



Miscellaneous

Systematic correlation of environmental exposure and physiological and self-reported behaviour factors with leukocyte telomere length

Chirag J Patel,* Arjun K Manrai, Erik Corona, and Isaac S Kohane

Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

*Corresponding author. Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, MA 02115, USA. E-mail: chirag_patel@hms.harvard.edu

Accepted 9 February 2016

Abstract

Background: It is hypothesized that environmental exposures and behaviour influence telomere length, an indicator of cellular ageing. We systematically associated 461 indicators of environmental exposures, physiology and self-reported behaviour with telomere length in data from the US National Health and Nutrition Examination Survey (NHANES) in 1999–2002. Further, we tested whether factors identified in the NHANES participants are also correlated with gene expression of telomere length modifying genes.

Methods: We correlated 461 environmental exposures, behaviours and clinical variables with telomere length, using survey-weighted linear regression, adjusting for sex, age, age squared, race/ethnicity, poverty level, education and born outside the USA, and estimated the false discovery rate to adjust for multiple hypotheses. We conducted a secondary analysis to investigate the correlation between identified environmental variables and gene expression levels of telomere-associated genes in publicly available gene expression samples.

Results: After correlating 461 variables with telomere length, we found 22 variables significantly associated with telomere length after adjustment for multiple hypotheses. Of these variables, 14 were associated with longer telomeres, including biomarkers of polychlorinated biphenyls (PCBs; 0.1 to 0.2 standard deviation (SD) increase for 1 SD increase in PCB level, $P < 0.002$) and a form of vitamin A, retinyl stearate. Eight variables associated with shorter telomeres, including biomarkers of cadmium, C-reactive protein and lack of physical activity. We could not conclude that PCBs are correlated with gene expression of telomere-associated genes.

Conclusions: Both environmental exposures and chronic disease-related risk factors may play a role in telomere length. Our secondary analysis found no evidence of association between PCBs/smoking and gene expression of telomere-associated genes. All correlations between exposures, behaviours and clinical factors and changes in telomere length will require further investigation regarding biological influence of exposure.

Key words: environment-wide association study, exposome, genome-wide association study, telomere length, aging, gene-environment interaction

Key Messages

- Identification of indicators of human physiology, behaviour and environmental exposures that potentially influence telomere length may be important to understanding the biological ageing process.
- In a systematic association study of 461 variables, we found 22 environmental exposures and physiological factors associated with telomere length in participants of the National Health and Nutrition Examination Survey (NHANES), including environmental exposures such as PCBs, vitamin A and cadmium, in addition to physiological indicators such as trunk fat and C-reactive protein.
- Lower pulse rate and higher maximum respiratory rate (VO2Max), indicators of physical fitness, were correlated with longer telomere length.

Introduction

A priority in human biology research is the identification of environmental, behavioural and physiological correlates of biological ageing. A biological ‘hallmark of ageing’¹ includes telomere length. Telomeres are defined as repetitive nucleotide sequences located physically at the end of a chromosome, whose function is to protect the integrity of genomic content (as reviewed in²). However, over time and as a function of cell division, telomeres shorten and the life spans of cells decrease.³ It is hypothesized that factors other than chronological age may play a role in telomere length. For example, genome-wide association studies (GWAS), a way to prioritize genetic variants for further biological investigation, have found genetic variants associated with telomere length (e.g.⁴). As reviewed in⁵, behaviour and environmental exposure factors, such as smoking, obesity, diet and stress, may also play a role in telomere length. It is important to identify potentially modifiable factors that may influence the biological ageing process as indicated by telomere length, for a more complete view of progression of telomere length other than chronological age.

Prioritization and/or hypothesis generation regarding the role of physiological, behavioural and environmental factors in telomere length and ageing phenotypes can benefit from the previously proposed systematic analyses^{6,7} to search for environmental indicators associated with traits and disease. Previous investigations have scanned for associations between environmental exposure, behavioural and clinical factors putatively correlated with health-related phenotypes and outcomes, such as type 2 diabetes,^{8,9} serum lipids,¹⁰ blood pressure,¹¹ metabolic syndrome,¹² endometrial cancer,¹³ all-cause mortality,¹⁴ family income¹⁵ and preterm birth.¹⁶ Here, we search for environmental, behavioural and physiological indicators

correlated with telomere length in a cohort of non-institutionalized participants of the National Health and Nutrition Examination Survey (NHANES), a survey representative of the USA. We claim that such an approach can prioritize physiological, environmental and self-reported indicator factors for future investigation into the causal nature (if any) for identified associations. Second, in a secondary analysis, we investigated whether identified environmental exposures are associated with gene expression changes of genes that regulate telomere length.

Methods

An overview of our methodology is depicted in [Figure 1](#).

Study sample

The study sample for our analysis comes from two independent waves of the continuous National Health and Nutrition Examination Survey (NHANES), the 1999–2000¹⁸ and 2001–02¹⁹ surveys ([Figure 1A](#)), representative of the non-institutionalized population of the USA. Overall, the total samples sizes of the 1999–2000 and 2001–02 surveys were 9965 and 11 032, respectively.

Needham and colleagues have previously described the measurement of mean telomere length (MTL) of the leukocyte DNA from NHANES participants.²⁰ In short, a sample of leukocyte DNA was taken from the NHANES participants. This sample was purified and telomere length was measured with the quantitative real-time polymerase chain reaction (PCR) relative to a standard reference DNA. These participants were a random selection of the population surveyed in 1999–2000 and 2001–02 who had consented to give blood for DNA purification. The total

