Controlling for baseline telomere length biases estimates of the rate of telomere attrition

Melissa Bateson[1*] Dan T. A. Eisenberg[2] and Daniel Nettle[1]

1 Centre for Behaviour and Evolution and Institute of Neuroscience, Newcastle University, UK.

2 Department of Anthropology, University of Washington, USA.

*Corresponding author: Professor Melissa Bateson

Address for correspondence: Institute of Neuroscience, Newcastle University, Henry Wellcome Building, Framlington Place, Newcastle upon Tyne, NE2 4HH.

Phone: +44 (0) 191 2085056

Email for correspondence: melissa.bateson@ncl.ac.uk

Word count for Abstract and main text: 6750

MeSH indexing:

Publication type:

Journal article/ Statistics, Meta-analysis

MeSH Terms:

Molecular epidemiology

Longitudinal studies

Telomere/ Length, Shortening

Error/ Outcome Measurement

Smoking

Data analyses/ Statistical

Data interpretation/ Statistical

Bias/ Statistical, Systematic, Epidemiologic

Computing/ Statistical

Meta-analysis

Humans

Abstract

Background

Longitudinal studies have an important role in telomere epidemiology. In analysing effects of exposures on change in leukocyte telomere length (LTL), it is common to control for baseline LTL. However, collider bias arising from measurement error could cause overestimation of the difference in LTL attrition between groups with different exposures. We evaluated this using smoking as a test case.

Methods

We simulated LTL data to ask whether controlling for baseline LTL biases estimates of the difference in LTL attrition between smokers and non-smokers. We tested predictions from our simulation in a meta-analysis of previously-published longitudinal cohorts.

Results

Our simulations show that if baseline LTL is shorter in smokers and LTL measurement error is non-zero, then controlling for baseline LTL overestimates the difference in LTL attrition between smokers and non-smokers. The size of this bias increases synergistically with increasing baseline difference and increasing LTL measurement error. Supporting these simulation results, the estimated difference in LTL attrition between smokers and non-smokers in empirical data is greater when models control for baseline LTL and the size of this discrepancy is positively correlated with LTL measurement error.

Conclusions

The false-positive error rate for reports of effects of smoking on telomere attrition is likely to exceed 5%. The bias responsible is not specific to smoking and will affect all exposures for which baseline differences in LTL exist. To avoid bias, models of LTL attrition should not control for baseline LTL. Many claims of accelerated LTL attrition in individuals exposed to adversity need to be re-assessed.

Key words: telomere length, telomere attrition, longitudinal, measurement error, regression to the mean, collider bias

Key messages

- Analysis of longitudinal changes in leukocyte telomere length (LTL)—a widely studied biomarker of human health—is emerging as a common method for establishing whether exposure to toxins, disease, stress or other forms of adversity causes accelerated LTL attrition.
- We show that the common strategy of statistically controlling for baseline LTL in analyses of LTL change introduces bias and is likely to yield false-positive results.
- Based on our findings, we recommend that models of LTL change should not control for baseline LTL.
- Many previous claims of accelerated LTL attrition in individuals exposed to toxins, disease, stress or adversity may be false positive results and consequently need to be re-assessed.

Introduction

Leukocyte telomere length (LTL)—the mean number of TTAGGG sequence repeats at the end of leukocyte chromosomes—is emerging as a widely studied biomarker of human health. Many cross-sectional studies of LTL demonstrate that mean LTL is shorter in individuals that have been exposed to diverse forms of adversity.[1] Recent meta-analyses show that LTL tends to be shorter in individuals who are smokers,[2,3] are more sedentary,[4,5] are obese,[6] were subjected to childhood trauma[7] or psycho-social stress,[8] suffer from schizophrenia,[9,10] post-traumatic stress disorder,[11] anxiety or depression[12,13] or have higher perceived stress.[14] These studies have been widely assumed to support the hypothesis that the exposure increases the rate of LTL attrition. However, a cross-sectional association between an exposure and LTL does not necessarily imply a causal link between the exposure and telomere attrition: further evidence for causation is required.[15] A common source of evidence used to support the hypothesis that an exposure causes increased telomere attrition comes from studies demonstrating that the same exposures associated with shorter LTL cross-sectionally are also associated with faster LTL attrition within individuals over time. To obtain such evidence, telomere attrition is estimated from longitudinal datasets in which LTL is measured at least twice in each individual, first at baseline (LTL_b) and again at follow-up (LTL_{fu}; see Box 1 for a list abbreviations). The best estimate of the change in telomere length for a given individual is then simply the difference between the baseline and follow-up measurements (ΔLTL; where negative values indicate telomere attrition). Multiple regression approaches are typically used to estimate the associations between exposure variables and the rate of telomere attrition.[16–22] In the current paper we address the question of how these statistical models should be constructed in order to obtain unbiased estimates. As we explain below, there are strong theoretical reasons to predict that the current practice of controlling statistically for LTL_b biases estimates of the difference in ΔLTL between groups of individuals with different exposures and increases the probability of false-positive results. While our discussion is relevant to all of the exposures listed above (and also other factors implicated in accelerated telomere attrition including age[16,18,19,23] and male sex[24]), here we use the comparison of smokers and non-smokers to illustrate the impact of different analytic strategies.

Box 1: List of abbreviations

LTL	Leukocyte telomere length.
LTL_b	True LTL at the baseline timepoint. Units are bp (base pairs).
LTL_{fu}	True LTL at a follow-up timepoint. Units are bp.
ΔLTL	True change in LTL between baseline and follow-up (calculated as LTL_{fu}-LTL_b); telomere attrition is thus a negative value of ΔLTL. Units are $bp.year^{-1}$.
$mLTL_b$	Measured LTL at the baseline timepoint. Units are bp.
$mLTL_{fu}$	Measured LTL at a follow-up timepoint. Units are bp.
mΔLTL	Measured change in LTL between baseline and follow-up (calculated as $mLTL_{fu}$-$mLTL_b$). Units are $bp.year^{-1}$.
$error_b$	LTL measurement error at baseline.
$error_{fu}$	LTL measurement error at follow-up.
CV	Coefficient of variation (standard deviation/mean) of measurement error. Expressed as %.

