

## OTHER ORIGINAL ARTICLES

# Systematic evaluation of environmental factors: persistent pollutants and nutrients correlated with serum lipid levels

Chirag J Patel,<sup>1,2</sup> Mark R Cullen,<sup>3</sup> John PA Ioannidis<sup>4,5,6</sup> and Atul J Butte<sup>1,2\*</sup>

<sup>1</sup>Department of Pediatrics, Division of Systems Medicine, Stanford University School of Medicine, Stanford, CA, USA,

<sup>2</sup>Lucile Packard Children's Hospital, Palo Alto, CA, USA, <sup>3</sup>Department of Medicine, Division of General Medical Disciplines, Stanford University School of Medicine, Stanford, CA, USA, <sup>4</sup>Department of Medicine, Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA, USA, <sup>5</sup>Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA and <sup>6</sup>Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA USA

\*Corresponding author. Department of Pediatrics, Stanford School of Medicine, 1265 Welch Road, Stanford, CA 94305, USA. E-mail: [abutte@stanford.edu](mailto:abutte@stanford.edu)

---

**Accepted** 5 January 2012

**Background** Both genetic and environmental factors contribute to triglyceride, low-density lipoprotein-cholesterol (LDL-C), and high-density lipoprotein-cholesterol (HDL-C) levels. Although genome-wide association studies are currently testing the genetic factors systematically, testing and reporting one or a few factors at a time can lead to fragmented literature for environmental chemical factors. We screened for correlation between environmental factors and lipid levels, utilizing four independent surveys with information on 188 environmental factors from the Centers of Disease Control, National Health and Nutrition Examination Survey, collected between 1999 and 2006.

**Methods** We used linear regression to correlate each environmental chemical factor to triglycerides, LDL-C and HDL-C adjusting for age, age<sup>2</sup>, sex, ethnicity, socio-economic status and body mass index. Final estimates were adjusted for waist circumference, diabetes status, blood pressure and survey. Multiple comparisons were controlled for by estimating the false discovery rate and significant findings were tentatively validated in an independent survey.

**Results** We identified and validated 29, 9 and 17 environmental factors correlated with triglycerides, LDL-C and HDL-C levels, respectively. Findings include hydrocarbons and nicotine associated with lower HDL-C and vitamin E ( $\gamma$ -tocopherol) associated with unfavourable lipid levels. Higher triglycerides and lower HDL-C were correlated with higher levels of fat-soluble contaminants (e.g. polychlorinated biphenyls and dibenzofurans). Nutrients and vitamin markers (e.g. vitamins B, D and carotenes), were associated with favourable triglyceride and HDL-C levels.

**Conclusions** Our systematic association study has enabled us to postulate about broad environmental correlation to lipid levels. Although subject to confounding and reverse causality bias, these findings merit evaluation in additional cohorts.

**Keywords** Lipids, cholesterol, environment, pollutants, nutrients, GWAS, EWAS

## Introduction

Serum lipid levels are risk factors for coronary heart disease (CHD), atherosclerosis, type 2 diabetes and stroke. Both genetic and environmental factors influence lipid level phenotypes. Lipid level variation can be influenced by genetics.<sup>1,2</sup> On the other hand, environmental factors also play a role. For example, lifestyle factors such as physical exercise, smoking and diet have well-documented relationships with lipid levels.<sup>3–7</sup> There are a number of reports connecting lipid levels, cardiovascular disease, type 2 diabetes and the metabolic syndrome with specific persistent pollutants, such as dioxins, organochlorinated pesticides, dibenzofurans and polychlorinated biphenyls (PCBs).<sup>8–14</sup> Other less tangible environmental factors, such as air pollution,<sup>15</sup> may also have an adverse relationship with lipid levels.

Although extensive efforts are underway to dissect genetic components with genome-wide association studies (GWASs)<sup>16</sup>, similar studies to systematically identify specific environmental factors are lacking. Results of epidemiologic studies, which typically test one or a few factors at a time, may be further distorted by selective reporting of subsets of analyses, outcomes and adjustments. It has been postulated that this contributes to a fragmented, ultimately unreliable literature.<sup>17–22</sup> More importantly, the phenomenon of environmental exposure is complex and influenced by differences in individuals, time, place and other exposures.<sup>19</sup> Humans are exposed to not a few, but many environmental adverse or protective factors simultaneously. Due to this complexity, the net effects due to environmental factors on human health may be miscalculated when considering a few factors at a time. Protective effects of environmental factors are not usually considered in context of adverse effects of other co-existing factors, potentially leading to lack of physiological coherence and public health relevance. Health surveys and bio-monitoring projects,<sup>23</sup> in which multiple environmental factors are simultaneously measured, provides an opportunity to hypothesize about how a system of environmental factors relate to disease and other characteristics among the general population.<sup>24</sup>

We conduct a more systematic approach to associating multiple environmental chemical factors with serum lipid levels, similar to a GWAS,<sup>25</sup> utilizing the National Health and Nutrition Examination Survey (NHANES), a nationally representative health survey.<sup>26</sup> Instead of testing a few associations at a time, we evaluate 188 environmental factors for association to lipid levels while accounting for the multiplicity of comparisons. The emerging significant associations are then validated in an independent NHANES dataset. Further, we conduct

systematic sensitivity analyses among the measured confounders to estimate bias. We term this method an 'Environment-wide Association Study', or EWAS.

Using such an analytic procedure, we have found and validated 29, 9 and 17 markers for environmental chemical factors correlated with triglycerides, low-density lipoprotein-cholesterol (LDL-C) and high-density lipoprotein-cholesterol (HDL-C), respectively, including a spectrum of persistent organic pollutants, nutrients and vitamins. Many of these factors have been explored before in association with related diseases, such as type 2 diabetes, obesity, lipid levels and the metabolic syndrome.<sup>10–14,27</sup> However, each of these studies addresses issues of model adjustment, variable coding and assessment of effects in different ways, possibly leading to conflicting study results.<sup>28</sup> In this systematic study, we propose one type of analytic process to unify and standardize these analyses. Specifically, we assess how environmental factors are correlated among themselves and with changes in serum lipid levels, while consistently adjusting for other factors such as age, sex, ethnicity, socio-economic status (SES) and body mass index (BMI).

## Methods

### Data

We downloaded all available NHANES laboratory and questionnaire data for 1999–2000, 2001–02, 2003–04 and 2005–06 surveys. Laboratory data included serum and urine measures of environmental factors and clinical measures including lipid levels. Each survey is an independent, non-overlapping sampling of participants representative of the general US population. We analysed factors that were a direct measurement of an extrinsic environmental factor (e.g. amount of pesticide or heavy metal in urine or blood). We did not consider intrinsic physiological measures (e.g. red blood cell count or albumin) or responses to questionnaires except for sensitivity analyses.

We used three of the four surveys (1999–2000, 2001–02, 2005–06) for testing for multiple environmental factors in association with lipid levels and reserved one survey (2003–04) for validation testing of findings. As each survey had a different set and number of environmental factors measured, we selected 2003–04 as the validation survey as it had the largest number of shared factors with each of the other surveys, maximizing the number of factors that could be validated.

We eliminated 119 factor variables from our analyses whose majority of observations were under the National Centers for Health Statistics (NCHS) documented limit of detection or, for categorical factor variables, varied little. Specifically, we omitted continuous variables if

99% of the observations were deemed below the threshold limit of detection. For categorical factor variables, we omitted those that had 99% of the observations belonging to one category. After the elimination of these 119 such factor variables, we were left with 169 variables from the 1999–2000 survey, 182 from 2001–02, 96 from 2005–06 and 258 from the 2003–04 (validation) survey. Next, we selected factors from each survey that were present in the validation survey. This left us with a total of 188 unique factors that could be validated, of which 126 were from the 1999–2000 survey, 157 from 2001–02 and 65 from 2005–06. Using a categorization provided by NHANES, we binned these factors into 26 ‘classes’ of related factors (Figure 1A, Supplementary Table 1, available as Supplementary Data at *IJE* online).

Different environmental factors were measured in varying numbers of participants: 109–3610 (median 938), 101–3388 (median 896) and 222–7485 (median 1958) individuals for triglyceride, LDL-C and HDL-C levels, respectively (Figure 1B). Individuals are selected randomly based on their demographic characteristics for the complex, stratified survey.<sup>29</sup> Serum triglyceride levels were measured in the morning after >8.5 h of fasting. LDL-C levels were derived from total cholesterol and direct HDL-C measurements were derived from the Friedewald calculation.<sup>30</sup>

### Correlation between factors

We computed the pair-wise partial Pearson correlation coefficient between each environmental factors using the test and validation surveys separately, adjusting for age and BMI in addition to creatinine levels for urinary measures. Since we had 188 environmental factors, the total number of possible pairs of factors (and correlations) equals 17 578 ( $188 \times 187/2$ ); however, 4455 (25%) of all possible pairs of factors were not measured in the same overlapping individuals and, as a result, their correlations could not be computed. We assessed correlations between factors in the test and validation cohorts separately and compared their relative strength by estimating percentiles of the entire distribution of correlations. We also compared correlations within classes (or ‘intra-class’ correlation) and between classes (‘inter-class’) correlations. For factors measured in more than one of the test surveys, their coefficient was combined using a meta-analytic random effects method.

### Correlation of environmental factors with lipid levels

The systematic analysis encompasses multiple steps (Figure 1C–G), which we term an ‘environment-wide association study’ (EWAS). First, survey-weighted linear regressions are performed, whereby  $\log_{10}$ -transformed lipid levels are dependent variables, modelled as a function of each environmental factor and age, age<sup>2</sup>, sex, BMI, ethnicity and SES (Figure 1C). For SES, we used the tertile of poverty index (participant’s household income divided by the time-adjusted poverty threshold), as previously described.<sup>25</sup> Ethnicity was coded in five groups

(Mexican American, Non-Hispanic Black, Non-Hispanic White, Other Hispanic, Other). We used *R survey* module for all survey-weighted analyses<sup>31</sup> with appropriate pseudo-strata, pseudo-sampling units and weights to accommodate the complex sampling of the data.

Chemical exposure data arising from mass spectrometry or absorption measurements were log-transformed. We used z-scores [standard deviations (SDs) from the mean] to compare effect sizes; specifically, effect sizes for these variables denote change in lipid levels for a change in 1 SD of exposure. For binary variables, such as presence/absence assays for infectious agents, effect sizes denote change in lipid levels for those with presence of a factor vs those without.