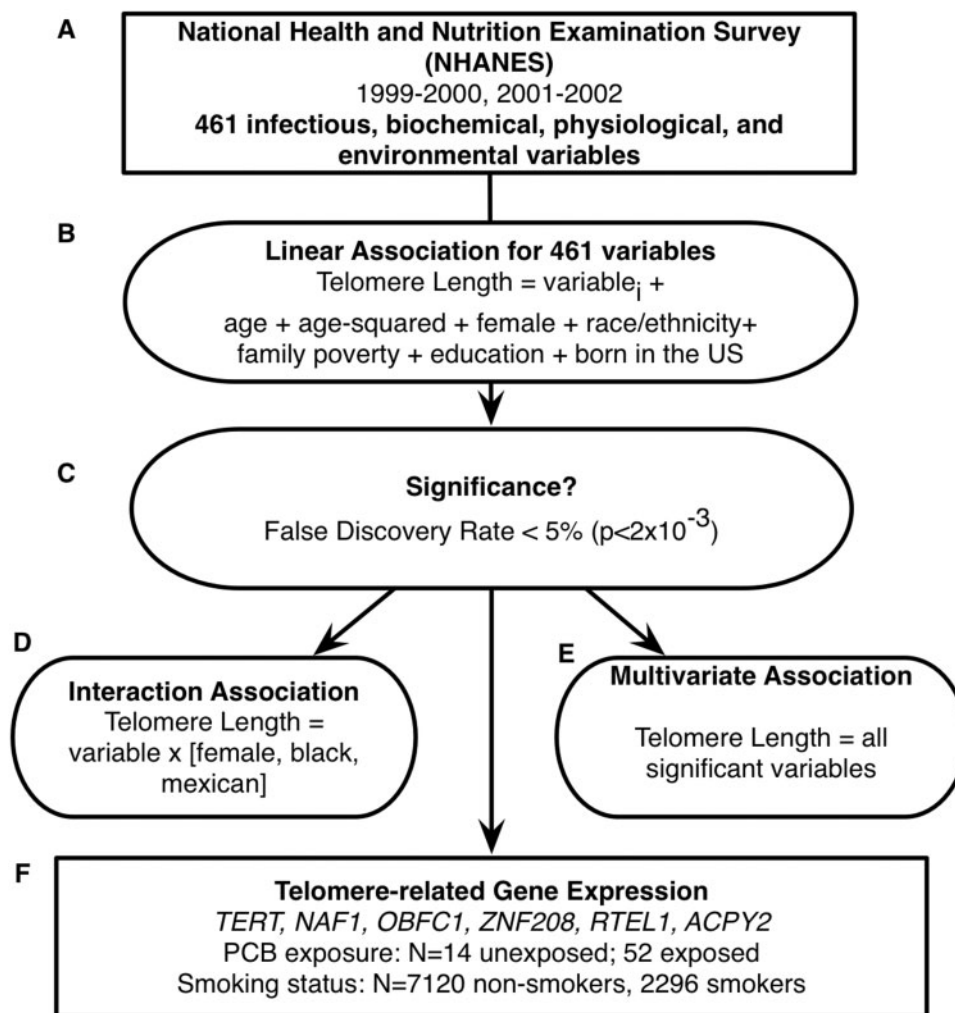


Figure 1. Method to search for physiological, environmental, and behaviour factors associated with mean telomere length (MTL). A) National Health and Nutrition Examination Survey of years 1999–2002. B) Scanning 461 variables for association iteratively in MTL. i denotes i th variable (out of 461) associated with MTL. C) Ascertaining statistical significance via false discovery rate. D) Interaction testing of FDR-significant variables with female sex, black, and Mexican race/ethnicity groups. E) Multivariate linear model predicting MTL as a function of FDR-significant variables. F) Estimating correlations between expression in genes that modulate telomere length in tissue samples exposed to PCB or smoking.

available numbers of individuals with a telomere measurement were 3567 and 4260 in the 1999–2000 and 2001–02 surveys, respectively.

Physiological factors, environmental exposures and self-reported behavioural factors

We assessed all available ($N=461$) indicators of physiological state, environmental exposure and self-reported behavior as potential correlates (referred to as ‘factors’) of telomere length in this investigation (for examples see Table 1; for complete list see Table S1, available as Supplementary data at IJE online).

The 86 physiological factors included: (i) 13 body measures/adiposity; (ii) two on blood pressure; (iii) one on pulse/heart rate; (iv) 20 on blood cell parameters; and (v) 50 on serum/urine-based biochemical measures. The 50

biochemical factors included 12 indicators of kidney function, 8 of liver function, 7 of cardiovascular health, 5 on iron status, 2 on bone health, 2 on inflammation, 4 metabolic indicators, 6 hormone measures, 3 on prostate health, and 1 on nutritional status.

We considered 256 environmental exposure biomarker factors, such as environmental chemicals, indicators of infectious agents, and nutrients assayed from serum and urine. These included 1 serum marker of nicotine metabolism (cotinine), 7 types of dioxins, 10 furans, 7 diacyl metabolites, 20 heavy metals, 23 hydrocarbons, 17 nutrients, 1 phenol, 12 phthalates, 35 polychlorinated biphenyls, 33 pesticides, 6 phytoestrogens, 11 polyfluorochemicals and 28 volatile compounds. We assessed 45 infectious agents factors: 37 bacterial and 8 viral. All continuous biomarkers had a right-skewed distribution and were log-transformed and z-standardized.

Table 1. Examples of environmental, physiological and behavioural variables in NHANES

Category	Variable Type	Number	Examples	
Environmental exposure (serum- or urine-based)	Bacterial infection	37	Urinary <i>Chlamydia</i> (yes/no) <i>H. pylori</i> (log per 1 SD)	
	Cotinine	1	Cotinine (log per 1 SD)	
	Diakyl	7	Urinary dimethylphosphate (log per 1 SD)	
	Dioxins	7	2,3,7,8-tetrachlorodibenzodioxin (log and per 1 SD)	
	Furans	10	2,3,7,8-tetrachlorodibenzofuran (log and per 1 SD)	
	Heavy metals		20	Serum cadmium (log and 1 SD) Serum lead (log and 1 SD)
		Hydrocarbons	23	Urinary 1-hydroxyfluorene (log and per 1 SD)
	Nutrients		17	Serum folate (log and per 1 SD) Serum vitamin D (log and per 1 SD)
		PCBs	35	Serum (polychlorinated biphenyls) PCB170 (log and per 1 SD)
	Pesticides	33	Serum heptachlorepoxyde (log and per 1 SD)	
	Phenols	1	2-isopropoxyphenol (log per 1 SD)	
	Phthalates	12	Mono-n-butyl phthalate (log per 1 SD)	
	Phytoestrogens	6	Urinary enterolactone (log and per 1 SD)	
	Polyfluorochemicals	11	Perfluorodecanoic acid (log per 1 SD)	
	Viral infection	8	Hepatitis B antibody (yes/no)	
	Volatile compounds	28	Blood benzene (log per 1 SD)	
	Total		256	
	Physiological	Biochemistry	44	Urinary creatinine (log per 1 SD) Triglycerides (log per 1 SD)
			Blood	20
Blood pressure		3	Systolic blood pressure (per 1SD)	
Body measures		13	Body mass index (log per 1 SD)	
Hormone		6	Thyroxine (log per 1 SD)	
Total		86		
Self-reported behaviour		Alcohol use	4	Drink 5 in a day (yes/no)
	Food component recall	70	Dietary nutrient intake levels derived from Food Frequency Questionnaire (FFQ) (adjusted for caloric intake)	
		Physical fitness	15	[Health.gov] guideline activity levels from self-reported activity (ordinal) Estimated VO2max (per 1 SD)
	Smoking behaviour	23	Current cigarette smoker (yes/no)	
	Smoking family	4	Does anyone smoke in home (yes/no)?	
	Social support	3	Number of close friends (per 1 SD)	
	Total	119		
Grand total	461			

Further, we considered 119 self-reported or tested assessments of behavioural factors, such as food intake, physical fitness, smoking, alcohol use and social support. Specifically, we considered 70 specific self-reported nutrients based on an in-person interview that used the United States Department of Agriculture and Department of Health and Human Services food recall questionnaires.^{21–23} These food and nutrient consumption factors were linearly adjusted by total caloric intake and z-standardized. We considered 27 questions concerning smoking (e.g. ‘Current smoker?’, ‘Past smoker’). We considered 15 indicators of physical activity; 14 of these indicators were those used to assess the maximal oxygen uptake, such as blood pressure taken during different stages of physical

activity performed on a treadmill. The remaining variable for physical activity was a self-reported indicator, estimated by deriving metabolic equivalents for self-reported leisure and normal-time activities,²⁴ and treated as an ordinal factor based on [Health.gov] physical activity guideline categories as previously described.^{10,25} Last, we considered three and four self-reported indicators of social support and alcohol use, respectively.