Researchers often have a strong intuition that it is important to control for baseline variation in the outcome variable of interest in analyses of change. In the current context, this implies including LTL_b as a covariate (i.e. a continuous predictor variable for which a regression coefficient is estimated) in analyses of the association between smoking and ΔLTL (models 2 and 3 in Table 1). We have found eleven studies that report the association between smoking and ΔLTL and all of these control for

LTL_b in their multiple regression models by including it as a covariate.[16–21,23,25–28] What are the arguments in favour of controlling for LTL_b in this way?

In a highly-cited paper, Vickers & Altman[29] consider the best analytic approach for controlled trials of an intervention with baseline and follow-up measurement. They show that analysis of covariance (which controls for baseline measurement in an analysis of change) yields the largest estimate (of the models they compared) for the effect of the intervention on the measured outcome variable. They argue that analysis of covariance is generally the most powerful analytic approach, and that the efficiency gains from controlling for baseline will be greatest when the correlation between baseline and follow-up measurements is low. This paper is cited as the justification for controlling for LTL_b in at least one study of the factors associated with ΔLTL.[28] In studies of telomere dynamics, the correlation between baseline and follow-up telomere measurements is often low (for example, Bendix et al.[16] report a Pearson correlation of only 0.38), apparently providing a strong argument for controlling for LTL_b in analyses of ΔLTL.

However, although controlling for differences in LTL_b can increase regression coefficients and hence improve statistical power, there is an established epidemiological literature showing that this practice can yield biased estimates and hence spurious, false-positive results. One scenario in which bias occurs is when the outcome variable is measured with error.[30] For example, Glymour et al.[31] examined the consequences of controlling for baseline cognitive function in asking whether educational attainment affects change in cognitive function in old age. They showed that baseline control induces a spurious statistical association between education and change in cognitive function because of measurement error. More generally, they conclude that when exposures are associated with baseline health status, an estimation bias arises if there is measurement error in health status.

In the case of LTL, two meta-analyses have confirmed that smokers have shorter LTL than non-smokers in cross-sectional datasets.[2,3] Thus, longitudinal datasets are likely to show a baseline association between smoking and LTL. It is also well established that measurement error is a major problem in telomere epidemiology. In large-scale studies, LTL is most commonly measured via a quantitative PCR-based method[32] and less frequently via a more expensive Southern blot-based method[33]. Both methods involve error, and while the magnitude of this error varies between studies, evidence suggests that the Southern blot method is typically more precise, with one comparison estimating the inter-assay coefficient of variation (CV) as 6.45% for qPCR and 1.74% for Southern blot[34]. Much higher reported inter-assay CVs for both methods are not uncommon (e.g. 9.3% for qPCR[21] and 2.8% for Southern blot[35]). Therefore, controlling for LTL_b in analysis of the association between smoking and ΔLTL appears to meet the criteria for bias identified by Glymour et al.[31]

In order to formally establish whether an analysis is likely to be biased, epidemiologists advocate construction of a directed acyclic graph. This is a diagram representing the causal relationships among a set of variables that can be used to identify the correct analytic strategy.[30,31,36] We used this approach to represent one possible hypothesis for the relationships among smoking, LTL_b and ΔLTL. Figure 1 represents the null hypothesis that smoking does not affect ΔLTL; we assumed instead, that the association between smoking and LTL_b is brought about by both variables being caused by exposure to a third variable. We assumed that this third variable is exposure to early-life adversity,[15] but it could equally be a genetic difference. To reflect the presence of error in the measurement of LTL we distinguish between true and measured values of LTL and ΔLTL; measured values are indicated with a prefix of m. Although we are ultimately interested in true LTL and ΔLTL, these are latent variables that are not directly accessible. Any analysis must therefore use mLTL and mΔLTL.

We assume that $mLTL_b$ is positively related to true LTL_b and baseline measurement error ($error_b$), and that mΔLTL is positively related to ΔLTL and follow-up measurement error ($error_{fu}$). However, mΔLTL must also be negatively related to $error_b$ (see the Supplementary Material for a proof of why this follows). This is due to regression to the mean: the phenomenon whereby subjects measured with an extreme error, negative or positive, at baseline will on average tend to be measured with a less extreme error at follow-up, generating the negative correlation between $mLTL_b$ and mΔLTL that is commonly observed in longitudinal telomere datasets.[37]

In Figure 1, a path connects smoking with $mLTL_b$ via early-life adversity and LTL_b. Early-life adversity is assumed to cause both smoking and LTL_b (in a directed acyclic graph, a path is a series of lines connecting two variables, regardless of arrow direction). Thus, as long as early-life adversity is not controlled for, a negative association will be present between smoking and $mLTL_b$. A path also connects smoking with mΔLTL via early-life adversity, LTL_b, $mLTL_b$ and $error_b$. On this path, $mLTL_b$ is caused by both LTL_b and $error_b$ and is therefore what is termed a 'collider', a common effect of our outcome and predictor variables (mΔLTL and smoking respectively). In the parlance of directed acyclic graphs, a collider blocks a path, meaning that smoking is independent of mΔLTL under our null hypothesis. However, controlling statistically for $mLTL_b$ unblocks the path between smoking and mΔLTL and hence introduces a spurious association between smoking and mΔLTL. This latter phenomenon is known as 'collider bias'.[38,39] In summary, it follows from the assumptions embodied in Figure 1 that controlling for $mLTL_b$ should inflate estimates of the association between smoking and mΔLTL via collider bias. The size of this bias should depend on both the presence of an association between smoking and LTL_b and the size of the LTL measurement error.

In the remainder of this paper we test above predictions with two complementary approaches. First, we use a simulation model to show numerically that controlling for $mLTL_b$ biases estimates of the association between smoking and mΔLTL and that the size of the bias depends on size of the LTL measurement error. By using realistic values in our simulation we determine the likely importance of any bias. Second, we use meta-analysis of seven previously-published empirical datasets to test the major assumptions and predictions of our simulation model in real LTL data.