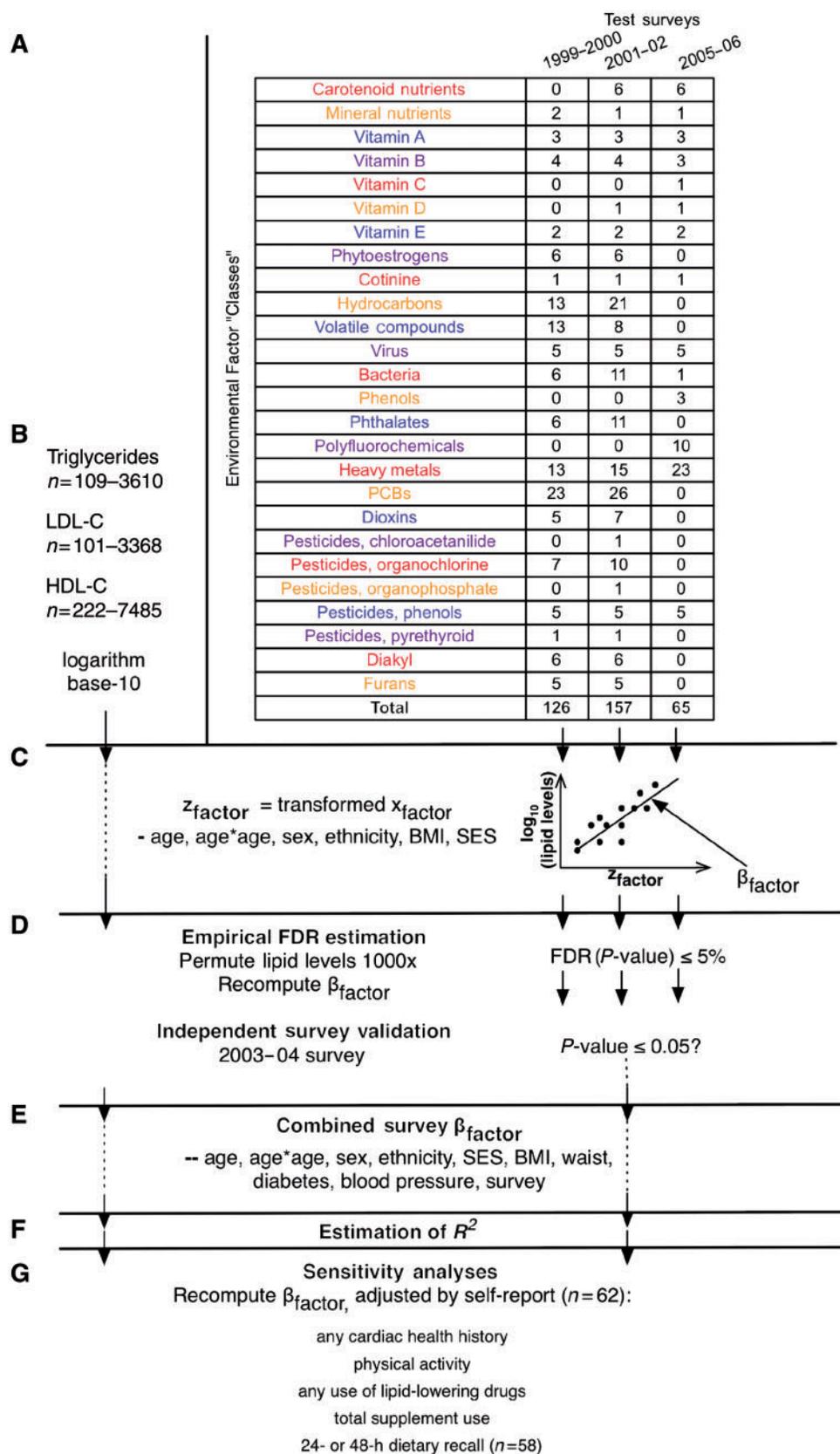
We calculated the false discovery rate (FDR), the estimated proportion of false discoveries made vs the number of total discoveries made for a given significance level  $\alpha$ , to control for multiple hypothesis testing (Figure 1D).<sup>32</sup> We created a ‘null distribution’ of regression test statistics for each survey separately, permuting the triglycerides, HDL-C and LDL-C levels 1000 times and refitting the linear regression models, collecting the test statistics for the coefficients corresponding to the environmental factor. In other words, the distributions of the lipid levels were not changed, but randomly assigned to different individuals in the survey.

The FDR is the ratio of the number of coefficients called significant at a given level  $\alpha$  in the null distribution and the number of results called significant from our real screen (Supplementary Methods, available as Supplementary Data at *IJE* online). We used a FDR of < 5% to select significant associations. We used an independent survey, the 2003–04 survey, to validate significant findings (Figure 1D). We considered a significant factor as ‘tentatively validated’ if it was significant ( $P < 0.05$ ) in the validation survey.

We then fit a final linear regression model with data combined from the four independent NHANES surveys for a tentatively validated environmental factor, attaining an overall estimate and *P*-value (Figure 1E). We utilized the larger sample size to adjust for additional co-variables that we were unable to adjust for in the single survey analyses (due to limited residual degrees of freedom) that also influence lipid levels. In addition to initial covariates, we also adjusted for waist circumference, type 2 diabetes status (approximated by fasting blood glucose  $\geq 126$  mg/dl), systolic and diastolic blood pressure (mm Hg) and survey. To estimate how much of the variance was described by each environmental factor, we estimated the change in the coefficient of determination ( $R^2$ ) adding that factor vs a model including only the adjusting factors (Figure 1F). We also performed regressions on untransformed lipid levels to estimate raw effect size.

### Sensitivity analyses

We conducted sensitivity analyses to account for recent food, alcohol, supplements, medications, exercise and



**Figure 1** Summary of environmental factors and analytic method. (A) Summary of the 26 factor classes and the number of factors within them for each NHANES test survey. (B) 100-7500 individuals had their HDL-C, LDL-C and triglyceride levels

(continued)

history of cardiovascular health (Figure 1G). In total, 62 questionnaire items were used (Supplementary Table 2, available as Supplementary data at *IJE* online). To evaluate the impact of these 62 adjusting variables, we recomputed the regression models by adding each variable to our final model one by one and observed the change in the effect size for each environmental factor. We also built a model adjusting for lipid-lowering drugs, supplement use, exercise and self-report cardiovascular-related disease simultaneously. More details can be found in the Supplementary Data (available at *IJE* online).

### Power calculations

We estimated<sup>33</sup> that our analyses had >80% median power for all surveys for detection of 5% change in HDL-C and LDL-C and 10% change of triglyceride levels for *P*-values corresponding to an FDR of 5%.

## Results

### Demographic and baseline associations with lipid levels

Tables 1–3 describe the multivariate relationship between baseline demographics and lipid levels. As expected,<sup>34</sup> demographics, BMI, ethnicity and SES are associated with lipid levels. For example, consistent positive correlations existed between age and triglycerides (5–10% higher per 10 years,  $P < 0.02$ ) and BMI and triglycerides (2% higher per 1 unit of BMI,  $P < 0.004$ ) and consistent negative correlations between Black ethnicity and triglycerides (13% lower vs White,  $P < 0.001$ ).<sup>35</sup> Consistent polynomial relationships existed between age and both HDL-C and LDL-C. Negative correlations existed between BMI and HDL-C (1% lower per BMI unit,  $P < 0.0001$ ). In addition, SES was associated with HDL-C (1–5% lower for lower vs higher tertile,  $P < 0.03$ ). These indicated that BMI, ethnicity, sex, SES, age and age<sup>2</sup>, were all covariates that needed to be controlled in our first stage analysis.

### Factor correlations

We computed the partial Pearson correlation between each pair of environmental chemical factors tested where pair-wise data were available. Of the 17 578 possible correlations, 13 123 correlations could be

computed ('Methods' section). These 13 123 correlations were adjusted for BMI and age in addition to creatinine levels for urinary measures. We computed the correlations in the test surveys and verified these correlations among the validation survey ( $P < 0.05$  for both test and validation surveys and with same sign); after this verification step, we were left with a total of 11 672 confirmed correlations (Figure 2). The 5th, 10th, 90th and 95th percentiles of verified partial Pearson correlations were  $-0.11$ ,  $-0.07$ ,  $0.26$  and  $0.38$ , respectively (lower left panel histogram, Figure 2).

The intra-class partial correlations were higher than between-class correlations in both the test surveys and validation survey (test surveys: mean  $\rho = 0.26$ , *t*-test  $P < 1e-10$ ; validation survey: mean  $\rho = 0.27$ , *t*-test  $P < 1e-10$ ). Specifically, the intra-class correlation for class PCBs was 0.41, 0.42 for dioxins, 0.5 for carotenoid nutrients, 0.2 for heavy metals, 0.2 for hydrocarbons, 0.3 for phytoestrogens, 0.3 for phthalates and 0.2 for phenols for the test surveys. We observed similar patterns among the validation survey. We observed several instances of large inter-class correlations, such as inverse correlations between carotenoid and vitamin E factors (*trans*- $\beta$ -carotene and  $\gamma$ -tocopherol,  $\rho = -0.3$ ). We also observed positive correlations between cotinine and the heavy metals lead and cadmium ( $\rho > 0.3$ ) and hydrocarbons 2- and 3-hydroxyfluorene ( $\rho > 0.5$ ). Similarly, we observed gross inter-class correlations between classes such as furans and dioxins (mean  $\rho = 0.4$ ), PCBs and dioxins (mean  $\rho = 0.2$ ), PCBs and organochlorine pesticides (mean  $\rho = 0.2$ ) and phthalates and hydrocarbons (mean  $\rho = 0.2$ ).