Due to NHANES protocol, many variables had a differing number of eligible participants. In the 1999–2000 survey, there were 179 to 3126 eligible participants (median 1046); in 2001–02 there were 131 to 3995 (median 1349) (Table S2, available as [Supplementary data](#) at *IJE* online).

Correlating environmental, physiological, and self-reported factors with telomere length

We associated each of 461 environmental exposure, physiological state, and self-reported behavioural factors with telomere length in the 1999–2000 and 2001–02 survey samples (Figure 1A). Specifically, we modelled each of the 461 factors as an independent variable, and mean telomere length variable as the dependent variable, in 461 separate survey-weighted linear regression models (R survey package²⁶), using weights for the 1999–2000 and 2001–02 surveys as specified by the National Centers for Health Statistics²⁷ (Figure 1B).

We adjusted for sex, race/ancestry (Mexican, Black, Other Hispanic and Other), education (less than high school, high-school equivalent, greater than high school), income-to-poverty ratio (as a continuous variable), born in the USA, age and age squared. Table S2 shows prevalence of each of these adjusting covariates in the population. We acknowledge that our choice of adjustment variables is arbitrary, but our main goal was to find environmental, physiological and self-reported behavioural factors that were independent on race, sex, income-to-poverty ratio (as a measure of socioeconomic status), education and age in telomere length. Our reasoning for choosing these variables was the following. First, as we have shown previously, many environmental exposures are correlated with income-to-poverty ratio.¹⁵ The income-to-poverty ratio is an approximate measure of socioeconomic status and is the household income divided by the national poverty level (assessed at the time of survey). Second, Needham *et al.*²⁰ hypothesized that socioeconomic status is also correlated with telomere length, and we desired to find associations that were independent of socioeconomic status. Last, many environmental exposure and physiological factors are correlated with age; we desired to search for factors that are associated with telomere length independently of age. Furthermore, these adjustments were consistent with previous investigations of MTL in NHANES.²⁸

Further, we conducted two sensitivity analyses to test the estimates in two other modelling scenarios. In model A, we re-estimated associations and *P*-values only adjusting for sex and race/ancestry (removing age, age-squared, income-to-poverty ratio, education and born in the USA). In model B, we re-estimated associations and *P*-values adjusting for sex, age, age squared and race/ancestry. We hypothesized that any differences in association sizes between the ‘full’ model (adjusting for age, age-squared, sex, race/ethnicity, poverty-to-index, education, and born in the USA) and model A would indicate confounding by age and socioeconomic differences. Any differences between the full model and model B would indicate confounding by age.

Similar to the above, we conducted a third sensitivity analysis that adjusted for only sex, age and age squared to examine potential for ‘collider bias’. Briefly, collider variables are those that ‘block’ an association between a variable and the outcome. For example, education is a potential ‘collider’ variable between the association between age and MTL, because younger participants (below 18) are less likely to have higher education (at high school or greater) and vice versa. In the sensitivity analysis, any large difference (>20%) between the association estimated in a model adjusting only for age, age squared and sex and the full model (adjusting for age, age squared, sex, race/ethnicity, poverty-to-index, education and born in the USA) may indicate a potential collider bias among the adjustment variables (such as education blocking an association between age and MTL).

Continuous biomarker factors that had a right-skewed distribution were log-transformed and z-standardized as previously described.^{8,10} For each of the 461 factors, we attained an association size, *P*-value and *R*² using survey-weighted linear regression. MTL was also standardized (the observed values were mean subtracted and divided by the standard deviation).

We used the Benjamini-Hochberg False Discovery Rate (FDR,¹⁷) and deemed an FDR of 5% (*P*-value of 0.002) as significant (Figure 1C). We used the FDR to prioritize findings to achieve the goal of controlling the number of ‘negative associations’ with respect to the number of claims. Put in a different way, if a collection of findings were reported at an FDR of 5%, then the reader would have some assurance that at most 5% of the significant findings were concluded to be false-positives. Results that are significant at a nominal *P*-value threshold do not have this property. Further, we also estimated the variance explained (*R*²) as an alternative way for the reader to prioritize findings.

We assessed the non-parametric correlations among the set of factors that had an FDR < 5%, specifically bi-serial correlations between binary factors and Spearman correlations when considering continuous factors. We visualized these pairwise correlations in a heat map and arranged the factors using a hierarchical clustering algorithm²⁹ as previously described.¹¹

We determined whether the set of significant associations (FDR < 5%) interacted with gender and Mexican American or African American race/ethnicity (Figure 1D). We modelled the interaction with a multiplicative term between the exposure variable and an indicator variable for gender, Mexican American or African American race/ethnicity. We built a multivariate model consisting of all factors that were significant in the scanning step (FDR < 5%) to estimate variance explained (Figure 1E).

Assessing expression of genes that are associated with telomere length

We attempted to estimate the association of two of the set of significant environmental exposure factors, smoking and PCBs, on the expression of genes that play a role in telomere length (Figure 1F; and [Supplementary Methods](#), available as [Supplementary data](#) at *IJE* online). We chose to investigate gene expression changes associated with PCBs because its correlation with MTL had little evidence in the literature. Further, PCBs had public gene expression data available. Second, we chose smoking as a ‘contrast’ to PCBs as it had a number of documented associations with telomere length in the literature and the sample size for the gene expression experiment was large. Because few of the gene expression datasets for both PCB and smoking exposure were observational in nature, we sought to find associations, not causal relationships.

In all, 28 genes have been previously implicated in mean telomere length in genome-wide association studies (GWASs): *TERT*, *DHX35*, *WDR65*, *PELI2*, *KPNA5*, *BRUNOL4*, *PIKC3C*, *OBFC1*, *TERT*, *KRT80*, *AK097143*, *ASCC2*, *DKK2*, *PAPSS1*, *LOC100128993*, *FXYD4*, *RASGEF1A*, *SYT16*, *TMPRSS7*, *TRDMT1*, *ACYP2*, *ZNF676*, *SLC44A4*, *CXCR4*, *CTC1*, *CSNK2A2*, *C5orf42* and *RTEL1*⁴. We sought evidence for differences in gene expression levels of a subset of these genes in cell lines that were exposed to polychlorinated biphenyls and a human population that reported smoking behaviour.

We combined seven publicly available datasets from the Gene Expression Omnibus (GEO) on polychlorinated biphenyl exposure sampled from peripheral blood mononuclear cells and cell lines (Affymetrix Human Arrays; [Table S3](#), available as [Supplementary data](#) at *IJE* online). These datasets were a combination of experimental, where exposures to PCBs were randomly assigned to cell lines (four out of seven), or induced peripheral blood mononuclear cells (three out of seven). One out of the seven was observational. We combined data from all experiments and assessed the significance of differences in mean and standard deviation in gene expression levels for 24 of the 29 GWAS-implicated genes (48 individual Affymetrix probes): *TERT*, *NAF1*, *ACYP2*, *OBFC1*, *ZNF208*, *RTEL1*, *ASCC2*, *CELF4*, *C5orf42*, *CSNK2A2*, *CTC1*, *FAM162B*, *KRT80*, *PELI2*, *LOC100128993*, *LRRC31*, *LRRC34*, *MYNN*, *PAPSS1*, *FXYD4*, *SLC44A4*, *SYT16*, *TRDMT1* and *CFAP57*. To test differences in mean, we used a linear regression model adjusting for experiment number (as a categorical variable). To test non-equality of standard deviation, we used Levene’s test of equality of variances.