Simulation model

The advantage of a simulation approach is that it is possible to generate datasets for which the true values of latent variables (in this case LTL_b and ΔLTL) are known. We can then verify how adding different magnitudes of measurement error and using different statistical analysis strategies affect estimates of the difference in ΔLTL between smokers and non-smokers. We simulated longitudinal LTL datasets in which we set the true differences between smokers and non-smokers in LTL_b, ΔLTL and the LTL measurement error ($error_b$ and $error_{fu}$) based on realistic values obtained from the literature. We then used these simulated datasets to calculate the size of biases in the estimates for the difference in ΔLTL between smokers and non-smokers obtained from different statistical models in which we varied whether we controlled for LTL_b.

We compared the four statistical models given in Table 1. Model 1 is the basic model in which mΔLTL is predicted by smoking status with no statistical control for $mLTL_b$. Model 1 is rarely found in the telomere epidemiology literature, but is sometimes seen in the analysis of randomised controlled trials of interventions such as physical exercise.[e.g. 40] Model 2 includes control for $mLTL_b$ by adding $mLTL_b$ as a covariate. Model 2 represents the approach recommended by Vickers & Altman[29] and most commonly adopted in the current telomere epidemiology literature.[e.g. 17–23,25–27] Model 3 is a less common variant of model 2 in which the outcome variable is $mLTL_{fu}$ as opposed to mΔLTL.[e.g.

[16,28,41,42] Model 4 is a repeated-measures equivalent of model 1 in which the outcome variable is mLTL and timepoint (baseline versus follow-up) is entered as a categorical predictor;[e.g. 43] in this model inclusion of the interaction between timepoint and smoking is necessary to test the hypothesis that mΔLTL differs between smokers and non-smokers. Note that models 1 and 4 contain no control for mLTLb, in that $mLTL_b$ is not included on the right-hand side of the model equation, whereas models 2 and 3 control for $mLTL_b$ by including it as a covariate and estimating its regression coefficient.

Methods

We simulated LTL datasets under four different scenarios for the true differences in LTL_b and ΔLTL between smokers and non-smokers: (A) No difference in LTL_b and no difference in ΔLTL; (B) No difference in LTL_b, but a true difference in ΔLTL; (C) A true difference in LTL_b, but no difference in ΔLTL; and (D) A true difference in LTL_b and a true difference in ΔLTL (Table 2). The parameter values used in each scenario were taken from Aviv et al.,[26] who report a small, but significant, difference in $mLTL_b$ between smokers and non-smokers of 141 bp and a non-significant mΔLTL between smokers and non-smokers of -2 $bp.year^{-1}$. We chose this study because LTL was measured using Southern blot and the reported inter-assay CV is only 1.4%. Thus, the LTL measurements are likely to be reasonable estimates of the true values.

The simulation of LTL values was implemented in the statistical computing language R (script available at the following DOI: 10.5281/zenodo.1009086). In each replicate simulation, values of LTL_b were generated for 2000 participants (1000 non-smokers and 1000 smokers) by drawing independent random samples from normal distributions with means and standard deviations given in Table 2. Each participant was then assigned a value of $\Delta LTL.year^{-1}$ by again drawing an independent random sample from normal distributions for ΔLTL with means and standard deviations given in Table 2. This rate of change was applied for 10 years starting with the true LTL_b to yield a true LTL_{fu} for each participant. We assumed that each participant experienced a constant value of ΔLTL over the follow-up interval. Measurement error was introduced into both LTL_b and LTL_{fu} by assuming that mLTL was an independent random sample from a normal distribution with the mean equal to the true LTL and the standard deviation equal to the true LTL*CV/100 where CV is the coefficient of variation of the measurement error expressed as a percentage. Measured ΔLTL for each participant was calculated as the difference between $mLTL_b$ and $mLTL_{fu}$. We assumed values of CV of 0, 1, 2, 4, 8, and 16%, and generated 1000 replicate data sets for each value of CV in each of the four scenarios (A, B, C and D). Note that while these CV values describe various levels of measurement error within our simulations, these specific CV values cannot be straightforwardly compared to the CVs from laboratory measures reported in empirical papers due to varying zero-points (see[44] for discussion of the comparability of CVs).

We modelled the dataset from each replicate with the four different models summarised in Table 1. Models 1, 2 and 3 are variants of the general linear model and were fitted using the 'lm' function in the R base package, whereas model 4 is a general linear mixed-effects model and was fitted using the 'lmer' function in the R package 'lme4'.[45]

To compare the estimates of the difference in mΔLTL between smokers and non-smokers produced by the different models we extracted the β coefficients for the 'Smoking' variable produced by models 1, 2 and 3 and the 'Time point×Smoking' variable for model 4. To analyse type 1 errors (the probability of incorrectly rejecting the null hypothesis of no difference in ΔLTL between smokers and non-smokers in scenarios where there was no true difference) and statistical power (the probability of correctly rejecting the null hypothesis of no difference in ΔLTL in scenarios where there was a true

difference) we additionally recorded whether the β coefficient was significantly different from zero (at $p < 0.05$ as widely employed) in each analysis. Summarised output from one run of the simulation is available at the following DOI: 10.5281/zenodo.1009086. These data were used to create Figures 2, 3 and S1.

To test the sensitivity of our results to various assumptions, we conducted the following additional simulations. First, to examine sensitivity to the size of the difference in LTL_b between smokers and non-smokers in scenarios C and D, we re-ran the simulation with differences of: 0, 100, 200, 400, 800 and 1600 bp. (Our rationale for including differences up to 1600 bp was that assuming age-related attrition of 40 bp.year^{-1} a 1600-bp difference would be expected between 20 and 60-year olds, meaning that for analyses of the effect of age on ΔLTL.year^{-1} this value would be realistic). Second, to examine sensitivity to the size of the study, we re-ran the simulation with the following numbers of participants (half smokers and half non-smokers): 200, 400, 800, 1600, 3200 and 6400. Third, to examine sensitivity to the true difference in ΔLTL between smokers and non-smokers in scenarios B and D we re-ran the simulation with a true difference of -20 bp.year^{-1} (ΔLTL of -50 bp.year^{-1} in smokers and -30 bp.year^{-1} in non-smokers). Finally, to examine sensitivity to the assumption that LTL measurement error is proportional to LTL, we re-ran the simulation with non-proportional measurement error. We used the following standard deviation values to calculate the measurement error: 0, 70, 140, 280, 560 and 1120 bp.