### Environment associations with lipid levels

For triglyceride levels, 15 out of 126, 29 out of 157 and 12 out of 65 factors passed the requested threshold of significance (FDR < 5%) for the 1999–2000, 2001–02 and 2005–06 surveys, respectively (Figure 3A). For LDL-C, 2 out of 131, 10 out of 162 and 9 out of 65 were significant, respectively (Figure 3B). For HDL-C, 1 out of 131, 26 out of 162 and 15 out of 65 were significant (Figure 3C). We tentatively validated significant findings from our screen, by searching for whether any of the factors significant in any of the three studies above were also significant in the fourth independent 2003–04

### Figure 1 Continued

measured for each of these factors in each survey; these lipid levels were log transformed to assume normality for least squares regression. (C) Each of these 126, 157 and 65 factors was tested for association with the logarithm base 10 of HDL-C, LDL-C and triglyceride levels with a linear regression model adjusted for age, age<sup>2</sup>, sex, BMI, ethnicity and SES. (D) We estimated the FDR by permuting the lipid levels and re-computing the linear models; an FDR of 0.05 was considered significant. We deemed a factor to be tentatively validated if it was found to be significant in the validation survey with  $P \leq 0.05$  and an effect in the same direction. (E) We estimated a final coefficient for tentatively validated factors by combining all surveys and adjusting for age, age<sup>2</sup>, sex, ethnicity, SES, BMI, waist circumference, type 2 diabetes status (fasting blood glucose  $\geq 126$  mg/dl), blood pressure and survey. (F) We estimated the coefficient of determination ( $R^2$ ) for the final, combined models. (G) We recomputed our final models, adding 62 self-report variables one by one to attempt to check the validity of the environmental effect

**Table 1** Estimates of multivariate linear regression model predicting  $\log_{10}$ (triglycerides) as a function of sex, age, age<sup>2</sup>, ethnicity (in reference to Whites), an estimate of SES (in reference to high SES) and BMI for each survey

Triglycerides	1999–2000 ( <i>n</i> = 3002)		2001–02 ( <i>n</i> = 3610)	
	Estimate (95% CI)	<i>P</i> -value	Estimate (95% CI)	<i>P</i> -value
Sex (vs male)	−0.026 (−0.061 to 0.008)	0.1	−0.061 (−0.089 to −0.033)	0.002
<b>Age (10 years)</b>				
Age	0.044 (0.011–0.076)	0.02	0.052 (0.019–0.085)	0.009
Age <sup>2</sup>	−0.00017 (−0.00047 to 0.00013)	0.2	−0.00026 (−0.00061 to 8.7e−05)	0.1
<b>Ethnicity (vs White)</b>				
Black	−0.14 (−0.19 to −0.098)	9 × 10 <sup>−4</sup>	−0.13 (−0.17 to −0.079)	0.001
Mexican–American	0.011 (−0.049 to 0.071)	0.6	0.0088 (−0.036 to 0.053)	0.6
Other Hispanic	−0.034 (−0.075 to 0.0074)	0.08	0.038 (−0.1 to 0.18)	0.5
Other	−0.027 (−0.098 to 0.044)	0.3	0.03 (−0.046 to 0.11)	0.4
<b>SES (vs high tertile)</b>				
SES (medium)	0.011 (−0.032 to 0.055)	0.5	0.018 (−0.011 to 0.047)	0.2
SES (low)	0.027 (−0.018 to 0.072)	0.2	0.037 (−0.0034 to 0.077)	0.07
BMI (10 U)	0.11 (0.078–0.15)	9 × 10 <sup>−4</sup>	0.093 (0.059–0.13)	0.001
	2005–06 ( <i>n</i> = 2912)		Validation: 2003–04 ( <i>n</i> = 3449)	
Sex (vs male)	−0.043 (−0.072 to −0.013)	0.01	−0.056 (−0.078 to −0.033)	0.001
<b>Age (10 years)</b>				
Age	0.046 (0.016–0.076)	0.01	0.078 (0.055–0.1)	3 × 10 <sup>−4</sup>
Age <sup>2</sup>	−2e−04 (−0.00052 to 0.00011)	0.2	−0.00061 (−0.00082 to −0.00039)	8 × 10 <sup>−4</sup>
<b>Ethnicity (vs White)</b>				
Black	−0.13 (−0.16 to −0.1)	9 × 10 <sup>−5</sup>	−0.13 (−0.17 to −0.097)	2 × 10 <sup>−4</sup>
Mexican–American	0.019 (−0.014 to 0.051)	0.2	0.0074 (−0.033 to 0.048)	0.7
Other Hispanic	−0.02 (−0.081 to 0.041)	0.4	0.03 (−0.046 to 0.11)	0.4
Other	0.044 (−0.019 to 0.11)	0.1	−0.012 (−0.089 to 0.065)	0.7
<b>SES (vs high tertile)</b>				
SES (medium)	0.0044 (−0.031 to 0.04)	0.8	0.005 (−0.028 to 0.038)	0.7
SES (low)	0.026 (−0.017 to 0.069)	0.2	0.035 (−0.0063 to 0.077)	0.08
BMI (10 U)	0.095 (0.071–0.12)	2 × 10 <sup>−4</sup>	0.099 (0.069–0.13)	4 × 10 <sup>−4</sup>

95% confidence interval (CI) and *P*-value of associations are also shown. *n* is unweighted sample size.

survey at *P* < 0.05. We found 29, 9 and 17 tentatively validated factors for triglycerides, LDL-C and HDL-C, respectively (Figure 3A–C).

The data were combined across surveys for each tentatively validated factor and estimates were further adjusted for waist circumference, type 2 diabetes status, blood pressure and survey, in addition to age, age-squared, BMI, age, sex, SES and ethnicity. The variance ascribed to baseline co-variables was 22–25% (triglycerides), 15–16% (LDL-C) and 23–26% (HDL-C). Each of the tentatively validated environmental factors described an additional 0.7–18.4% (triglycerides), 1.8–14.1% (LDL-C) and 0.4–4.0% (HDL-C) of the variance in lipid levels (Supplementary Tables 3–5, available as Supplementary Data at *IJE* online).

Effects for the tentatively validated associations are shown in Figure 4. We present here some of them in more detail. Effect sizes for continuous variables are for 1 SD of log-transformed value of the environmental factor.

### Vitamins A and E: unfavourable association with lipid levels

For all three lipids, we found a consistent association for lipid-soluble, anti-oxidant vitamins, such as vitamins A, E, and carotenoids (Figure 4 A–C and Supplementary Figure 1, available as Supplementary Data at *IJE* online). For example, a form of vitamin A, retinol, was positively associated with triglycerides ( $P = 6 \times 10^{-21}$ , effect = 10% or 25 mg/dl higher

**Table 2** Estimates of multivariate linear regression model predicting log<sub>10</sub>(LDL-C) as a function of sex, age, age<sup>2</sup>, ethnicity, SES and BMI for each survey

LDL-C	1999–2000 (n = 2743)		2001–02 (n = 3318)	
	Estimate (95% CI)	P-value	Estimate (95% CI)	P-value
Sex (vs male)	−0.015 (−0.034 to 0.0043)	0.1	−0.016 (−0.028 to −0.0049)	0.01
<b>Age (10 years)</b>				
Age	0.059 (0.042–0.076)	7 × 10 <sup>−4</sup>	0.058 (0.042–0.073)	2 × 10 <sup>−4</sup>
Age <sup>2</sup>	−0.00046 (−0.00066 to −0.00027)	0.003	−0.00048 (−0.00065 to −0.00032)	7 × 10 <sup>−4</sup>
<b>Ethnicity (vs White)</b>				
Black	−0.015 (−0.036 to 0.0065)	0.1	−0.0087 (−0.032 to 0.014)	0.4
Mexican–American	−0.012 (−0.028 to 0.0047)	0.1	−0.019 (−0.037 to −0.0013)	0.04
Other Hispanic	−0.013 (−0.034 to 0.0081)	0.2	−0.015 (−0.041 to 0.011)	0.2
Other	−0.0098 (−0.041 to 0.021)	0.4	0.01 (−0.036 to 0.056)	0.6
<b>SES (vs high tertile)</b>				
SES (medium)	−0.0028 (−0.027 to 0.021)	0.8	0.0044 (−0.019 to 0.028)	0.6
SES (low)	0.0094 (−0.014 to 0.033)	0.3	0.01 (−0.013 to 0.033)	0.3
BMI (10 U)	0.022 (0.006–0.038)	0.02	0.014 (0.0048–0.023)	0.01
		2005–06 (n = 2853)	Validation: 2003–04 (n = 3389)	
Sex (vs male)	−0.0024 (−0.018 to 0.013)	0.7	−0.0054 (−0.022 to 0.012)	0.4
<b>Age (10 years)</b>				
Age	0.098 (0.079–0.12)	4 × 10 <sup>−5</sup>	0.065 (0.049–0.081)	1 × 10 <sup>−4</sup>
Age <sup>2</sup>	−9e−04 (−0.0011 to −7 × 10 <sup>−4</sup> )	9 × 10 <sup>−5</sup>	−0.00055 (−0.00072 to −0.00038)	4 × 10 <sup>−4</sup>
<b>Ethnicity (vs White)</b>				
Black	−0.012 (−0.04 to 0.015)	0.3	−0.012 (−0.028 to 0.0033)	0.1
Mexican–American	0.0014 (−0.023 to 0.026)	0.9	0.0082 (−0.012 to 0.028)	0.3
Other Hispanic	0.035 (−0.011 to 0.081)	0.1	−0.019 (−0.097 to 0.059)	0.6
Other	−0.026 (−0.067 to 0.015)	0.2	−0.026 (−0.055 to 0.003)	0.07
<b>SES (vs high tertile)</b>				
SES (medium)	−0.013 (−0.035 to 0.0089)	0.2	0.0087 (−0.014 to 0.032)	0.4
SES (low)	−0.0023 (−0.02 to 0.015)	0.7	0.013 (−0.01 to 0.036)	0.2
BMI (10 U)	0.0078 (−0.0034 to 0.019)	0.1	0.015 (−0.0029 to 0.034)	0.08

95% CI: 95% confidence interval.

triglycerides per 1 SD) in all surveys examined. Another form of vitamin A, retinyl palmitate, was also positively associated with triglycerides ( $P = 6 \times 10^{-21}$ , effect = 10%) and LDL-C ( $P = 4 \times 10^{-13}$ , effect = 5% or 6 mg/dl). Retinyl stearate was negatively associated with HDL-C ( $P = 4 \times 10^{-5}$ , effect = −3% or −1 mg/dl).