We downloaded an observational dataset consisting of lymphocyte gene expression on 2296 smokers and 7120

non-smokers from a public data repository, ArrayExpress. In these experiments, gene expression was assayed on an Illumina Human 6 bead expression array.³⁰ We tested for differences in logarithm of expression mean levels of 19 genes: *ACYP2*, *BRUNOL4*, *CSNK2A2*, *CXCR4*, *DHX35*, *DKK2*, *DNMT2*, *FXYD4*, *KPNA5*, *OBFC1*, *PAPSS1*, *PELI2*, *PIKC3C*, *RASGEF1A*, *TERT*, *TMPRSS7*, *RTEL1*, *NAF1* and *ASCC2* (*ASC1p100*), using a two-tailed unpaired t test.

Results

[Table S2](#) (available as [Supplementary data](#) at *IJE* online) shows the age, sex, race/ethnicity, education and income-to-poverty characteristics of the NHANES participants with telomere measurements; 52 was the average age of the participants. Next, 52%, 17% and 30% of the participants attained less than a high school education, high school equivalent, and greater than high school level education, respectively. The average income to poverty ratio (household income divided by the poverty threshold at the time of the survey) was 2.2; 64% of the participants were born the USA; and 56%, 24%, 15%, 3% and 2% reported to be of Mexican, Non-Hispanic White, Black, Other Hispanic or Other race/ethnicity, in the participant sample.

[Table S4](#) (available as [Supplementary data](#) at *IJE* online) shows a univariate and multivariable linear regression model predicting MTL as a function of demographic adjusting variables. We first report the univariate associations between the adjusting variables and MTL. A 1-year increment in age was negatively correlated with MTL (−0.02 for a 1-year increment in age in 1 SD change in telomere length; 95% CI: [−0.04, −0.01]). Black race/ethnicity was associated with higher levels of MTL relative to Whites [0.25 higher than the White referent; 95% CI: (0.14, 0.37)]. Participants with high school [−0.09 lower than the referent; 95% CI: (−0.2, −0.01)] and less than high school education [−0.21 lower than the referent; 95% CI: 9–0.3, −0.13] had lower MTL levels relative to participants with greater than high school educated participants.

In the multivariate model, females had longer telomeres than males (after adjustment with covariates). Second, after adjusting for covariates, Blacks had on average shorter telomeres than White participants. Third, participants with less than a high school education had a shorter MTL than those with greater than a high school education, and higher family poverty-income ratio was associated with shorter MTL. Together, these demographic adjustment variables explained 14% of MTL variability.

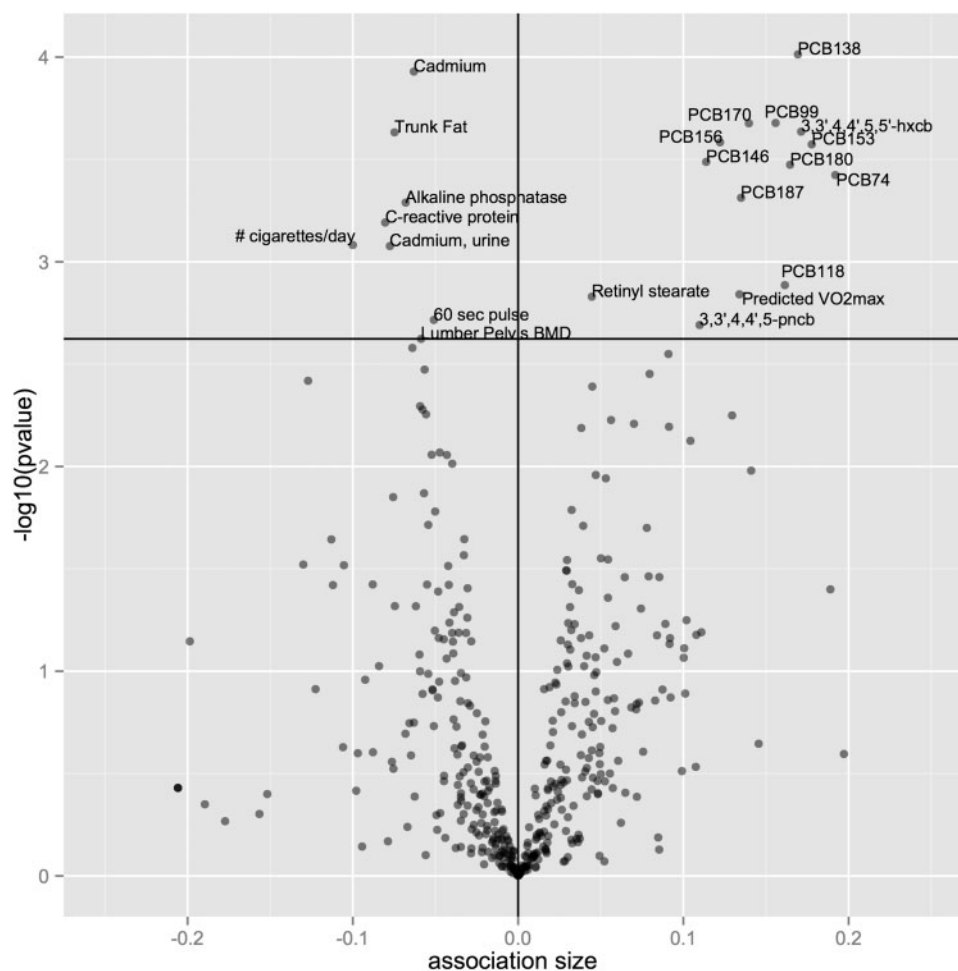


Figure 2. $-\log_{10}(\text{pvalue})$ vs. association size for 461 factors in mean telomere length (MTL). X-axis is the SD change in MTL for a 1 SD unit change in variables. Y-axis is the $-\log_{10}(\text{pvalue})$ of association. Black horizontal line denotes p-value of 0.002 (FDR threshold of 5%).

Correlating environmental, physiological and self-reported behavioural factors with telomere length

Of the 461 factors associated with MTL, 22 achieved significance at the $\text{FDR} < 5\%$ corresponding to a P -value of 0.002 (Figure 2, Table 1; Table S5, available as Supplementary data at *IJE* online).