Results

In scenario A, in which there is no difference in either LTL_b or ΔLTL between smokers and non-smokers, all models correctly estimate the true difference in ΔLTL as zero (Figure 2A). However, in scenario C, in which there is a difference in LTL_b, but no difference in mΔLTL, while models 1 and 4 correctly estimate the difference in ΔLTL as zero, models 2 and 3 overestimate it at non-zero values of measurement error, and this overestimation increases as LTL measurement error increases (Figure 2C). In scenario B, in which there is no difference in LTL_b, but a true difference in ΔLTL, all models correctly estimate the difference in ΔLTL at around -2 bp.year^{-1} (Figure 2C). However, in scenario D, in which there is a difference in LTL_b and a true difference in ΔLTL of -2 bp.year^{-1}, while models 1 and 4 correctly estimate the difference in ΔLTL, models 2 and 3 overestimate it at non-zero values of measurement error, and this overestimation increases as measurement error increases (Figure 2D). The magnitude of the bias produced by models 2 and 3 in scenarios C and D is the same, and is hence independent of the presence of a true difference in ΔLTL.

In scenario A, the probability of type 1 errors based on a sample size of 2000 is around 0.05 for all models (Figure 3A). However, in scenario C, the type 1 error rates for models 2 and 3 are greater than 0.05 and rise as CV increases, reflecting the exaggerated estimates of difference in ΔLTL seen in Figure 2C (Figure 3C).

In scenario B, the power to correctly reject the null hypothesis of no difference in ΔLTL based on a sample size of 2000 is approximately the same for all models and decreases with increasing CV (Figure S1B). The low power reflects the small true effect size of only -2 bp.year^{-1}. In scenario D, the power of models 1 and 4 decreases with increasing CV, but the power of models 2 and 3 increases with increasing CV, reflecting the exaggerated estimates of difference in ΔLTL seen in Figure 2D (Figure S1D).

In terms of both accuracy of parameter estimates (Figure 2) and precision of parameter estimates (Figures 3 and S1), models 1 and 4 were identical to each other and different from models 2 and 3 which were identical to each other. Thus, the models fell into two groups determined by whether or

not they control for $mLTL_b$. Since models 3 and 4 are redundant, henceforth, we only describe results for models 1 (no control for LTL_b) and 2 (control for LTL_b).

Varying the difference in LTL_b in scenarios C and D confirmed that there is a synergistic interaction between difference in LTL_b and CV on the size of the bias arising from model 2 (Figure 4A). At high, but realistic, values of the difference in LTL_b and CV, the bias led to near-certain type 1 errors in scenario C (Figure 4B).

Varying the numbers of participants in the simulation had no impact on the accuracy of the parameter estimates: biases in scenarios C and D were identical to those seen in Figure 2 at all study sizes (Figure S2). There was no impact on the probability of type 1 errors in scenario A, but an increased probability of type 1 errors produced by model 2 in scenario C (Figure S3). Increasing the number of participants increased the power to reject the null hypothesis in scenarios B and D, but this increase was greater with model 2 in scenario D due to the exaggerated parameter estimates (Figure S4).

Increasing the true difference in ΔLTL from -2 to -20 $bp.year^{-1}$ in scenarios B and D had no impact on the size of the biases observed: the difference between the parameter estimates for models 1 and 2 was the same as that seen in Figure 2 (Figure S5). Concomitantly, there was no impact on the probability of type 1 errors (Figure S6). Model 1 correctly estimates the difference in ΔLTL at around -20 $bp.year^{-1}$ in scenarios B and D (Figure S5). The larger true effect size results in a huge increase in power in scenarios B and D compared to that seen in Figure S1 (Figure S7).

Changing the way in which we implemented measurement error from error that was proportional to LTL to non-proportional error had no impact on the size of the biases observed in scenarios C and D (Figure S8), the probability of type 1 errors in scenarios A and C (Figure S9) or power in scenarios B and D (Figure S10).

Discussion

As long as there was no true difference in baseline LTL_b between smokers and non-smokers, then all of the statistical models that we applied accurately estimated the difference in ΔLTL between smokers and non-smokers. However, if there was even a small difference in LTL_b between smokers and non-smokers *and* LTL measurement error was non-zero, then controlling for LTL_b biased estimates of the difference in ΔLTL between smokers and non-smokers. Specifically, the difference in ΔLTL was overestimated and the size of this overestimation increased synergistically with increases in the difference in LTL_b and in LTL measurement error. This bias translated into a type 1 (i.e. false-positive) error rate of above the usually-accepted 5% level when there was no true difference in ΔLTL. This rise in the false-positive error rate was exacerbated in studies with larger numbers of participants. The apparent improvement in power provided by models 2 and 3 in scenario D, seen in Figures S1, S4, S7 and S10, and noted by Vickers & Altmanm,[29] is an artefact of biased parameter estimates.

It is worth pointing out that scenario B is unlikely to be very common, unless LTL_b is measured early in life, before the participants have started smoking. Likewise, scenario A is not typical, given the abundant cross-sectional evidence that smokers have shorter telomeres than non-smokers.[2,3] Thus, the scenarios likely to be empirically widespread are exactly those (C and D) where bias will occur if LTL_b is controlled for.

We parameterised our simulation for a comparison of smokers and non-smokers. However, for variables where the difference in LTL_b is larger than 141 bp, as could be the case for a comparison of

different ages or races, our simulations suggest that false-positive error rates for associations with ΔLTL could approach 100% if LTL_b is controlled for (Figure 4).

In conclusion, given that LTL measurement error is never zero, our simulations suggest that models of types 2 and 3, which control for LTL_b, should be avoided in the analysis of factors associated with ΔLTL. In contrast, models 1 and 4 yield accurate parameter estimates. Models 1 and 4 yield equivalent results with two telomere measurements, but model 4 will be the preferred option if more than two telomere measurements are available.