We observed a consistent association between forms of vitamin E ( $\alpha$  and  $\gamma$  tocopherol) and lipid levels.  $\alpha$ -tocopherol strongly correlated with higher triglyceride and LDL-C levels [effect = 35% ( $P = 8 \times 10^{-20}$ ) and 16% ( $P = 7 \times 10^{-19}$ ) or 67 and 16 mg/dl, respectively].  $\gamma$ -tocopherol was also correlated with higher triglycerides (effect = 17% higher,  $P = 10^{-17}$ ) and LDL-C (6% higher,  $P = 3 \times 10^{-14}$ ) levels, but also with lower HDL-C (effect = −2%,  $P = 6 \times 10^{-6}$ ). Tocopherols are highly lipophilic and their absorption is enhanced by

triglycerides, though both were significant despite controlling for BMI and waist circumference.

### Carotenoids: favourable association with HDL-C and triglycerides and unfavourable association with LDL-C

Both isomers of  $\beta$ -carotene, *cis*- and *trans*-, were associated with lower triglyceride levels ( $P = 10^{-6}$ , effect = −7% or 12 mg/dl;  $P = 10^{-8}$ , effect = −10% or 16 mg/dl, respectively). However, both isomers of carotene, in addition to other carotenoids such as  $\beta$ -cryptoxanthin and lycopene, were consistently associated with higher levels of both HDL-C and LDL-C. The effect was 5% ( $P = 3 \times 10^{-12}$ ) and 6% ( $P = 5 \times 10^{-11}$ ) for HDL-C and LDL-C levels, respectively for *cis*- $\beta$ -carotene and 3% ( $P = 10^{-10}$ ) and 12% ( $P = 8 \times 10^{-17}$ ) for lycopene.

**Table 3** Estimates of multivariate linear regression model predicting  $\log_{10}(\text{HDL-C})$  as a function of sex, age, age<sup>2</sup>, ethnicity, SES and BMI for each survey

HDL-C	1999–2000 ( <i>n</i> = 6386)		2001–02 ( <i>n</i> = 7485)	
	Estimate (95% CI)	<i>P</i> -value	Estimate (95% CI)	<i>P</i> -value
Sex (vs male)	0.068 (0.059–0.077)	$3 \times 10^{-5}$	0.076 (0.066–0.087)	$9 \times 10^{-6}$
<b>Age (10 years)</b>				
Age	0.02 (0.0082–0.032)	0.009	0.027 (0.018–0.035)	$4 \times 10^{-4}$
Age <sup>2</sup>	−0.00012 (−0.00022 to $-1.6 \times 10^{-5}$ )	0.03	−0.00018 (−0.00027 to $-8.5 \times 10^{-5}$ )	0.004
<b>Ethnicity (vs White)</b>				
Black	0.047 (0.03–0.064)	0.002	0.048 (0.034–0.061)	$3 \times 10^{-4}$
Mexican–American	0.016 (−0.0019 to 0.034)	0.07	0.0067 (−0.0022 to 0.016)	0.1
Other Hispanic	−0.016 (−0.039 to 0.0069)	0.1	−0.0039 (−0.028 to 0.02)	0.7
Other	0.0073 (−0.027 to 0.042)	0.6	0.0056 (−0.018 to 0.029)	0.6
<b>SES (vs high tertile)</b>				
SES (medium)	−0.0054 (−0.021 to 0.01)	0.4	−0.017 (−0.034 to 0.00065)	0.06
SES (low)	−0.018 (−0.032 to −0.0035)	0.03	−0.03 (−0.048 to −0.012)	0.008
BMI (10 U)	−0.06 (−0.069 to −0.05)	$6 \times 10^{-5}$	−0.063 (−0.071 to −0.056)	$4 \times 10^{-6}$
	2005–06 ( <i>n</i> = 7278)		Validation: 2003–04 ( <i>n</i> = 6969)	
Sex × (vs male)	0.075 (0.063–0.086)	$2 \times 10^{-5}$	0.078 (0.067–0.088)	$8 \times 10^{-6}$
<b>Age (10 years)</b>				
Age	0.016 (0.0065–0.026)	0.008	0.017 (0.0084–0.026)	0.004
Age <sup>2</sup>	−1e−04 (−0.00022 to $1.3 \times 10^{-5}$ )	0.07	−1e−04 (−0.00019 to $-1.3 \times 10^{-5}$ )	0.03
<b>Ethnicity (vs White)</b>				
Black	0.042 (0.026 to −0.058)	0.001	0.041 (0.031–0.05)	$1 \times 10^{-4}$
Mexican–American	−0.00017 (−0.011 to 0.011)	1	−0.0048 (−0.019 to 0.0088)	0.4
Other Hispanic	0.0048 (−0.015 to 0.025)	0.6	−0.008 (−0.035 to 0.019)	0.5
Other	−0.0028 (−0.023 to 0.018)	0.7	−0.0039 (−0.033 to 0.025)	0.7
<b>SES (vs high tertile)</b>				
SES (medium)	−0.012 (−0.022 to −0.003)	0.02	−0.0098 (−0.019 to −0.00066)	0.04
SES (low)	−0.023 (−0.035 to −0.012)	0.003	−0.019 (−0.03 to −0.0076)	0.008
BMI (10 U)	−0.061 (−0.07 to −0.053)	$9 \times 10^{-6}$	−0.058 (−0.069 to −0.047)	$4 \times 10^{-5}$

95% CI: 95% confidence interval.

**Favourable lipid correlations with vitamins B, C, D, iron, mercury and enterolactone**

We found serum levels of folate (vitamin B), C, D, iron and mercury to be favourably associated with HDL-C (Figure 4C). Effect sizes of vitamin, iron and mercury levels on HDL-C were similar, ranging from 3 to 4% (1–2 mg/dl) higher HDL-C ( $P < 0.002$ ). Last, we found enterolactone, a product of lignan metabolism in the intestine, to be associated with 10% (17 mg/dl) lower triglyceride levels ( $P = 2 \times 10^{-7}$ ) (Figure 4A).

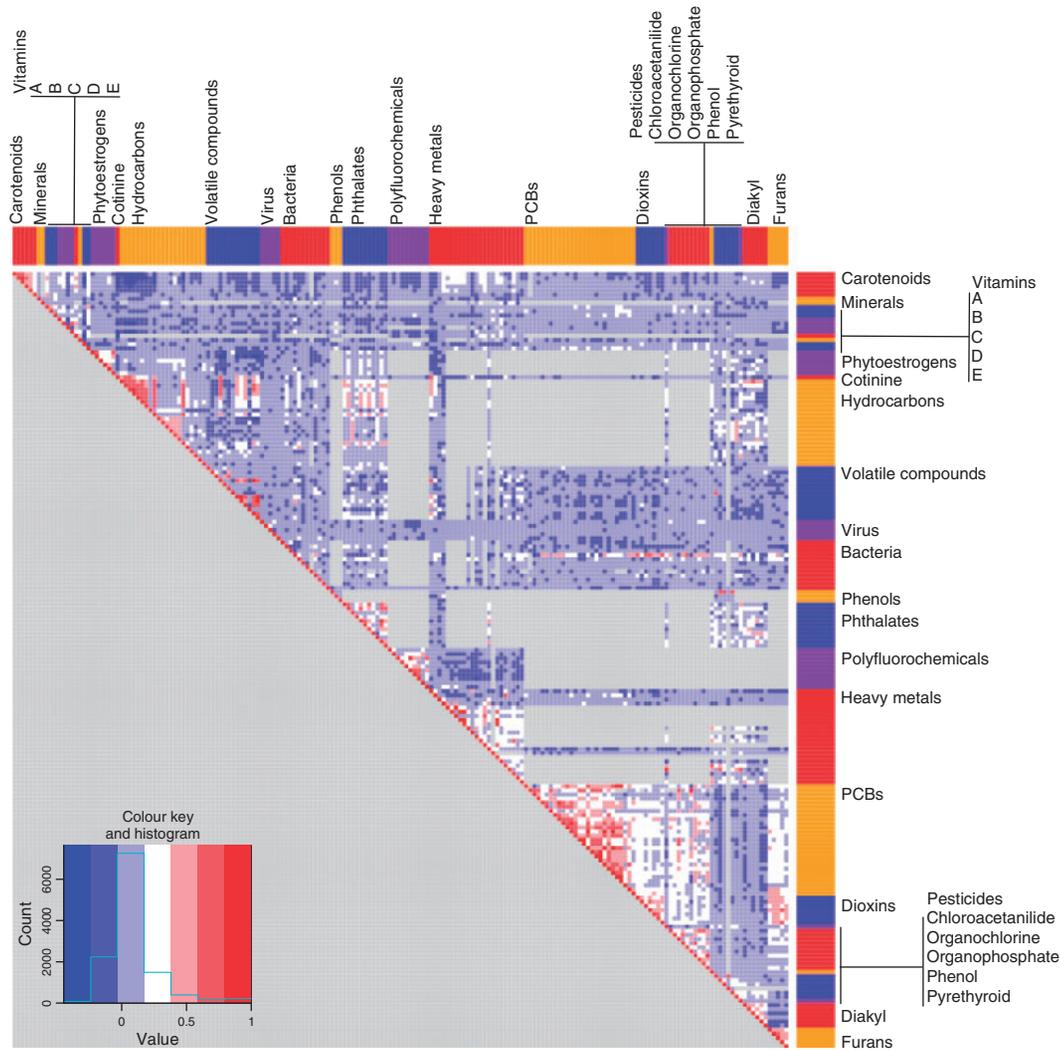
**Persistent pollutants: unfavourable association with triglycerides and HDL-C**

PCBs, dibenzofurans and organochlorine pesticides, all persistent organic pollutants, were unfavourably associated with both triglyceride and HDL-C levels

(Figure 4A and C). Seven PCB factors were tentatively validated and the most significant co-geners PCB74 and PCB170 were associated with 15% ( $P = 10^{-6}$ ) and 19% ( $P = 4 \times 10^{-6}$ ) higher triglyceride levels. Five organochlorine factors were tentatively validated, among which oxychlorodane and *trans*-nonachlor changes were linked to 29 and 30% higher ( $P = 5 \times 10^{-9}$ ,  $1 \times 10^{-8}$ ) triglyceride levels. Another organochlorine pesticide, heptachlor epoxide, was associated with 3% lower HDL-C ( $P = 0.006$ ).