First, we present results associated with shorter telomere length. We found that higher levels of indicators of trunk fat, inflammation (C-reactive protein), lumbar pelvis bone mineral density (BMD) and a marker of liver and bone disease, alkaline phosphatase, were all associated with shorter MTL. Per 1 SD increase in the independent variable, MTL decreased by (in units of SD): -0.08 [95% CI: $(-0.11, -0.04)$, $P = 2 \times 10^{-4}$] for trunk fat; -0.08 [95% CI: $(-0.12, -0.04)$, $P = 6 \times 10^{-4}$]; -0.07 [95% CI: $(-0.099, -0.037)$, $P = 5 \times 10^{-4}$] for alkaline phosphatase. These biomarkers of clinical physiology added modest amounts to adjusted R^2 ($< 1\%$ incremental increase in R^2). Further, some environmental indicators were also associated with shorter telomere length. Number of cigarettes

smoked per day (among current smokers) and urine and serum biomarkers of cadmium were associated with shorter telomere length. For a 1 SD increase, MTL decreased by (in units of SD): -0.08 [95% CI: $(-0.12, -0.04)$, $P = 8 \times 10^{-4}$] for urinary cadmium, and -0.06 [95% CI: $(-0.09, -0.04)$, $P = 1 \times 10^{-4}$] for serum cadmium. Urinary cadmium had the highest R^2 reported in the investigation (3.8%). Cadmium is a heavy metal commonly found in the matrix of cigarette smoke.³¹

We now consider factors associated with longer telomere length. MTL increased by 0.13 SD for predicted VO_2Max [95% CI: $(0.07, 0.20)$, $P = 1 \times 10^{-3}$] a measurement of predicted maximum rate of oxygen uptake (a function of body mass index, sex, a questionnaire item about physical activity, and age³²). MTL increased by 0.05 SD [95% CI: $(0.02, 0.07)$, $P = 2 \times 10^{-3}$] for a 1 SD increase in log-transformed serum retinyl stearate; 12 out of the 35 polychlorinated biphenyl (PCBs) congeners were associated with longer MTL (Figure 2, Table 1; Table S4). For example, MTL increased by 0.17 [95% CI: $(0.11, 0.23)$,

$P = 1 \times 10^{-4}$] for a 1 SD unit increase in log-transformed PCB138, and by 0.16 [95% CI: (0.09, 0.22), $P = 2 \times 10^{-4}$] for a 1 SD increase in log-transformed PCB99. Other PSB congeners identified at the FDR level of significance included PCB170, previously associated with T2D prevalence,⁸ PCB156, PCB74, PCB180, PCB146, PCB118 and PCB187 as well as 3,3-, 4,4', 5,5'/3,3', 4,4' and 5 types of PCBs.

We hypothesized that the positive relationship between PCBs and MTL was subject to collinearity with respect to age. To assess collinearity between age and the PCB congeners, we estimated the 'variance inflation factors', an index that measures how much the variance of the coefficient of the PCB terms in the regression model increases due to collinearity, or correlation, with other specified variables in the model. A VIF for PCBs can be estimated for each of the adjusting covariates and is technically defined as $1/(1 - R_j^2)$, where R_j^2 is the variance explained in the PCB factor as a function of a covariate, such as age. If, for example, the R_j^2 is 1, then the PCB factor can be completely explained by another covariate and the VIF will be large. On the other hand, if the R_j^2 is zero and the PCB factor cannot be explained by another variable, the VIF will be 1. A VIF above 5 to 10 is generally considered to be high, and one can conclude that the coefficient may be poorly estimated or unstable because of the dependence among the factor and the other co-variables.³³ In summary, we could not exclude the possibility of collinearity among the PCB and age variables; however, most of the variance inflation factors (VIFs) were modest in size (below 5³⁴). The VIFs for 12 PCB factors ranged from 2.0 (PCB099) to 5.7 (3,3',4,4',5 pentachloro-PCB), with 10 of 12 all under a VIF value of 5 (Table S6, available as Supplementary data at *IJE* online). When removing age and age squared from the model, the VIFs for the PCB factors ranged from 1.2 (PCB138) to 2.25 (3,3',4,4',5 pentachloro-PCB). The range of Pearson ρ between the PCB factors and age was 0.3 to 0.6.

We further estimated the association sizes in sensitivity analyses in two additional modelling scenarios, A (adjusting by sex and race/ancestry only) and B (adjusting by age, age squared, sex and race/ancestry) to examine whether there were any large influences between education, income and age in telomere length correlations (as specified in the full model above, adjusting for age, age-squared, sex, race/ancestry, income-poverty ratio, education and born in the USA). Association sizes for the 12 PCB exposure factors and retinyl stearate (a serum indicator of vitamin A) went from positive to negative between the full model and model A (Table S5, available as Supplementary data at *IJE* online) whereas they had the same sign between the full model and model B, indicating a large effect on age in associations between PCBs and telomere length. In other words, PCBs and

retinyl stearate had an overall negative association in telomere length; however, among age strata, these factors had an opposite association in the positive direction. The association size for lumbar pelvis bone mineral density also switched sign of association. In the full model reported above, the association between lumbar pelvis BMD was negative; however, in model A, the association between BMD and telomeres was positive. We concluded that age had a large influence on association sizes for these factors, whereas education and income-to-poverty ratio had a marginal influence on them. Furthermore, associations for the eight remaining factors that achieved FDR < 5% remained stable between the full model, model A and model B.

Next, we conducted a third sensitivity analysis that adjusted for only sex, age and age-squared to examine potential for 'collider bias' for the set of factors that achieved FDR less than 5%. We lacked evidence to suggest that the association sizes differed between estimates with only age, age squared and sex vs the full model (age, age squared, race/ethnicity, education, income-to-poverty ratio, born in the USA). First, most associations between the two models remained consistent in direction and size (Pearson ρ of 0.98, Figure S1, available as Supplementary data at *IJE* online). The associations that changed over 20% included urinary cadmium (-0.2 vs -0.1 in the full model), and the PCBs, including PCB118 (0.2 vs 0.16), PCB138 (0.2 vs 0.17), PCB146 (0.14 vs 0.11), PCB156 (0.15 vs 0.12) and PCB187 (0.17 vs 0.13). All the associations in the full model were smaller than those estimated in models only adjusting for age, age squared and sex.

Most of the correlations between significant factors were modest (Figure S2, available as Supplementary data at *IJE* online); the median of the absolute value of Spearman rho was 0.2 (interquartile range of 0.06 to 0.33). However, we observed a dense correlation pattern between the PCB factors (Figure S2, median ρ of PCBs was 0.81). Indicators of smoking, cadmium exposure, inflammation and trunk fat were all modestly correlated with one another (median Spearman $\rho = 0.11$).