Meta-analysis of empirical datasets

On the basis of our simulations we predict that in real longitudinal datasets, estimates of the difference in ΔLTL between smokers and non-smokers will depend on both the size of the measurement error and the modelling strategy adopted. Specifically, we predict that estimates of the difference in ΔLTL between smokers and non-smokers will be larger when they are derived from models controlling for $mLTL_b$, and that the size of this effect of modelling strategy will increase as measurement error increases.

Here we test these predictions using real data from seven published longitudinal cohorts. Our specific aims were as follows. First, we set out to confirm that there is substantial variation in LTL measurement error among the seven cohorts. Second, we tested whether the estimated association between smoking and mΔLTL is greater when the association is derived from a model controlling for LTL_b (model 2; see Table 1) compared with a model without control for LTL_b (model 1), and whether any discrepancy is explained by differences in LTL measurement error among cohorts.

Methods

We used data from participants in seven longitudinal cohorts whose LTL had been measured at least twice and for which data on smoking status were also available (Table 3). We restricted our analyses to those participants who were either current or never smokers at the time of the baseline LTL measurement (designated 'smokers' and 'non-smokers' respectively); those who had quit smoking at some point prior to the baseline measurement were excluded.

The first telomere measurement for each participant was designated as $mLTL_b$ and the second, or last where more than two were available (both the Lothian cohorts), as $mLTL_{fu}$. For each participant $\Delta LTL.year^{-1}$ was calculated as $(mLTL_{fu}-mLTL_b)/(age_{fu}-age_b)$ so that negative values indicate telomere attrition.

To characterise the LTL measurement error present in each cohort we did not use the CVs reported for the cohorts, because CV values are often incomparable across studies.[44] Instead, we used signatures of measurement error that can be directly calculated from the telomere measurements themselves, namely the correlation between $mLTL_b$ and $mLTL_{fu}$ and the correlation between $mLTL_b$ and mΔLTL.[46] All else being equal, the correlation between $mLTL_b$ and $mLTL_{fu}$ will be weaker the higher the measurement error, and the correlation between $mLTL_b$ and mΔLTL will be more negative the higher the measurement error.[37,47]

For each cohort, we modelled the difference in $m\Delta LTL.year^{-1}$ between smokers and non-smokers using models 1 and 2 (Table 1). These models yielded estimates of the standardised β coefficient for the association between smoking and $m\Delta LTL.year^{-1}$. To compare the difference in the estimates of this parameter between models 1 and 2 we calculated the difference in association ($\Delta\beta = \beta_{model\ 2} - \beta_{model\ 1}$). A more negative association between smoking and $m\Delta LTL.year^{-1}$ in model 2 compared to

model 1 will therefore be indicated by a more negative value of $\Delta\beta$. To compare the results obtained across the seven cohorts we used meta-regression, fitting linear regression models to the values obtained for each cohort weighting data points by the number of participants in each cohort.

Results

Descriptive statistics

The combined dataset included data from 1,768 adults, comprising 550 current smokers and 1,218 never-smokers at the baseline measurement. The mean age at baseline of the cohorts was 65.9±8.5 years (mean±sd; range: 53.4-80.2) and the mean follow-up interval was 8.5±1.2 years (mean±sd; range: 6.0-9.5).

Five cohorts measured LTL using the qPCR method and two used the Southern blot method. For all cohorts, the slope of the regression of $mLTL_{fu}$ on $mLTL_b$ is less than one (Figure 5A). However, the strength of the relationship differs markedly between cohorts, with Pearson correlation coefficients ranging from -0.01 to 0.97 (Table 3). For all cohorts, the slope of the regression of $m\Delta LTL.year^{-1}$ on $mLTL_b$ is negative (Figure 5B). There is a positive association between the correlation coefficient arising from the association between $mLTL_b$ and $mLTL_{fu}$ and the correlation coefficient arising from the association between $mLTL_b$ and $m\Delta LTL.year^{-1}$ (weighted linear regression: β±se = 0.76±0.18, t = 4.17, p = 0.0088; Figure 5C).

Effects of modelling strategy

We compared estimates (standardised β coefficients) of the difference in $m\Delta LTL.year^{-1}$ between smokers and non-smokers derived from models 1 and 2 (Table 3). Coefficients from models 1 and 2 are strongly positively correlated, but not identical (Figure 6A; weighted linear regression: β±se = 0.89±0.11, t = 8.15, p = 0.0005). There is a tendency for the coefficients from model 2 to be more negative, indicating a bigger estimated difference in $m\Delta LTL.year^{-1}$ compared to model 1 (paired t-test: t(6) = 1.87, p = 0.1106). This difference is greater if the comparison is restricted to the five cohorts measured with qPCR (paired t-test: t(4) = 3.87, p = 0.0180). There is a positive relationship between the correlation coefficient arising from the association between $mLTL_b$ and $mLTL_{fu}$ (a proxy for measurement error in the cohort) and $\Delta\beta$ (a measure of likely bias; weighted linear regression β±se = 0.11±0.04, t = 2.91, p = 0.0336; Figure 6B).

Discussion

Two proxies for LTL measurement error varied among the seven cohorts: there was variation in both the correlation between $mLTL_b$ and $mLTL_{fu}$ and the correlation between $mLTL_b$ and $m\Delta LTL$. Furthermore, these two proxies were correlated with each other as would be expected if they both reflect measurement error. When we estimated the difference in ΔLTL between smokers and non-smokers using two modelling strategies, model 1 (no baseline control) and model 2 (baseline control) produced different results: estimates derived from model 2 showed a more negative effect of smoking than those derived from model 1. Since there can only be one true difference in ΔLTL, the estimates derived from either model 1 or model 2 (or both) must be incorrect. The fact that controlling for LTL_b increases estimates of the effect of smoking rather than decreasing them suggests that LTL_b is not a proxy for positive confounders of the difference in ΔLTL between smokers and non-smokers, but instead introduces a bias. Indeed, the directed acyclic graph and simulation analyses both argue that controlling for LTL_b (model 2) yields biased estimates. Thus, it seems likely that model 2 is biased. This conclusion is strengthened by our finding that the size of the discrepancy between the estimates derived from models 1 and 2 is predicted by a proxy for the magnitude of the LTL measurement error present in the cohort.