**Markers for air pollution and nicotine: unfavourable association with HDL-C**

Several markers of air pollution and nicotine exposure were unfavourably associated with HDL-C (Figure 4C). The polyaromatic hydrocarbon markers of fluorene,



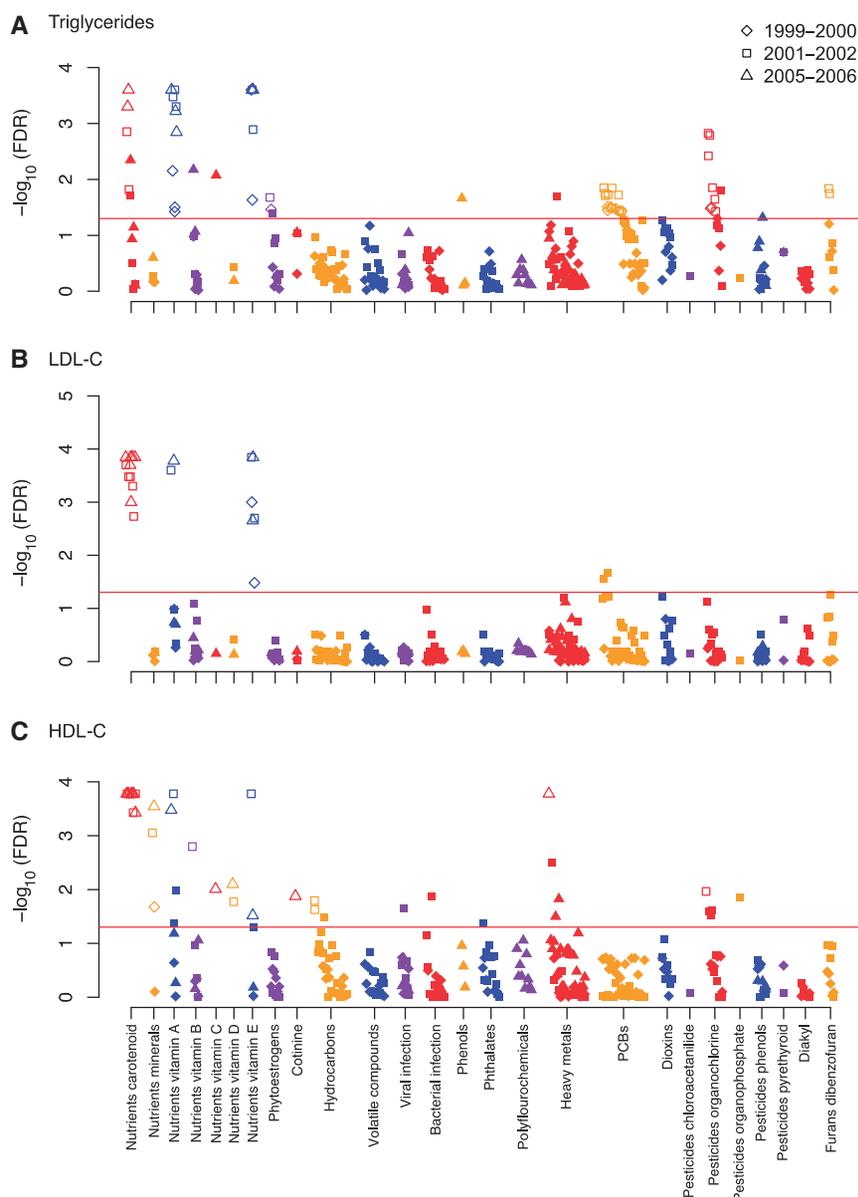
**Figure 2** Partial Pearson's correlation between environmental factors. Partial Pearson's correlation, adjusted by age and BMI (and creatinine for factors measured in urine) for each of the 188 factors were computed for each survey separately. We combined correlations between surveys using a meta-analytic random-effects estimate and displayed them in a heatmap (above), and ordered them by environmental 'class', coloured as in Figure 1A. Pairs of factors where correlations could not be computed are shown in grey

3-hydroxyfluorene and 2-hydroxyfluorene, were associated with 3% lower HDL-C ( $P=0.006$  and  $P=0.004$ ). Cotinine, a serum biomarker for nicotine, was associated also with 3% lower HDL-C ( $P=2 \times 10^{-6}$ ).

### Sensitivity analyses with further adjustments

For most questionnaire variable adjustments, we did not see a sizable difference in estimated coefficients or  $P$ -values for the environmental factors (Supplementary Figures 2–4, available as Supplementary Data at *IJE* online), including questionnaire items regarding self-reported cardiovascular-related disease status and use of drugs. Interestingly, some adjustments increased the effect size of the environmental factor. For example, the association of cotinine, 3- and 2-hydroxyfluorene with HDL-C strengthened after

adjustment for alcohol intake. Adjustment for fish and shellfish consumption strengthened the association between retinyl stearate and HDL-C and triglyceride levels. Conversely, the effect of vitamin C and folate in relation to HDL-C decreased when taking supplement count, total fiber intake and physical activity into account. Adjusting for supplement count decreased the effect of  $\gamma$ -tocopherol on HDL-C. Simultaneous adjustment for self-reported cardiovascular-related disease, supplement count, lipid-lowering drugs and physical activity strengthened the association between tocopherols and pollutant factors and triglycerides, while attenuating the association to  $\alpha$ -carotene (Supplementary Figure 2 available as Supplementary Data at *IJE* online). For HDL-C levels, effects of cotinine, mercury, 3- and 2-flourene, folate, vitamin C, vitamin D



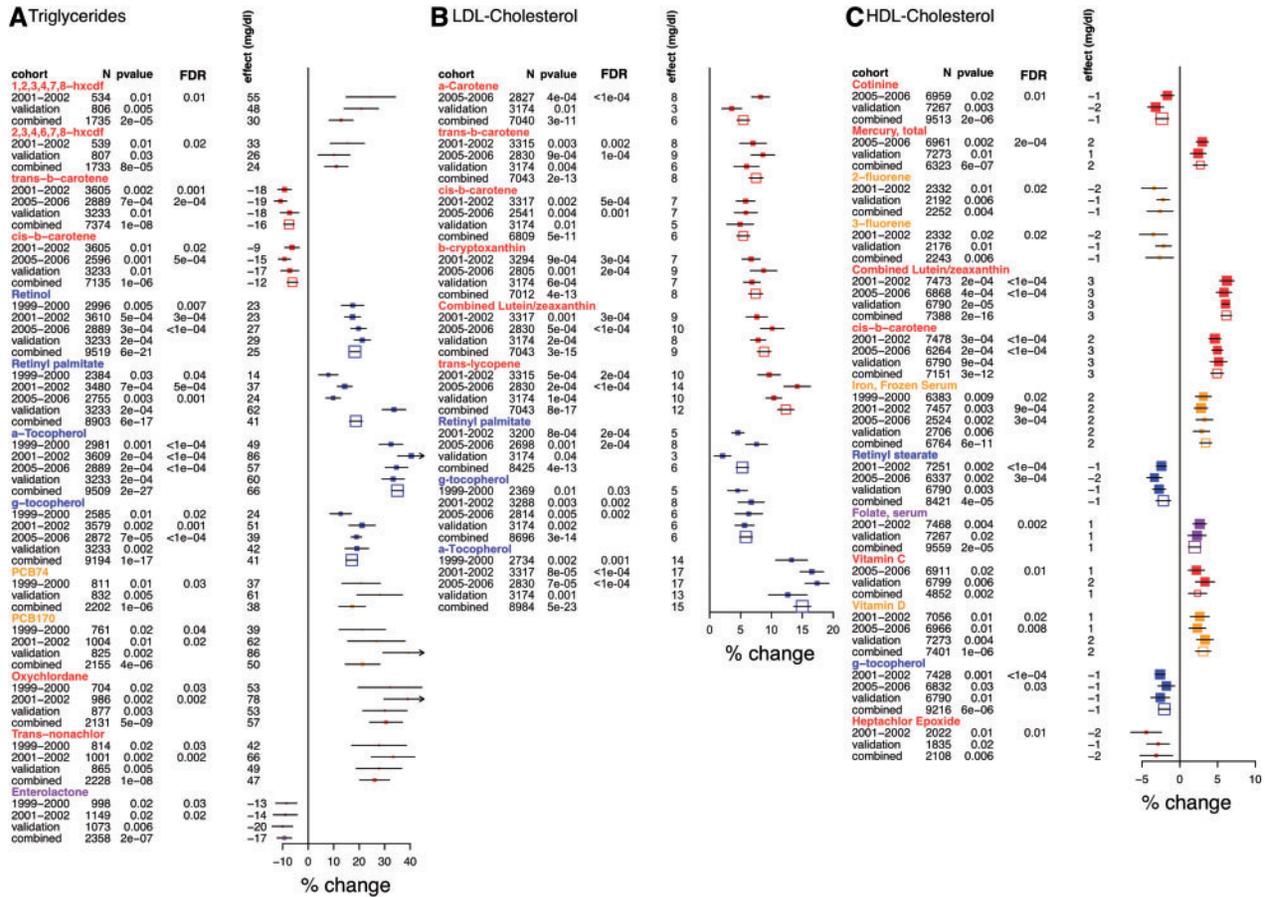
**Figure 3** Significance of association [ $-\log_{10}(\text{FDR})$ ] for each of 188 factors by survey in association to (A) triglycerides, (B) LDL-C, (C) HDL-C. Y-axis indicates  $-\log_{10}(\text{FDR})$  of the adjusted linear regression coefficient for each of the environmental factors. Colours represent different environmental classes as represented in Figure 1A. Red line corresponds to an FDR of 0.05. Findings validated in the 2003–04 survey are seen in the open markers

and  $\gamma$ -tocopherol were all attenuated >15% (Supplementary Figure 4 available as Supplementary data at *IJE* online). However, the direction and significance of the effects were preserved throughout.