We built a multivariate model to assess independent effects of a subset ($N=8$) of the 22 significant factors (Table 2). The eight factor variables that we modelled in a single multivariate model included PCB138, serum cadmium, trunk fat, alkaline phosphatase, C-reactive protein, retinyl stearate, pulse rate and lumbar pelvis BMD. We chose eight of the 22 factors to ensure a large sample size, as all factors were ascertained in different subsamples (Table 3). Further, because there was a high correlation between the PCB exposure factors, we chose only one of them, PCB138. We further eliminated number of cigarettes per day, as this variable was ascertained just among smokers. We eliminated urine cadmium as it was ascertained in

Table 2. Associations with mean telomere length in NHANES participants that were FDR < 5%

Description	Category	Estimate	95% CI	P	FDR	Adjusted R2 (full-reduced model)
PCB138 (1 SD of log)	PCBS	0.169	[0.105, 0.234]	9.7×10^{-5}	0.015	0.0092
Cadmium (1 SD of log)	Heavy metals	-0.063	[-0.088, -0.039]	1.2×10^{-4}	0.015	0.0039
PCB99 (1 SD of log)	PCBS	0.156	[0.092, 0.220]	2.1×10^{-4}	0.015	0.0122
PCB170 (1 SD of log)	PCBS	0.140	[0.082, 0.197]	2.1×10^{-4}	0.015	0.0074
3,3',4,4',5,5'-HxCB (1 SD of log)	PCBS	0.171	[0.100, 0.242]	2.3×10^{-4}	0.015	0.0114
Trunk fat (1SD)	Body measures	-0.075	[-0.106, -0.044]	2.3×10^{-4}	0.015	0.0053
PCB156 (1 SD of log)	PCBS	0.122	[0.071, 0.174]	2.6×10^{-4}	0.015	0.0073
PCB153 (1 SD of log)	PCBS	0.178	[0.103, 0.253]	2.7×10^{-4}	0.015	0.0081
PCB146 (1 SD of log)	PCBS	0.114	[0.065, 0.163]	3.2×10^{-4}	0.015	0.0037
PCB180 (1 SD of log)	PCBS	0.165	[0.094, 0.236]	3.4×10^{-4}	0.015	0.0068
PCB74 (1 SD of log)	PCBS	0.192	[0.108, 0.276]	3.8×10^{-4}	0.016	0.0164
PCB187 (1 SD of log)	PCBS	0.135	[0.074, 0.196]	4.9×10^{-4}	0.018	0.0057
Alkaline phosphatase (1 SD of log)	Biochemistry	-0.068	[-0.099, -0.037]	5.1×10^{-4}	0.018	0.0040
C-reactive protein (1 SD of log)	Biochemistry	-0.080	[-0.118, -0.043]	6.4×10^{-4}	0.021	0.0042
Average cigarettes/day during past 30 days (1 SD)	Smoking behaviour	-0.100	[-0.148, -0.052]	8.3×10^{-4}	0.024	0.0081
Cadmium, urine (1 SD of log)	Heavy metals	-0.078	[-0.115, -0.040]	8.4×10^{-4}	0.024	0.0379
PCB118 (1 SD of log)	PCBS	0.161	[0.080, 0.243]	1.3×10^{-3}	0.035	0.0108
Predicted VO2max (1 SD)	Physical fitness	0.134	[0.066, 0.202]	1.4×10^{-3}	0.036	0.0078
Retinyl stearate (1 SD of log)	Nutrients	0.045	[0.022, 0.067]	1.5×10^{-3}	0.036	0.0016
60-s pulse (1SD)	Blood pressure	-0.051	[-0.078, -0.024]	1.9×10^{-3}	0.044	0.0018
3,3',4,4',5-PNCB (1 SD of log)	PCBS	0.110	[0.051, 0.168]	2.0×10^{-3}	0.045	0.0083
Lumbar pelvis BMD (1 SD)	Body measures	-0.059	[-0.091, -0.027]	2.4×10^{-3}	0.050	0.0024

Last column is the additional variance explained (R^2) for the variable. Estimates are to be interpreted as 1 SD unit change in telomere length for a 1 SD unit change in environmental, behavioural or clinical variable.

a smaller sample size relative to other serum cadmium. We eliminated VO2Max as it was also ascertained in a smaller sample size relative to pulse rate. In the multivariate model ($N=8$ variables plus 11 adjustment variables), only PCB153 was significant (estimate of 0.13, $p=0.01$), and this model only explained an additional 1% of variance of MTL vs the baseline model containing only the 11 adjustment variables (Table 3). Further, trunk fat was the next most significant (estimate of -0.06 , $P=0.11$). When modelling only trunk fat and PCB138 plus the adjustment variables, trunk fat and PCB138 were significant ($P < 0.05$). Thus, trunk fat and PCB138 were associated with MTL independently of one another.

After correction for multiple hypotheses, we lacked evidence to conclude that sex, race (Mexican, Black) and income to poverty ratio modified the effect of the variables that achieved FDR < 5% in association with MTL (all FDR > 50%).

PCB and smoking exposure and telomere-related gene expression

We could not conclude that there was a significant difference in the mean level of candidate gene expression in PCB-

exposed vs non-exposed samples after considering multiple hypotheses (Figure S3 and Table S7, available as Supplementary data at *IJE* online, total $N=66$, exposed $N=52$), but one probe for the *SLC44A4* gene demonstrated Bonferroni-level of significance [non-parametric Wilcox $P=0.00096$ (Bonferroni corrected for 48 tests)]. The parametric test adjusting for experiment indicated that higher expression of the PCB exposure status was associated with higher expression of the probe; however, the parametric test was borderline significant [linear regression $P=0.1$ (Bonferroni adjusted), $P=0.002$]. Furthermore, and second, the second probe for *SLC44A4* failed to reach significance (uncorrected Wilcox $P=0.82$). Last, we could not discern any difference in mean or variance of candidate gene expression levels in smokers vs non-smokers (Figure S4 and Table S8, available as Supplementary data at *IJE* online).

Discussion

In this investigation, we searched for correlations between clinical, behavioural and environmental factors and MTL in an epidemiological survey representative of the US population during the years 1999–2002, extending an previously documented approach.⁷ We claim that, by

Table 3. Multivariate models predicting telomere length

Variable	Full model			Reduced model		
	Estimate	95% CI	<i>P</i>	Estimate	95% CI	<i>P</i>
PCB138 (1 SD change of log)	0.126	[0.046, 0.207]	0.013	0.158	[0.079, 0.236]	0.001
Serum cadmium (1 SD change of log)	0.013	[-0.069, 0.095]	0.756	.	.	.
Trunk fat (1 SD)	-0.063	[-0.128, 0.002]	0.088	-0.066	[-0.123, -0.009]	0.037
Alkaline phosphatase (1 SD of log)	-0.036	[-0.103, 0.031]	0.322	.	.	.
C-reactive protein (1 SD of log)	0.018	[-0.039, 0.075]	0.552	.	.	.
Retinyl stearate (1 SD of log)	0.031	[-0.018, 0.080]	0.248	.	.	.
Pulse rate (1 SD)	-0.039	[-0.082, 0.004]	0.110	.	.	.
Lumbar pelvis bone mineral density (1 SD)	0.03	[-0.026, 0.087]	0.321	.	.	.
Female sex	0.037	[-0.068, 0.143]	0.506	0.039	[-0.057, 0.134]	0.442
Race/ethnicity						
Black	0.11	[-0.080, 0.300]	0.286	0.179	[-0.014, 0.372]	0.088
Other ethnicity	-0.102	[-0.563, 0.359]	0.676	-0.063	[-0.467, 0.342]	0.765
Other Hispanic	0.259	[-0.022, 0.540]	0.104	0.422	[0.159, 0.685]	0.006
Mexican	0.129	[-0.218, 0.477]	0.484	-0.016	[-0.298, 0.266]	0.915
Income-to-poverty ratio	0.038	[-0.008, 0.083]	0.137	0.044	[0.0002, 0.088]	0.066
Age	-0.034	[-0.055, -0.012]	0.012	-0.031	[-0.047, -0.014]	0.002
Age-squared	0.0001	[-0.0001, 0.0003]	0.485	2.0×10^{-5}	[-0.0001, 0.0002]	0.757
Education (vs > HS)						
HS	-0.092	[-0.231, 0.047]	0.228	-0.067	[-0.188, 0.054]	0.295
<HS	-0.092	[-0.262, 0.079]	0.320	-0.003	[-0.140, 0.134]	0.963
Born outside USA (vs in USA)	0.076	[-0.198, 0.349]	0.600	0.073	[-0.174, 0.319]	0.571
<i>N</i>	1520			1928		
R ² (full minus baseline)	0.01			0.01		

HS, high school.