We do not report the statistical significance of the associations in Table 3. Our rationale was that the cohorts are small (47-539 participants) and the majority of the differences were therefore not significant. However, for the cohorts with indications of high measurement error, the likely bias arising from model 2 is sufficient to cause concerns over inference, especially if the studies were larger. For example, in the Hertfordshire Ageing Study, which has a baseline difference of -0.19 standard deviations and massive measurement error, the β coefficients for the difference in attrition from model 2 (likely biased) is more than double what it is for model 1 (unbiased).

In the Supplementary Material (Table S1 and Figure S11), we show, using the same datasets, that the above results for smoking generalise to two other variables, sex and body mass index, that are also associated with LTL in cross-sectional studies and have been suggested to cause differences in LTL attrition. [6,24]

General Discussion

We have used three separate lines of evidence to argue that controlling for LTL_b in analyses of ΔLTL biases estimates of the effects of exposures such as smoking. First, we used directed acyclic graphs to show that under a realistic set of assumptions, LTL_b is likely to be a collider on the path linking smoking and ΔLTL. Controlling for LTL_b is therefore predicted to introduce collider bias in the form of an overestimation of the true difference in ΔLTL between smokers and non-smokers. Second, we used a simple simulation model to confirm, again under a realistic set of assumptions, that controlling for LTL_b does indeed inflate estimates of the true difference in ΔLTL between smokers and non-smokers, but only when a true difference in LTL is present at baseline. The magnitude of this bias is positively related to the magnitude of TL measurement error. Third, we analysed data from seven longitudinal human cohorts and showed that, in line with our predictions, estimates of the difference in telomere attrition between smokers and non-smokers tended to be greater when LTL_b was included in statistical models as a covariate. Furthermore, the magnitude of this latter difference was predicted by LTL measurement error, as would be expected if the difference arises from collider bias.

We initially found it difficult to obtain an intuitive understanding of why controlling for LTL_b is problematic. Figure 7 is an attempt to provide a graphical explanation based on simulated data. The dark grey triangles and pale grey circles indicate LTL measurements for smokers and non-smokers respectively; the black triangles and circles are the means of the data for smokers and non-smokers respectively. All four panels depict LTL measurements from a scenario in which there is a true difference in LTL_b between smokers and non-smokers, but no true difference in ΔLTL (i.e. scenario C in our simulations). The left-hand two panels (A and C) show LTL measurements made without error (CV = 0%), whereas the right-hand two panels (B and D) show the same true LTL values depicted on the left, but now measured with error (CV = 6%). All four panels plot LTL_b on the x-axis, hence in all panels the mean LTL_b for smokers (black triangle) is to the left of the mean LTL_b for non-smokers (black circle). Panels A and B plot ΔLTL as the outcome variable and thus relate to a model 2-type analysis, whereas panels C and D plot LTL_{fu} as the outcome variable and thus relate to a model 3-type analysis.

Panels A and B of Figure 7 show the association between LTL_b and ΔLTL as a solid black regression line. When there is no measurement error (panel A), there is no relationship between LTL_b and ΔLTL (the slope is zero). However, when LTL measurement error is introduced (panel B), a negative relationship between LTL_b and ΔLTL occurs as a result of regression to the mean. Controlling for LTL_b

in an analysis of the association between smoking and ΔLTL means asking what the difference in ΔLTL between smokers and non-smokers is for a given value of LTL_b; this is conceptually equivalent to comparing the residuals from the regression of ΔLTL on LTL_b for smokers and non-smokers (the black line). In panel A, the residuals of the data from the regression line are identical for smokers and non-smokers, because the means for smokers and non-smokers lie on the line. However, in panel B the mean for smokers lies below the line, whereas the mean for non-smokers lies above the line. Hence in panel B residuals are on average negative for smokers and positive for non-smokers creating a spurious difference in the residual ΔLTL between smokers and non-smokers. This bias only occurs because the smokers have a mean LTL_b that is lower than that of non-smokers; it would not occur if there was no difference in LTL_b, because the black triangle and circle would then be in the same place. Panels C and D show the association between LTL_b and LTL_{fu} as a solid black regression line. When there is no measurement error (panel C), the slope of the relationship between LTL_b and LTL_{fu} is one. However, when LTL measurement error is introduced (panel D), a flatter relationship between LTL_b and LTL_{fu} results. Controlling for baseline LTL_b in an analysis of the association between smoking and LTL_{fu} causes a spurious difference in LTL_{fu} between smokers and non-smokers in panel D via an exactly analogous mechanism to that described for panel B.

Given first, that there are robust differences in LTL_b between smokers and non-smokers,[2,3] second, that LTL measurement error is often substantial[34 and Figure 5] and third, that most published analyses of the effect of smoking on ΔLTL or LTL_{fu} control for LTL_b, we suggest that the difference in ΔLTL between smokers and non-smokers is likely to have been overestimated in the literature. Reports of significantly accelerated LTL attrition in smokers compared to non-smokers should therefore be interpreted with caution.[e.g. 16,18,28] In a recent meta-analysis in which we re-analysed LTL data from 18 longitudinal cohorts without control for LTL_b, we found no evidence to support accelerated LTL attrition in adult smokers.[3] It is therefore likely that there is in fact no true difference in ΔLTL between smokers and non-smokers and that an alternative explanation needs to be sought for the robust difference reported in LTL_b.[15]

Our findings are likely to have much broader implications than the specific case of the effect of smoking on ΔLTL, analysed here. The bias we describe is relevant to estimating the effect of any factor that is associated with a true difference in TL at the time of baseline measurement on the rate of subsequent TL attrition. Indeed, our own analyses suggest that published analyses of the effects of sex and body mass index on ΔLTL are likely to be biased (see Supplementary Material). There is a growing literature based predominantly on cross-sectional data claiming that exposure to various forms of stress and adversity accelerates TL attrition.[1,48–57] While cross-sectional associations between exposure to stress and short TL do not prove that stress causes TL attrition,[15] longitudinal studies have started to emerge that appear to support a causal relationship.[20,22,42] Unfortunately, just as in the literature on effects of smoking, it is typical for analyses to control for TL_b in these latter studies, meaning that the results should be treated with caution. Re-analyses of these datasets is required to establish whether the claimed differences in TL attrition are in due to bias. We predict that removing TL_b as a control variable from the models used to analyse these data will not just increase the standard error of the estimates (as would be true if TL_b was an innocuous incidental variable that needs to be controlled for to increase power), but will systematically shift the parameter estimates for the effect of the exposure on TL attrition towards zero. Our findings are also relevant to areas of epidemiology outside of telomere biology and apply to the analysis of any similarly structured observational studies in which changes over time in imperfectly measured variables are examined. While this problem is understood by some epidemiologists,[e.g. 31] we hope that the current paper raises awareness of measurement error-induced collider bias more widely.

