both a set of exposure variables and a set of outcome variables. Routine approaches such as multiple linear regressions to analyze such data are usually challenged as they are plagued by the potential issues including multicollinearity and multiple testing. Since canonical correlation analysis (CCA) assesses the

☆ *Funding*: National Institutes of Health (R03ES016368). *Human subjects*: Five hundred and sixty four human subjects involved. *Animal subjects*: Not applicable.
* Corresponding author at: Department of Biostatistics and Epidemiology, East Tennessee State University, Johnson City, TN 37614, USA. Fax: +1 423 439 6491.
E-mail address: wut@etsu.edu (T. Wu).

0013-9351/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.envres.2008.11.002

correlation of two canonical variates (latent variables), one representing a set of the exposure variables and the other a set of outcome variables, it is potentially a useful method to evaluate the health effects of environmental exposures. Although canonical correlation has often been applied to social sciences and bioinformatics (Stevens, 1986; Pugh and Hu, 1991; Steinfath et al., 2007), the method has been rarely used in environmental health assessment.

The National Health and Nutrition Examination Survey conducted during 1999 and 2000 (NHANES 1999–2000) included a sub-project which collected detailed information on personal exposures to 10 VOCs (CDCa). NHANES 1999–2000 also collected blood samples and performed a number of biochemical liver tests (CDCa). In this study, we applied CCA to the NHANES data and explored if there were associations between the environmental VOC exposures and the biochemical liver tests.

2. Materials and methods

2.1. Data source and study sample

Data from NHANES 1999–2000 were used for this study. Briefly, NHANES 1999–2000 used a stratified multistage probability design to obtain a representative sample of the civilian non-institutionalized US general population. A total of 12,160 persons were asked to take part in NHANES 1999–2000. Interview and physical examination data were collected on 9282 of the eligible participants, and the overall response rate was 76.3% (CDCb). Since the NHANES data sets are accessible to the public, the Institutional Review Board of East Tennessee State University granted the exemption of review and approved the study.

The VOC Project of personal exposures to air toxics was conducted among a representative subsample of the NHANES participants between the ages of 20 and 59 years (CDCc). The project was designed to characterize exposures to these air toxics and to determine predictors of exposure. Among 851 subjects eligible for the VOC Project during 1999 and 2000, about 75% had measurements of personal VOC exposure.

To limit potential confounding, we excluded the subjects who had liver conditions, heart disease, stroke, cancer, or diabetes. These conditions were confirmed if a subject reported in NHANES that a physician had ever told him/her having these conditions. Those tested serum positive to hepatitis C virus (HCV) in NHANES were also excluded from the study. Of the 851 subjects who participated in the VOC project, 115 who had one or more of the above conditions were excluded from the study. We then excluded 170 subjects, who did not have VOC assessments and the two subjects with very extreme values in ALT (1163 U/L) or AST (827 U/L), resulting in a sample size of 564. However, depending on the variables analyzed, the sample size varied slightly because some of the variables of interest had missing values. A comparison of the study sample with the NHANES 1999–2000 participants in several characteristics are shown in Table 1.

2.2. Study variables

2.2.1. Demographic and behavioral characteristics

Demographic information such as gender, age, ethnicity, family income, education attainment, and health risk behaviors such as cigarette smoking and alcohol consumption were collected in NHANES 1999–2000 through interviews or questionnaires.

2.2.2. Biochemical liver tests

In NHANES 1999–2000, venous blood specimens were collected, centrifuged, refrigerated at 4–8 °C, and then shipped weekly to a central laboratory where they were tested upon arrival (CDCd). Serum markers of liver function assessed in NHANES 1999–000 include: albumin, total bilirubin (TB), alanine aminotransferase (ALT), aspartate aminotransferase (AST), lactate dehydrogenase (LDH), alkaline phosphatase (ALP), and γ-glutamyl transferase (GGT). Laboratory procedures and quality control of the measurements are detailed by the NHANES Laboratory Procedure Manual (CDCd). In this study, the biochemical liver tests served as continuous outcome variables in the analyses.