'Full model' denotes a model fitting up to 8 FDR-significant variables predicting 1 SD unit change in telomere length. 'Reduced model' only fits two of eight variables with a *P*-value < 0.1 in the full model. Estimates are to be interpreted as 1 SD unit change in telomere length for a 1 SD unit change in environmental, behavioural or clinical variable. Increase in variance of telomere length explained when adding all FDR-significant variables is 1%. *N* denotes sample size. R² (full minus baseline) denotes additional variance explained in MTL for eight or two FDR-significant variables. Sample size and variance explained are depicted in bold print.

associating all possible variables in MTL while accounting for multiple comparisons, we increase the possibilities of discovery while avoiding publication bias, in contrast to the typical epidemiological approach of testing a few hypotheses at a time. We emphasize that this approach does not yield a list of environmental, physiological and behavioural factors that have a causative relationship with MTL. However, the approach has prioritized a list of factors to examine in future investigations for causality in MTL.

We found 22 factors correlated with MTL at FDR < 5%, including clinical indicators of obesity (trunk fat), liver and bone health (alkaline phosphatase), frailty (bone mineral density), adverse environmental exposure (serum/urine cadmium) and inflammation (C-reactive protein: CRP). Our method was also able to uncover the positive association between serum PCB levels and longer MTL. Last, we attempted to study the biological relevance of a subset of our findings by querying publicly available toxicogenomics datasets to test whether exposure status is

associated with changes in expression of genes that modulate telomere length, such as *SLC44A4*.

We found that higher levels of PCBs were correlated with longer MTL. There are a few non-exclusive scenarios that may explain the finding. First, the correlations may be confounded. Second, they may be spurious (type I error). Third, variables may 'switch signs' and have opposite association sizes in the context of collinear variables such as age (Table S5). Despite these challenges to inference, our results are consistent with the one other population-based study.³⁵ Specifically, in an investigation of healthy Koreans, Shin and colleagues found a consistent positive association between PCB exposures and telomere length, adjusting for age, body mass index, smoking and alcohol,³⁵ with correlation coefficients ranging from 0.25 to 0.35.

We claim that these results are consistent with the hypothesized role of longer telomeres and PCB in cancer. Based on animal studies, PCBs are categorized as a potential carcinogens.³⁶ We speculate that PCBs play a role in cancer risk via telomere lengthening or modulating the

function of telomerase-related genes, such as those identified in a previous GWAS.⁴ We note that there is contentious role in carcinogenesis for telomeres, and both longer and shorter telomere length are correlated with different cancer phenotypes. Longer telomeres have been causally associated with risk for melanoma, using a technique known as ‘Mendelian randomization’.³⁷ Short telomeres have shown both null associations in a large observational study [$N=47\,102$, hazard ratio (HR)=0.95, 95% CI: (0.80, 1.11),³⁸] and strong effects [$N=137$, HR=2.16, 95% CI: (1.6, 2.9)³⁹] in cancer risk. We hypothesize that PCB exposure has a role in carcinogenesis and cell survival; however, we emphasize that this investigation is far from assessing the causal nature between PCBs and MTL.

Our investigation has uncovered physiological indicators of cardiovascular-related disease risk. Specifically we found that physiological factors such as trunk fat, inflammation (CRP), physical activity and liver function factors are associated with shorter telomeres. CRP is a marker of chronic inflammation due to chronic disease, such as obesity (and also infection). Alternatively, these disease/clinical indicators may be a result or indication of psychological stress.⁴⁰ Further, chronic diseases, such as type 2 diabetes (T2D), have also been hypothesized to be associated with shorter MTL.⁴¹ However, we stress that these results may not be causal, but an indicator of chronic phenomena. For example, Du and colleagues conclude that genetic predisposition to obesity or T2D is not causally related to telomere length,⁴² leaving the possibility that the relationship between trunk fat and other chronic disease indicators is reverse causal or marks another process associated with ageing. Our findings between serum/urine cadmium and shorter telomere length affirm a recent candidate metals exposure study of NHANES participants.²⁸ Serum cadmium levels increase as humans age, and sources include ambient air pollution (through fossil fuel combustion), diet and smoking.³¹

Our study has a number of limitations. First, the correlations we estimated should not be interpreted as causal and are correlative. Many other confounding biases, including residual confounding (with known confounders) and unknown confounders may exist, even though we adjusted for numerous demographic factors (e.g.^{5,20}). Further, the PCB associations may be due to collinearity of a factor with age.

Second, another drawback includes potential for reverse causation. The NHANES is a cross-sectional survey and changes in MTL may come before exposure, also known as ‘reverse causality’. It is difficult to assert how changes in telomere length come before exposure, and a way avoid reverse causation and confounding bias includes Mendelian randomization (MR)⁴³ and also as referenced above. In

Mendelian randomization, genotypes associated with physiological, environmental and even self-reported behavioural factors are ‘stand-in’ instrumental variables. Because genotypes are assigned randomly at birth and are, in principle, static, biases such as reverse causality and confounding are avoided. However, the MR approach requires genotype indicators of exposure (variants associated with PCB exposure levels in GWAS, for example). Whereas GWAS for MTL have been documented, genotypes that are associated with differences in levels of environmental exposure are just emerging.

Third, another drawback includes potential ‘collision’ among our adjustment variables, resulting in collider bias. For example, as a reviewer has pointed out, education is associated with age (younger participants are less likely to have higher than a high school education and vice versa). In these scenarios, any potential association between age and telomere length is ‘blocked’ by education. Whereas we acknowledge potential collider bias in our analyses, the association sizes between significant factors adjusted only for age, agesquared and sex did not differ from association sizes estimated from our main models (age, agesquared, sex, education, income and born in the USA).

Fourth, we acknowledge that power for detection for the physiological, environmental and self-reported behavioural indicator factors varied due to type of factor (e.g. categorical and continuous) and sample sizes. Therefore, we have explained variance (R^2) for re-prioritization of factors.

Despite limitations, the approach described here gives a comprehensive population-based view of clinical, behavioural and environmental exposure variables associated with telomere length, useful for hypothesis generation for future investigation. We emphasize that the correlations between some environmental exposure factors, such as PCBs and cadmium, and telomere length require further epidemiological replication in other cohorts. At the same time, experiments must also proceed to investigate if (and how) exposures and behaviours influence genes that regulate telomere length.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by a National Institute of Environmental Health Sciences [grant numbers 1K99ES023504, 1R21ES025052], and the Pharmaceutical Research and Manufacturers Association of America (PhRMA).

Conflict of interest: None of the authors report a conflict of interest.