Conclusions

Controlling statistically for baseline telomere length incorrectly inflates estimates of the difference in telomere attrition between smokers and non-smokers, and the size of this bias is positively related to the size of telomere measurement error. This bias is not restricted to smoking and will occur for any factor that, like smoking, is associated with shorter telomeres at the time of the baseline measurement. On the basis of our analyses we recommend that models of telomere attrition should not control for baseline telomere length by including it as a covariate. Given that the majority of previous analyses of factors affecting telomere attrition control for baseline telomere length in this way, many claims of accelerated telomere attrition in individuals that are older, fatter or exposed to various forms of adversity could be false-positive results that need to be re-assessed.

Supplementary Data

Supplementary Material accompanies this article.

Acknowledgements

We thank the following individuals for granting us permission to reuse data from their longitudinal cohorts: Athanase Benetos (ADELAHYDE and Evolution de la Rigidité Artérielle), Yoav Ben-Shlomo (Caerphilly Cohort Study), Cyrus Cooper (Hertfordshire Ageing Study), Diana Kuh (MRC National Survey of Health and Development), Ian Deary (Lothian Birth Cohorts 1921 and 1936).

Funding

This work was supported by the National Centre for the Replacement Refinement & Reduction of Animals in Research (NC/K000802/1 to MB), the European Research Council (AdG 666669, COMSTAR to DN) and the United States National Science Foundation (BCS-1519110 to DTAE).

Conflicts of interest: none declared.

References

1. Pepper G V., Bateson M, Nettle D. Telomeres as integrative markers of exposure to stress and adversity: A systematic review and meta-analysis. *R Soc Open Sci*. 2018;**5**:180744.

2. Astuti Y, Wardhana A, Watkins J, Wulaningsih W. Cigarette smoking and telomere length: A systematic review of 84 studies and meta-analysis. *Environ Res*. Elsevier Inc.; 2017;**158**:480–489.

3. Bateson M, Aviv A, Bendix L, et al. Smoking does not accelerate leukocyte telomere attrition: a meta-analysis of 18 longitudinal cohorts. *Zenodo*. **pre**-**print**(DOI: 10.5281/zenodo.1463971).

4. Denham J, O'Brien BJ, Charchar FJ. Telomere length maintenance and cardio-metabolic disease prevention through exercise training. *Sport Med*. Springer International Publishing; 2016;**46**(9):1213–1237.

5. Mundstock E, Zatti H, Louzada FM, et al. Effects of physical activity in telomere length: Systematic review and meta-analysis. *Ageing Res Rev*. Elsevier B.V.; 2015;**22**:72–80.

6. Mundstock E, Sarria EE, Zatti H, et al. Effect of obesity on telomere length: systematic review

and meta-analysis. *Obesity*. 2015;**23**(11):2165–2174.

7. Li Z, He Y, Wang D, Tang J, Chen X. Association between childhood trauma and accelerated telomere erosion in adulthood: A meta-analytic study. *J Psychiatr Res*. Elsevier Ltd; 2017;**93**:64–71.

8. Hanssen LM, Schutte NS, Malouff JM, Epel ES. The relationship between childhood psychosocial stressor level and telomere length: a meta-analysis. *Heal Psychol Res*. 2017;**5**(1).

9. Rao S, Kota LN, Li Z, et al. Accelerated leukocyte telomere erosion in schizophrenia: Evidence from the present study and a meta-analysis. *J Psychiatr Res*. 2016;**79**:50–56.

10. Polho GB, De-Paula VJ, Cardillo G, Santos B dos, Kerr DS. Leukocyte telomere length in patients with schizophrenia: A meta-analysis. *Schizophr Res*. Elsevier B.V.; 2015;**165**(2–3):195–200.

11. Li X, Wang J, Zhou J, Huang P, Li J. The association between post-traumatic stress disorder and shorter telomere length: A systematic review and meta-analysis. *J Affect Disord*. Elsevier B.V.; 2017;**218**(March):322–326.

12. Schutte NS, Malouff JM. The association between depression and leukocyte telomere length: a meta-analysis. *Depress Anxiety*. 2015;**32**:229–238.

13. Ridout KK, Ridout SJ, Price LH, Sen S, Tyrka AR. Depression and telomere length: A meta-analysis. *J Affect Disord*. 2016;**191**:237–247.

14. Mathur MB, Epel E, Kind S, et al. Perceived stress and telomere length: A systematic review, meta-analysis, and methodologic considerations for advancing the field. *Brain Behav Immun*. Elsevier Inc.; 2015;**54**:158–169.

15. Bateson M, Nettle D. Why are there associations between telomere length and behaviour? *Philos Trans R Soc Lond B Biol Sci*. 2018;**373**:20160438.

16. Bendix L, Thinggaard M, Fenger M, et al. Longitudinal changes in leukocyte telomere length and mortality in humans. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2014;**69 A**(2):231–239.

17. Ehrlenbach S, Willeit P, Kiechl S, et al. Influences on the reduction of relative telomere length over 10 years in the population-based bruneck study: Introduction of a well-controlled high-throughput assay. *Int J Epidemiol*. 2009;**38**(6):1725–1734.

18. Huzen J, Wong LSM, Veldhuisen DJ Van, et al. Telomere length loss due to smoking and metabolic traits. *J Intern Med*. 2014;**275**:155–163.