2.2.3. Personal exposure to VOCs

Personal exposure to 10 VOCs including benzene, chloroform, ethylbenzene, tetrachloroethene, toluene, trichloroethene, *o*-xylene, *m*-,*p*-xylene, 1,4-dichlorobenzene, and methyl *tert*-butyl ether (MTBE) were assessed through the use of passive exposure monitors (or badges) worn by participants for a period of 48–72 h. The actual duration of exposure ranged from 43.3 to 76.0 h with a mean of 56.4 h. The number of cubic meters for each subject was derived from the duration the exposure monitor was used. The value for each analyte was expressed as the weight of the analyte per cubic meter. If a VOC test result was below the limit of detection, a value was imputed in NHANES 1999–2000 using the detection limit adjusted for the actual duration the badge was exposed divided by the square root of two. The procedures of sample collection and lab tests were detailed in the NHANES documentation (CDCc).

2.3. Statistical analysis

2.3.1. Descriptive analysis

Descriptive statistics including mean, standard deviation, and proportion were used to describe the characteristics of both the study sample and the NHANES participants. Descriptive statistics including mean, median, percentiles, minimum, and maximum values were used to show the distributions and numerical characteristics of the seven liver function test results and the exposure levels of the 10 VOCs.

2.3.2. Canonical correlation analysis (CCA)

CCA is an exploratory statistical method to assess correlations between two sets of variables (Stevens, 1986). One of the assumptions of CCA for the significance testing of canonical correlations is that the variables in both sets follow a multivariate normal distribution. Therefore, in this study, we transformed the VOC and liver function test variables to Blom normal scores from their ranks to assure that the multivariate normality is not violated. The transformation formula was $Z_i = \Phi^{-1}((r_i-3/8)/(n+1/4))$, where Φ^{-1} is the inverse cumulative normal (PROBIT) function, r_i is the rank of the ith observation, and n is the number of observations for the ranking variable (Blom, 1958; Tukey, 1962).

The fundamental principle behind CCA is the creation of a number of canonical variates, each consisting of a linear combination of one set of variables (X_i), which has the form:

$$U_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p$$

and a linear combination of the other set of variables (Y_i), which has the form:

$$V_i = b_{i1}Y_1 + b_{i2}Y_2 + \cdots + b_{iq}Y_q$$

The goal is to determine the coefficients, or canonical weights (a_{ij} and b_{ij}), that maximize the correlation between canonical variates U_i and V_i. The first canonical correlation, Corr(U_1, V_1), is the highest possible correlation between any linear combination of the variables in the exposure set and any linear combination of the variables in the outcome set. Further pairs of maximally correlated linear combinations are chosen in turn, and they are orthogonal to those already identified. The maximum number of canonical correlation is equal to the number of variables in the smaller set, which is seven in this study (the number of biochemical liver tests of interest). Significance test of a canonical correlation coefficient was performed using likelihood ratio test. Since this was an exploratory study of the health effects of VOC exposures, the significance level of the test was set at 0.10, instead of 0.05, to limit the chance of failing to detect an effect.

Structure correlation coefficients, also called canonical loadings, are used to interpret the importance of each original variable in the canonical variates. A structure correlation is the correlation of a canonical variate with the variable in its set. Variables that are highly correlated with a canonical variate should be considered more important when deriving a meaningful interpretation of the related canonical variate. This way of interpreting canonical variates is the same as the interpretation of factors in factor analysis (Shafto et al., 1997). As a rule of thumb, an absolute value of 0.3 or greater in canonical loading was used to select the variables that are thought to have a meaningful interpretation of the related canonical variate (Lambert and Durand, 1975; Thompson, 1984). We chose a cutoff value of 0.35 to select important loadings in this study.

2.3.3. Linear regression analysis

Restricted to the important VOC and liver function test variables derived from CCA, multiple linear regressions were used to examine the relationship between an individual VOC variable and an individual biochemical liver test, without and with adjustment for sex, age, BMI, ethnicity, educational level, family income, alcohol drinking, and cigarette smoking. These covariates are described in Table 1. Logarithm-transformations of ALP, GGT, LDH, and total bilirubin were performed to reduce the skewness of the variables. The level of significance was set at 0.10. All the analyses described above were performed using SAS statistical software (SAS Institute, 1989; Afifi et al., 2004).

3. Results

3.1. Characteristics of the study participants

The proportion of men in the sample was 44.33% and women 55.67%. Non-Hispanic whites accounted for 40.78%, Mexican