## Discussion

By combing through a large number of environmental exposures using a systematic approach, we have found and validated multiple previously known environmental chemical factors correlated with serum lipid levels beyond the level of false discovery. Populations are

exposed to many environmental factors, both harmful and beneficial. It is possible that by studying a few of these factors, we may miss major factors that truly influence disease. Further, by examining multiple factors, we may capture the relative effect of different factors as compared with others. This approach gives a broader, inclusive perspective of benefits and harms that may enhance the interpretation and overall public health relevance of this literature. Such an investigation is made possible by health survey data assaying multiple environmental factors; these surveys are critical to understanding their relationship with characteristics in the general population.<sup>24</sup>



**Figure 4** Forest plots for validated environmental factors associated with (A) triglycerides, (B) LDL-C, (C) HDL-C. Survey (labelled as 1999–2000, 2001–02, 2005–06, filled points) denotes the NHANES survey in which the specific factor was found to be significant (FDR < 0.05) in a model adjusting for age, age<sup>2</sup>, SES, ethnicity, sex, BMI. ‘Validation’ indicates the estimates found for the significant factor in the validation survey. Combined survey (unfilled points) denotes the estimate attained when combining all surveys available for exposure in a model adjusting for age, age-squared, SES, ethnicity, sex, BMI, waist circumference, type 2 diabetes status, blood pressure and survey. Percent change (x-axis) is the percent change of lipid level for a change in 1 SD of logged exposure value. Effect size (in mg/dl) attained when fitting the untransformed lipids to the model. Symbols proportional to sample size and colours represent different environmental classes as represented in Figure 1A. For triglycerides and HDL-C, only the top most significant factors for each factor class is shown; forest plots of all validated factors are seen in Supplementary Figure 1, available as Supplementary Data at *IJE* online

By using transparent reporting and estimation of the FDR, this approach bypasses the problem of selectively testing and reporting one or a few associations at a time that has been debated as a source of biased results and false positives in epidemiological studies.<sup>19,20,22,28,36</sup> We use the breadth of environmental factor and phenotypic measures to conduct extensive sensitivity and correlation analyses that are critical given the complex physiological web of correlation apparent in environmental epidemiological study.<sup>19</sup> Relatedly, such a systematic display of a large number of associations (Figure 3) may enable us to create hypotheses regarding how multiple chemical factors might jointly contribute to phenotypic states.<sup>37</sup> While we have focused here on modelling main effects, a next analytical step might include evaluating how mixtures of environmental factors are connected to lipid levels. However, assessing interactions between

environmental factors will add another layer of complexity, potentially requiring more power for the study. We acknowledge that the approach has drawbacks. First, in our current scenario, some factors were present in more surveys than others and, therefore, have additional opportunity for tentative validation (defined as an FDR of < 5% in test cohorts and  $P < 0.05$  in validation cohort), potentially leading to a bias in factors found. Secondly, the method calls for multiple testing on different types of factors without consideration of priors and a strict FDR threshold is applied, giving way to the possibility of false negatives.<sup>38,39</sup> Nevertheless, just as systematic genome-wide studies have had utility in finding novel genetic loci associated with complex disease,<sup>16</sup> this EWAS strategy provides an opportunity to find novel markers of exposure and prioritize their validation in follow-up studies.

Our findings reveal complex relationships between serum lipid levels and fat-soluble antioxidant vitamins A and E and carotenoids. Randomized studies and meta-analyses<sup>40–44</sup> have shown these vitamins to have no benefits or even confer harm when given in high doses, in contrast to previous favourable associations in observational studies.<sup>45,46</sup> The unfavourable lipid profile that we observed with vitamin E forms is consistent with observational data, and possibly consistent with the randomized evidence on clinical outcomes.<sup>47</sup>

We observed an association of vitamins B (folate), C and D, mercury and iron, to higher HDL-C levels. Folate<sup>48</sup> and vitamin D<sup>49</sup> have previously been associated with higher HDL-C. Fish, a source of cardioprotective omega-3 fatty acids, are also a large source of mercury;<sup>50</sup> however, we did not observe a large change in effect size of mercury when accounting for consumption of fish. These nutrients and metals may be to some extent surrogate markers of 'healthy diet' behaviours; however, what exactly constitutes a 'healthy diet' is currently very difficult to define, in contrast to earlier claims.<sup>51,52</sup> The strength of the association for these dietary markers is similar on HDL-C, ranging from 1 to 3 mg/dl for a standardized change per factor. These are small effects and it is unclear whether cumulatively they could have a much larger impact in raising HDL-C level, given the correlations between these markers (Figure 2).

We also identified enterolactone to be strongly associated with favourable triglyceride levels in this study. Enterolactone is a metabolite of lignans, which are found in foods such as flaxseed and have been associated with favourable cholesterol profiles in this form.<sup>53,54</sup> Again, it is unclear what role, if any, this marker plays as a surrogate of 'healthy diets' and effects on heart disease have been inconsistent.<sup>55</sup>

We found markers of hydrocarbons, 2- and 3-hydroxyfluorene to be strongly associated with unfavourable HDL-C levels. Although others have shown the association of these metabolites to self-report cardiovascular disease with the NHANES data,<sup>56</sup> to our knowledge the association with HDL-C is novel. Relatedly, we also found a marker of nicotine, cotinine, to have a similar association with HDL-C. Particulate matter air pollution, composed of many types of hydrocarbons and smoking long have been a major concern for cardiovascular-related diseases.<sup>15,57,58</sup> It is well-known that smoking influences HDL-C levels<sup>59,60</sup> and acute and chronic exposures to tobacco smoke have been shown to decrease HDL-C substantially.<sup>61</sup> The high correlation of the hydrocarbon markers to cotinine suggests that these associations might all indicate exposure to cigarette smoke.

We also have reconfirmed the correlation between banned-use persistent pollutants, such as organochlorine pesticides, dibenzofurans and polychlorinated biphenyls, with adverse lipid profiles, such as large increase of triglycerides and large decrease in HDL-C. These environmental factors have already been implicated in other metabolic-related and

cardiovascular diseases and among several populations. For example, PCB170 and heptachlor epoxide have been associated with type 2 diabetes and hypertension in these surveys.<sup>25,62</sup> Similarly, PCBs and dibenzofurans have been associated with metabolic syndrome in a Japanese population.<sup>63</sup>

We acknowledge that these associations might be confounded due to the fat solubility of these pollutants. Nevertheless, there have been efforts to elucidate causal relationships using different analytic methods and ecological data. For example, in a recent study considering causal pathways and confounding bias via structural equation modeling, investigators found a relationship between polychlorinated biphenyls and lipid levels consistent with forward causality for a native population with high exposure of these pollutants in upstate New York.<sup>8</sup> Another study found an ecological relationship between cardiovascular-related hospitalization rates in areas close to PCB pollution.<sup>9</sup> Nonetheless, the etiological relationships between persistent pollutants and the metabolic syndrome, type 2 diabetes and cardiovascular diseases remains elusive.<sup>64,65</sup> However, current etiological speculation includes the role of these pollutants interfering with PPARs, transcription factors known to be involved in lipid homeostasis,<sup>66</sup> and/or influencing change in DNA methylation.<sup>67,68</sup> Persistent pollutants were recently associated with atherosclerosis in the elderly in Sweden, independent of serum lipid levels, suggesting a direct pollutant effect on atherosclerosis.<sup>69</sup> Further investigation of these pollutants and consideration of other phenotypes along the causal pathway for cardiovascular-related diseases is warranted.

Elucidating both influence of persistent pollutants on lipids and quantifying their amount in serum lipids remains an issue of debate.<sup>70,71</sup> For example, there are methods to quantify persistent pollutants via adjustment with serum lipids;<sup>72,73</sup> but differing methods of adjustment of these factors could lead to conflicting results.<sup>70</sup> Porta *et al.*,<sup>71</sup> in investigating the influence of organochlorine pesticides on pancreatic carcinoma, indicate that linear adjustment may be inappropriate in some cases. Assessments between persistent pollutants and serum lipid levels as described here may address some of these issues.

Factor variability must be characterized to ensure their adequate analytical modelling. For example, we considered BMI as having a confounding role and included it as an adjustment in our models. However, BMI may lie on the causal pathway towards adverse lipid profiles. The inter-relationship among lipid levels, BMI and persistent pollutant factors is complex,<sup>74</sup> differing in the context of sex and demographics,<sup>75</sup> clinical characteristics<sup>76</sup> and after changes in weight (e.g. after bariatric surgery<sup>77</sup> and overall weight gain<sup>78</sup>). Inter-individual differences in pharmacokinetics also play a role in this complex relationship.<sup>79</sup> The choice of how to model adjusting variables ultimately influences inferences. Long-term longitudinal investigations<sup>79,80</sup> and causal

inference methods may be more suitable to understand causal pathways, if any, underlying the correlation of these environmental factors with lipid levels and other phenotypes.

There are some important limitations in a study using cross-sectional measurements and the observed correlations are far from causal. These associations may reflect a complex web of physiological correlation and/or reverse causality. For example,  $\alpha$ -tocopherol and carotenes are transported in serum with HDL and LDL<sup>81–83</sup> and accurate measurement of serum  $\alpha$ -tocopherol is dependent on serum lipids.<sup>84</sup> In this regard, the strong association between  $\alpha$ -tocopherol and LDL and triglycerides might be considered a true positive result. On the other hand, given the lack of evidence for  $\gamma$ -tocopherol or retinol associating with lipoprotein complexes, their association might be due to reverse causality, or increased anti-oxidant consumption among those who know about their adverse lipid level profile. However, given that vitamin E consumption has been found to increase mortality in meta-analysis,<sup>44</sup> the large effect sizes suggest that prospective studies may be scrutinized for any potentially adverse effects of vitamin E on lipid levels and other metabolic disorders, such as type 2 diabetes.<sup>25</sup>

Like vitamins, we must consider how the distribution of persistent pollutants among biomolecules in serum may influence our analyses. Persistent pollutants have a unique signature in plasma LDL and HDL; for example, PCBs are primarily carried in LDL whereas their metabolites are evenly carried in both LDL and HDL.<sup>85</sup> Furthermore, ascertaining levels of pollutants found in tissue other than serum may be eventually required to understand pathology. To this end, there are reports of concordance between concentrations of persistent pollutants found in different adipose tissues, such as between breast and abdominal adipose tissue.<sup>86</sup> There appears to be concordance between levels of Dichlorodiphenyldichloroethylene (DDE) found in serum and breast adipose tissue;<sup>87</sup> however, relative estimation varies based on the type of adjustment methods used.

Another issue includes the measurement of pollutant environmental factors themselves. For example, limits of detection varied across different NHANES surveys. To address this, we filtered out variables that had a majority of undetectable measurements;

however, results may be biased due to imbalance in measurement techniques and differing thresholds. In the future, factor measurement should be standardized, as proposed by the PhenX project, to ensure comparability of results among different studies and cohorts.<sup>88</sup> Environmental exposure biomonitoring data from other public health surveys might be able to aid in this effort and the National Academy of Science Committee on Human Biomonitoring for Environmental Toxicants lists examples of such efforts.<sup>23</sup>

Despite these limitations, we have shown here a systematic approach to create robust hypotheses regarding association of environmental factors with disease. Further studies should focus on elucidating their role in disease, if any.