References

- Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of ageing. *Cell* 2013;**153**:1194–217.
- O'Sullivan RJ, Karlseder J. Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol* 2010;**11**:171–81.
- Blackburn EH. Telomeres. *Trends Biochem Sci* 1991;**16**:378–81.
- Codd V, Nelson CP, Albrecht E *et al*. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet* 2013;**45**:422–27.
- Shammas MA. Telomeres, lifestyle, cancer, and ageing. *Curr Opin Clin Nutr Metab Care* 2011;**14**:28–34.
- Patel CJ, Ioannidis JP. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J Epidemiol Community Health* 2014;**68**:1096–100.
- Patel CJ, Ioannidis JP. Studying the elusive environment in large scale. *JAMA* 2014;**311**:2173–74.
- Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 2010;**5**:e10746.
- Hall MA, Dudek SM, Goodloe R *et al*. Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank. *Pac Symp Biocomput* 2014: 200–11.
- Patel CJ, Cullen MR, Ioannidis JP, Butte AJ. Systematic evaluation of environmental factors: persistent pollutants and nutrients correlated with serum lipid levels. *Int J Epidemiol* 2012;**41**:828–43.
- Tzoulaki I, Patel CJ, Okamura T *et al*. A nutrient-wide association study on blood pressure. *Circulation* 2012;**126**:2456–64.
- Lind PM, Riserus U, Salihovic S, Bavel B, Lind L. An environmental wide association study (EWAS) approach to the metabolic syndrome. *Environ Int* 2013;**55**:1–8.
- Merritt MA, Tzoulaki I, Tworoger SS *et al*. Investigation of dietary factors and endometrial cancer risk using a nutrient-wide association study approach in the EPIC and Nurses' Health Study (NHS) and NHSII. *Cancer Epidemiol Biomarkers Prev* 2015;**24**:466–71.
- Patel CJ, Rehkopf DH, Leppert JT *et al*. Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey. *Int J Epidemiol* 2013;**42**:1795–810.
- Patel CJ, Cullen MR, Ioannidis JPA, Rehkopf DH. Systematic assessment of the correlation of household income with infectious, biochemical, physiological factors in the United States, 1999–2006. *Am J Epidemiol* 2014;**181**:171–79.
- Patel CJ, Yang T, Hu Z *et al*. Investigation of maternal environmental exposures in association with self-reported preterm birth. *Reprod Toxicol* 2013;**45C**:1–7.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;**57**:289–300.
- Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Data, 1999–2000*. http://www.cdc.gov/nchs/nhanes/nhanes99_00.htm (31 January 2009, date last accessed).
- Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Data, 2001–2002*. <http://www.cdc.gov/nchs/nhanes/nhanes01-02.htm> (31 January 2009, date last accessed).
- Needham BL, Adler N, Gregorich S *et al*. Socioeconomic status, health behaviour, and leukocyte telomere length in the National Health and Nutrition Examination Survey, 1999–2002. *Soc Sci Med* 2013;**85**:1–8.
- Blanton CA, Moshfegh AJ, Baer DJ, Kretsch MJ. The USDA Automated Multiple-Pass Method accurately estimates group total energy and nutrient intake. *J Nutr* 2006;**136**:2594–99.
- U.S. Department of Agriculture, Agricultural Research Service, Beltsville Human Nutrition Research Center *et al*. *What We Eat in America, NHANES 2001–2002*; 2001 1/1/2013. <http://www.ars.usda.gov/News/docs.htm?docid=13793>. (1 October 2015, date last accessed).
- U.S. Department of Agriculture, Agricultural Research Service, Beltsville Human Nutrition Research Center *et al*. *What We Eat in America, NHANES 1999–2000*; 1999 1/1/2013. <http://www.ars.usda.gov/News/docs.htm?docid=13793>. (1 October 2015, date last accessed).
- Ainsworth BE, Haskell WL, Whitt MC *et al*. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;**32**(Suppl 9):S498–504.
- US Department of Health and Human Services. *2008 Physical Activity Guidelines for Americans*. Washington DC: US Department of Health and Human Services, 2008.
- Lumley T. *Survey: analysis of complex survey samples*. Department of Biostatistics, University of Washington, 2014.
- CDC, National Center for Health Statistics (NCHS). *National Health and Nutrition Examination Survey Analytic Guidelines*. 2003. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf (19 February 2010, date last accessed).
- Zota AR, Needham BL, Blackburn EH *et al*. Associations of cadmium and lead exposure with leukocyte telomere length: findings from national health and nutrition examination survey, 1999–2002. *Am J Epidemiol* 2015;**181**:127–36.
- Gordon A. *Classification*. 2nd edn. Boca Raton, FL: Chapman and Hall/CRC, 1999.
- Charlesworth JC, Curran JE, Johnson MP *et al*. Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. *BMC Med Genom* 2010;**3**:29.
- Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry. *ATSDR - Toxicological Profile: Cadmium*. 2012. <http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id = 48&tid = 15> (21 February 2013, date last accessed).
- Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE. Prediction of functional aerobic capacity without exercise testing. *Med Sci Sports Exerc* 1990;**22**:863–70.
- Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis*. Hoboken, NJ: Wiley, 2006.
- Montgomery MP, Kamel F, Saldana TM, Alavanja MC, Sandler DP. Incident diabetes and pesticide exposure among licensed pesticide applicators: Agricultural Health Study, 1993–2003. *Am J Epidemiol* 2008;**167**:1235–46.

35. Shin JY, Choi YY, Jeon HS *et al.* Low-dose persistent organic pollutants increased telomere length in peripheral leukocytes of healthy Koreans. *Mutagenesis* 2010;25:511–16.
36. US Environmental Protection Agency. *Basic Information | Polychlorinated Biphenyls | Wastes*. <http://www.epa.gov/epa-waste/hazard/tsd/pubs/about.htm> (5 February 2009, date last accessed; website out of action for improvement).
37. Iles MM, Bishop DT, Taylor JC *et al.* The effect on melanoma risk of genes previously associated with telomere length. *J Natl Cancer Inst* 2014;106. doi: 10.1093/jnci/dju267.
38. Weischer M, Nordestgaard BG, Cawthon RM, Freiberg JJ, Tybjaerg-Hansen A, Bojesen SE. Short telomere length, cancer survival, and cancer risk in 47102 individuals. *J Natl Cancer Inst* 2013;105:459–68.
39. Willeit P, Willeit J, Kloss-Brandstatter A, Kronenberg F, Kiechl S. Fifteen-year follow-up of association between telomere length and incident cancer and cancer mortality. *JAMA* 2011;306:42–44.
40. Epel ES, Blackburn EH, Lin J *et al.* Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci U S A* 2004;101:17312–15.
41. Willeit P, Raschenberger J, Heydon EE *et al.* Leucocyte telomere length and risk of type 2 diabetes mellitus: new prospective cohort study and literature-based meta-analysis. *PLoS One* 2014;9:e112483.
42. Du M, Prescott J, Cornelis MC *et al.* Genetic predisposition to higher body mass index or type 2 diabetes and leukocyte telomere length in the Nurses' Health Study. *PLoS One* 2013; 8:e52240.
43. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.