## Supplementary Data

Supplementary data are available at *IJE* online.

## Funding

The National Library of Medicine (grant numbers T15 LM 007033 and R01 LM009719); National Institute of General Medical Sciences (grant number R01 GM079719); National Institutes of Health Clinical and Translational Science Award (UL1 RR025744); Lucile Packard Foundation for Children's Health, Howard Hughes Medical Institute.

## Acknowledgements

We thank the NHANES participants, the National Center for Health Statistics and the Centers for Disease Control and Prevention. We thank the Hewlett Packard Foundation for computer infrastructure and Alex Skrenchuk and Boris Oskotsky, PhD, of Stanford Biomedical Informatics Research (Stanford University School of Medicine) for computer support. All authors conceived of the study. C.J.P conducted the analyses. All authors participated in interpretation of results and writing the manuscript.

**Conflict of interest:** None declared.

### KEY MESSAGES

- Unlike current-day genomic studies, systematic studies to dissect the connection between environmental factors and disease risk factors, such as serum lipid levels, are lacking.
- Here, we systematically correlate serum and urine biomarkers of environmental factors with cholesterol and triglyceride levels, uncovering a spectrum of persistent pollutants and nutrients connected with the risk factors.
- Such an investigation is one way to prioritize factors for follow-up validation study.

## References

- 1 Heller DA, de Faire U, Pedersen NL, Dahlen G, McClearn GE. Genetic and environmental influences on serum lipid levels in twins. *N Engl J Med* 1993;**328**: 1150–56.
- 2 Costanza MC, Cayanis E, Ross BM *et al.* Relative contributions of genes, environment, and interactions to blood lipid concentrations in a general adult population. *Am J Epidemiol* 2005;**161**:714–24.
- 3 Kris-Etherton PM, Krummel D, Russell ME *et al.* The effect of diet on plasma lipids, lipoproteins, and coronary heart disease. *J Am Diet Assoc* 1988;**88**:1373–400.
- 4 Schaefer EJ. Lipoproteins, nutrition, and heart disease. *Am J Clin Nutr* 2002;**75**:191–212.
- 5 Varady KA, Jones PJ. Combination diet and exercise interventions for the treatment of dyslipidemia: an effective preliminary strategy to lower cholesterol levels? *J Nutr* 2005;**135**:1829–35.
- 6 Craig WY, Palomaki GE, Haddow JE. Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data. *BMJ* 1998;**317**:784–88.
- 7 Kraus WE, Houmard JA, Duscha BD *et al.* Effects of the amount and intensity of exercise on plasma lipoproteins. *N Engl J Med* 2002;**347**:1483–92.
- 8 Goncharov A, Haase RF, Santiago-Rivera A *et al.* High serum PCBs are associated with elevation of serum lipids and cardiovascular disease in a Native American population. *Environ Res* 2008;**106**:226–39.
- 9 Sergeev AV, Carpenter DO. Hospitalization rates for coronary heart disease in relation to residence near areas contaminated with persistent organic pollutants and other pollutants. *Environ Health Perspect* 2005;**113**: 756–61.
- 10 Lee DH, Steffes MW, Sjodin A, Jones RS, Needham LL, Jacobs DR Jr. Low dose organochlorine pesticides and polychlorinated biphenyls predict obesity, dyslipidemia, and insulin resistance among people free of diabetes. *PLoS One* 2011;**6**:e15977.
- 11 Lee DH, Lee IK, Porta M, Steffes M, Jacobs DR Jr. Relationship between serum concentrations of persistent organic pollutants and the prevalence of metabolic syndrome among non-diabetic adults: results from the National Health and Nutrition Examination Survey 1999–2002. *Diabetologia* 2007;**50**:1841–51.
- 12 Lee DH, Steffes MW, Sjodin A, Jones RS, Needham LL, Jacobs DR Jr. Low dose of some persistent organic pollutants predicts type 2 diabetes: a nested case-control study. *Environ Health Perspect* 2010;**118**:1235–42.
- 13 Lee DH, Lee IK, Song K *et al.* A strong dose-response relation between serum concentrations of persistent organic pollutants and diabetes: results from the National Health and Examination Survey 1999–2002. *Diabetes Care* 2006;**29**:1638–44.
- 14 Park SK, Son HK, Lee SK *et al.* Relationship between serum concentrations of organochlorine pesticides and metabolic syndrome among non-diabetic adults. *J Prev Med Public Health* 2010;**43**:1–8.
- 15 Brook RD, Rajagopalan S, Pope CA III *et al.* Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation* 2010;**121**:2331–78.
- 16 Teslovich TM, Musunuru K, Smith AV *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;**466**:707–13.
- 17 Kavvoura FK, Liberopoulos G, Ioannidis JP. Selection in reported epidemiological risks: an empirical assessment. *PLoS Med* 2007;**4**:e79.
- 18 Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;**19**:640–48.
- 19 Ioannidis J, Loy EY, Poulton R, Chia KS. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci Transl Med* 2009;**1**:7ps8.
- 20 Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst* 2008;**100**:988–95.
- 21 Dwan K, Altman DG, Arnaiz JA *et al.* Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008;**3**:e3081.
- 22 Young SS. Acknowledge and fix the multiple testing problem. *Int J Epidemiol* 2010;**39**:934; author reply 934–35.
- 23 Committee on Human Biomonitoring for Environmental Toxicants. *Human Biomonitoring for Environmental Chemicals*. Washington, DC: National Academies Press, 2006.
- 24 Porta M, Puigdomenech E, Ballester F *et al.* Monitoring concentrations of persistent organic pollutants in the general population: the international experience. *Environ Int* 2008;**34**:546–61.
- 25 Patel C, Bhattacharya J, Butte A. A Environment-wide Association Study (EWAS) to type 2 diabetes. *PLoS One* 2010;**5**:e10746.
- 26 Centers for Disease Control and Prevention (CDC). *National Health and Nutrition Examination Survey, 2009*. <http://www.cdc.gov/nchs/nhanes/> (9 January 2011, date last accessed).
- 27 Wang L, Gaziano JM, Norkus EP, Buring JE, Sesso HD. Associations of plasma carotenoids with risk factors and biomarkers related to cardiovascular disease in middle-aged and older women. *Am J Clin Nutr* 2008;**88**: 747–54.
- 28 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;**2**:e124.
- 29 Ezzati TM, Massey JT, Waksberg J, Chu A, Maurer KR. Sample design: Third National Health and Nutrition Examination Survey. *Vital Health Stat 2* 1992;**113**:1–35.
- 30 Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972;**18**:499–502.
- 31 Lumley T. *Survey: analysis of complex survey samples*. 3.14 ed, 2009.
- 32 Storey JD. A direct approach to False Discovery Rates. *J R Statist Soc B* 2002;**64**:479–98.
- 33 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. Hillsdale, NJ: Erlbaum, 1988.
- 34 Fryar CD, Hirsch R, Eberhardt MS, Yoon SS, Wright JD. Hypertension, high serum total cholesterol, and diabetes: racial and ethnic prevalence differences in U.S. adults, 1999–2006. *NCHS Data Brief* 2010:1–8.
- 35 Ford ES, Li C, Zhao G, Pearson WS, Mokdad AH. Hypertriglyceridemia and Its pharmacologic treatment among US Adults. *Arch Intern Med* 2009;**169**:572–78.
- 36 Morgan TM, Krumholz HM, Lifton RP, Spertus JA. Nonvalidation of reported genetic risk factors for acute

- coronary syndrome in a large-scale replication study. *J Am Med Assoc* 2007;**297**:1551–61.
- <sup>37</sup> Sexton K, Hattis D. Assessing cumulative health risks from exposure to environmental mixtures - three fundamental questions. *Environ Health Perspect* 2007;**115**: 825–32.
- <sup>38</sup> Blair A, Saracci R, Vineis P *et al*. Epidemiology, public health, and the rhetoric of false positives. *Environ Health Perspect* 2009;**117**:1809–13.
- <sup>39</sup> Fallin MD, Kao WHL. Is 'X'-WAS the future for all of epidemiology? *Epidemiology* 2011;**22**:457–59.
- <sup>40</sup> Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P. Vitamin E supplementation and cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med* 2000;**342**: 154–60.
- <sup>41</sup> Omenn GS, Goodman GE, Thornquist M, Brunzell JD. Long-term vitamin A does not produce clinically significant hypertriglyceridemia: results from CARET, the beta-carotene and retinol efficacy trial. *Cancer Epidemiol Biomarkers Prev* 1994;**3**:711–13.
- <sup>42</sup> Redlich CA, Chung JS, Cullen MR, Blaner WS, Van Bennekum AM, Berglund L. Effect of long-term beta-carotene and vitamin A on serum cholesterol and triglyceride levels among participants in the Carotene and Retinol Efficacy Trial (CARET). *Atherosclerosis* 1999;**145**:425–32.
- <sup>43</sup> Vivekananthan DP, Penn MS, Sapp SK, Hsu A, Topol EJ. Use of antioxidant vitamins for the prevention of cardiovascular disease: meta-analysis of randomised trials. *Lancet* 2003;**361**:2017–23.
- <sup>44</sup> Miller ER 3rd, Pastor-Barriuso R, Dalal D, Riemersma RA, Appel LJ, Guallar E. Meta-analysis: high-dosage vitamin E supplementation may increase all-cause mortality. *Ann Intern Med* 2005;**142**:37–46.
- <sup>45</sup> Mente A, de Koning L, Shannon HS, Anand SS. A systematic review of the evidence supporting a causal link between dietary factors and coronary heart disease. *Arch Intern Med* 2009;**169**:659–69.
- <sup>46</sup> Willcox BJ, Curb JD, Rodriguez BL. Antioxidants in cardiovascular health and disease: key lessons from epidemiologic studies. *Am J Cardiol* 2008;**101**:75D–86D.
- <sup>47</sup> White E, Kristal AR, Shikany JM *et al*. Correlates of serum alpha and gamma-tocopherol in the women's health initiative. *Ann Epidemiol* 2001;**11**:136–44.
- <sup>48</sup> Semmler A, Moskau S, Grigull A *et al*. Plasma folate levels are associated with the lipoprotein profile: a retrospective database analysis. *Nutr J* 2010;**9**:31.
- <sup>49</sup> Jorde R, Figenschau Y, Hutchinson M, Emaus N, Grimnes G. High serum 25-hydroxyvitamin D concentrations are associated with a favorable serum lipid profile. *Eur J Clin Nutr* 2010;**64**:1457–64.
- <sup>50</sup> Smith KM, Barraj LM, Kantor M, Sahyoun NR. Relationship between fish intake, n-3 fatty acids, mercury and risk markers of CHD (National Health and Nutrition Examination Survey 1999–2002). *Public Health Nutr* 2009;**12**:1261–69.
- <sup>51</sup> Hu FB, Willett WC. Optimal diets for prevention of coronary heart disease. *J Am Med Assoc* 2002;**288**: 2569–78.
- <sup>52</sup> Joshipura KJ, Hu FB, Manson JE *et al*. The effect of fruit and vegetable intake on risk for coronary heart disease. *Ann Intern Med* 2001;**134**:1106–14.
- <sup>53</sup> Bassett CM, Rodriguez-Leyva D, Pierce GN. Experimental and clinical research findings on the cardiovascular benefits of consuming flaxseed. *Appl Physiol Nutr Metab* 2009;**34**:965–74.
- <sup>54</sup> Pan A, Yu D, Demark-Wahnefried W, Franco OH, Lin X. Meta-analysis of the effects of flaxseed interventions on blood lipids. *Am J Clin Nutr* 2009;**90**:288–97.
- <sup>55</sup> Park D, Huang T, Frishman WH. Phytoestrogens as cardioprotective agents. *Cardiol Rev* 2005;**13**:13–17.
- <sup>56</sup> Xu X, Cook RL, Ilacqua VA, Kan H, Talbott EO, Kearney G. Studying associations between urinary metabolites of polycyclic aromatic hydrocarbons (PAHs) and cardiovascular diseases in the United States. *Sci Total Environ* 2010;**408**:4943–48.
- <sup>57</sup> Pope CA III, Burnett RT, Thurston GD *et al*. Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease. *Circulation* 2004;**109**:71–77.
- <sup>58</sup> Miller KA, Siscovick DS, Sheppard L *et al*. Long-term exposure to air pollution and incidence of cardiovascular events in women. *N Engl J Med* 2007;**356**:447–58.
- <sup>59</sup> Wilson PW, Garrison RJ, Abbott RD, Castelli WP. Factors associated with lipoprotein cholesterol levels. The Framingham study. *Arteriosclerosis* 1983;**3**:273–81.
- <sup>60</sup> Njolstad I, Arnesen E, Lund-Larsen PG. Smoking, serum lipids, blood pressure, and sex differences in myocardial infarction. A 12-year follow-up of the Finnmark Study. *Circulation* 1996;**93**:450–56.
- <sup>61</sup> Moffatt RJ, Chelland SA, Pecott DL, Stamford BA. Acute exposure to environmental tobacco smoke reduces HDL-C and HDL2-C. *Prev Med* 2004;**38**:637–41.
- <sup>62</sup> Everett CJ, Mainous AG 3rd, Frithsen IL, Player MS, Matheson EM. Association of polychlorinated biphenyls with hypertension in the 1999–2002 National Health and Nutrition Examination Survey. *Environ Res* 2008;**108**:94–97.
- <sup>63</sup> Uemura H, Arisawa K, Hiyoshi M *et al*. Prevalence of metabolic syndrome associated with body burden levels of dioxin and related compounds among Japan's general population. *Environ Health Perspect* 2009;**117**: 568–73.
- <sup>64</sup> Longnecker MP, Daniels JL. Environmental contaminants as etiologic factors for diabetes. *Environmental Health Perspectives* 2001;**109**(Suppl 6):871–76.
- <sup>65</sup> Porta M. Persistent organic pollutants and the burden of diabetes. *Lancet* 2006;**368**:558–59.
- <sup>66</sup> Remillard RB, Bunce NJ. Linking dioxins to diabetes: epidemiology and biologic plausibility. *Environ Health Perspect* 2002;**110**:853–58.
- <sup>67</sup> Lee D-H, Jacobs DR, Miquel P. Hypothesis: a unifying mechanism for nutrition and chemicals as lifelong modulators of DNA hypomethylation. *Environ Health Perspect* 2009;**117**:1799–802.
- <sup>68</sup> Rusiecki JA, Baccarelli A, Bollati V, Tarantini L, Moore LE, Bonefeld-Jorgensen EC. Global DNA hypomethylation is associated with high serum-persistent organic pollutants in Greenlandic Inuit. *Environ Health Perspect* 2008;**116**:1547–52.
- <sup>69</sup> Lind PM, van Bavel B, Salihovic S, Lind L. Circulating levels of Persistent Organic Pollutants (POPs) and carotid atherosclerosis in the elderly. *Environ Health Perspect* 2011;**120**:38–43.
- <sup>70</sup> Schisterman EF, Whitcomb BW, Louis GM, Louis TA. Lipid adjustment in the analysis of environmental contaminants and human health risks. *Environ Health Perspect* 2005;**113**:853–57.

- <sup>71</sup> Porta M, Jarrod M, Lopez T *et al.* Correcting serum concentrations of organochlorine compounds by lipids: alternatives to the organochlorine/total lipids ratio. *Environ Int* 2009;**35**:1080–85.
- <sup>72</sup> Covaci A, Voorspoels S, Thomsen C, van Bavel B, Neels H. Evaluation of total lipids using enzymatic methods for the normalization of persistent organic pollutant levels in serum. *Sci Total Environ* 2006;**366**:361–66.
- <sup>73</sup> Bernert JT, Turner WE, Patterson DG Jr, Needham LL. Calculation of serum “total lipid” concentrations for the adjustment of persistent organohalogen toxicant measurements in human samples. *Chemosphere* 2007;**68**:824–31.
- <sup>74</sup> Dirinck E, Jorens PG, Covaci A *et al.* Obesity and persistent organic pollutants: possible obesogenic effect of organochlorine pesticides and polychlorinated biphenyls. *Obesity* 2011;**19**:709–14.
- <sup>75</sup> James RA, Hertz-Picciotto I, Willman E, Keller JA, Charles MJ. Determinants of serum polychlorinated biphenyls and organochlorine pesticides measured in women from the child health and development study cohort, 1963–1967. *Environ Health Perspect* 2002;**110**:617–24.
- <sup>76</sup> Porta M, Gasull M, Puigdomenech E *et al.* Distribution of blood concentrations of persistent organic pollutants in a representative sample of the population of Catalonia. *Environ Int* 2010;**36**:655–64.
- <sup>77</sup> Hue O, Marcotte J, Berrigan F *et al.* Increased plasma levels of toxic pollutants accompanying weight loss induced by hypocaloric diet or by bariatric surgery. *Obes Surg* 2006;**16**:1145–54.
- <sup>78</sup> Wolff MS, Britton JA, Teitelbaum SL *et al.* Improving organochlorine biomarker models for cancer research. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:2224–36.
- <sup>79</sup> Wolff MS, Anderson HA, Britton JA, Rothman N. Pharmacokinetic variability and modern epidemiology—the example of dichlorodiphenyltrichloroethane, body mass index, and birth cohort. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:1925–30.
- <sup>80</sup> Mendez MA, Garcia-Esteban R, Guxens M *et al.* Prenatal organochlorine compound exposure, rapid weight gain and overweight in infancy. *Environ Health Perspect* 2011;**119**:272–78.
- <sup>81</sup> Willett WC. *Nutritional Epidemiology*. New York: Oxford University Press, 1998.
- <sup>82</sup> Bender D. *Nutritional Biochemistry of the Vitamins*. Cambridge: University of Cambridge Press, 2003.
- <sup>83</sup> Ogihara T, Miki M, Kitagawa M, Mino M. Distribution of tocopherol among human plasma lipoproteins. *Clin Chim Acta* 1988;**174**:299–305.
- <sup>84</sup> Winbauer AN, Pingree SS, Nuttall KL. Evaluating serum alpha-tocopherol (vitamin E) in terms of a lipid ratio. *Ann Clin Lab Sci* 1999;**29**:185–91.
- <sup>85</sup> Noren K, Weistrand C, Karpe F. Distribution of PCB congeners, DDE, hexachlorobenzene, and methylsulfonyl metabolites of PCB and DDE among various fractions of human blood plasma. *Arch Environ Contam Toxicol* 1999;**37**:408–14.
- <sup>86</sup> Petreas M, Smith D, Hurley S, Jeffrey SS, Gilliss D, Reynolds P. Distribution of persistent, lipid-soluble chemicals in breast and abdominal adipose tissues: lessons learned from a breast cancer study. *Cancer Epidemiol Biomarkers Prev* 2004;**13**:416–24.
- <sup>87</sup> Lopez-Carrillo L, Torres-Sanchez L, Lopez-Cervantes M, Blair A, Cebrian ME, Uribe M. The adipose tissue to serum dichlorodiphenyldichloroethane (DDE) ratio: some methodological considerations. *Environ Res* 1999;**81**:142–45.
- <sup>88</sup> Hamilton CM, Strader LC, Pratt JG *et al.* The PhenX toolkit: get the most from your measures. *Am J Epidemiol* 2011;**174**:253–60.

## Commentary: A step towards more comprehensive analyses of life course effects of mixtures of environmental factors

Miquel Porta,<sup>1,2,3\*</sup> Magda Gasull<sup>1,2,3</sup> and José Pumarega<sup>1,2,3</sup>

<sup>1</sup>Hospital del Mar Research Institute-IMIM, Barcelona, Spain, <sup>2</sup>School of Medicine, Universitat Autònoma de Barcelona, Spain and <sup>3</sup>CIBER en Epidemiología y Salud Pública (CIBERESP), Spain

\*Corresponding author. Hospital del Mar IMIM UAB PRBB, Carrer del Dr. Aiguader 88, E-08003 Barcelona, Catalonia, Spain. E-mail: mporta@imim.es

Accepted 19 January 2012

With its strengths and weaknesses, the analysis ('association study') of the relationships between

environmental factors and lipid concentrations by Patel *et al.*<sup>1</sup> is of great interest for at least five reasons.