19. Müezzinler A, Mons U, Karina A, et al. Smoking habits and leukocyte telomere length dynamics among older adults: results from the ESTHER cohort. *Exp Gerontol*. Elsevier Inc.; 2015;**70**:18–25.

20. Révész D, Milaneschi Y, Terpstra EM, Penninx BWJH. Baseline biopsychosocial determinants of telomere length and 6-year attrition rate. *Psychoneuroendocrinology*. Elsevier Ltd; 2016;**67**:153–162.

21. Weischer M, Bojesen SE, Nordestgaard BG. Telomere shortening unrelated to smoking, body weight, physical activity, and alcohol intake: 4,576 general population individuals with repeat measurements 10 years apart. *PLoS Genet*. 2014;**10**(3).

22. Puterman E, Lin J, Krauss J, Blackburn EH, Epel ES. Determinants of telomere attrition over 1 year in healthy older women: stress and health behaviors matter. *Mol Psychiatry*. Nature Publishing Group; 2015;**20**(4):529–535.

23. Farzaneh-Far R, Lin J, Epel E, Lapham K, Blackburn E, Whooley MA. Telomere length trajectory and its determinants in persons with coronary artery disease: Longitudinal findings from the heart and soul study. *PLoS One*. 2010;**5**(1):e8612.

24. Gardner M, Bann D, Wiley L, et al. Gender and telomere length: systematic review and meta-analysis. *Exp Gerontol*. 2014 Mar;**51**:15–27.

25. Chen W, Gardner JP, Kimura M, et al. Leukocyte telomere length is associated with HDL cholesterol levels: The Bogalusa heart study. *Atherosclerosis*. 2009;**205**(2):620–625.

26. Aviv A, Chen W, Gardner JP, et al. Leukocyte telomere dynamics: Longitudinal findings among young adults in the Bogalusa Heart Study. *Am J Epidemiol*. 2009;**169**(3):323–329.

27. Toupance S, Labat C, Temmar M, et al. Short telomeres, but not telomere attrition rates, are associated with carotid atherosclerosis. *Hypertension*. 2017;**70**:420–426.

28. Ockenburg SL Van, Bos EH, Jonge P De, Harst P Van Der, Gans ROB, Rosmalen JGM. Stressful life events and leukocyte telomere attrition in adulthood: a prospective population-based cohort study. *Psychol Med*. 2015;**45**:2975–2984.

29. Vickers AJ, Altman DG. Statistics Notes: Analysing controlled trials with baseline and follow up measurements. *Bmj*. 2001;**323**(7321):1123–1124.

30. Glymour MM, Greenland S. Causal Diagrams. In: Rothman KT, Sander G, Lash TL, editors. *Mod Epidemiol*. 3rd. Philadelphia, USA.; 2008. p. 183–209.

31. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am J Epidemiol*. 2005;**162**(3):267–278.

32. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res*. 2002;**30**(10):47e–47.

33. Kimura M, Stone RC, Hunt SC, et al. Measurement of telomere length by the southern blot analysis of terminal restriction fragment lengths. *Nat Protoc*. Nature Publishing Group; 2010;**5**(9):1596–1607.

34. Aviv A, Hunt SC, Lin J, Cao X, Kimura M, Blackburn E. Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. *Nucleic Acids Res*. 2011;**39**(20):e134.

35. Steenstrup T, Hjelmborg JVB, Mortensen LH, Kimura M, Christensen K, Aviv A. Leukocyte telomere dynamics in the elderly. *Eur J Epidemiol*. 2013;**28**(2):181–187.

36. Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. *Epidemiology*. 1999;**10**(1):37–48.

37. Verhulst S, Aviv A, Benetos A, Berenson GS, Kark JD. Do leukocyte telomere length dynamics depend on baseline telomere length? An analysis that corrects for 'regression to the mean'. *Eur J Epidemiol*. 2013;**28**:859–66.

38. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;**14**(3):300–306.

39. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010;**39**(2):417–420.

40. Werner CM, Hecksteden A, Morsch A, et al. Differential effects of endurance, interval, and

resistance training on telomerase activity and telomere length in a randomized, controlled study. *Eur Heart J*. 2018;1–13.

41. Carlson LE, Beattie TL, Giese-Davis J, et al. Mindfulness-based cancer recovery and supportive-expressive therapy maintain telomere length relative to controls in distressed breast cancer survivors. *Cancer*. 2015;**121**(3):476–484.

42. Shalev I, Moffitt TE, Sugden K, et al. Exposure to violence during childhood is associated with telomere erosion from 5 to 10 years of age: a longitudinal study. *Mol Psychiatry*. Nature Publishing Group; 2013;**18**:576–81.

43. Shin YA, Lee JH, Song W, Jun TW. Exercise training improves the antioxidant enzyme activity with no changes of telomere length. *Mech Ageing Dev*. 2008;**129**(5):254–260.

44. Eisenberg DTA. Telomere Length Measurement Validity: The Coefficient of Variation (CV) is invalid and cannot be used to compare qPCR and Southern Blot telomere length measurement techniques. *Int J Epidemiol*. 2016;1–4.

45. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using {lmer}. *J Stat Softw*. 2015;**67**(1):1–48.

46. Nettle D, Seeker LA, Nussey D, Froy H, Bateson M. Consequences of measurement error in qPCR telomere data: A simulation study. *BioRXiv*. 2018;doi.org/10.1101/491944.

47. Steenstrup T, Hjelmborg JVB, Kark JD, Christensen K, Aviv A. The telomere lengthening conundrum - Artifact or biology? *Nucleic Acids Res*. 2013;**41**(13):1–7.

48. 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**(49):17312–5.

49. Verhoeven JE, Révész D, Wolkowitz OM, Penninx BWJH. Cellular aging in depression: Permanent imprint or reversible process? *BioEssays*. 2014;**36**:968–978.

50. Damjanovic AK, Yang Y, Glaser R, et al. Accelerated Telomere Erosion Is Associated with a Declining Immune Function of Caregivers of Alzheimer's Disease Patients. *J Immunol*. 2007;**179**(6):4249–4254.

51. Humphreys J, Epel ES, Cooper BA, Lin J, Blackburn EH, Lee KA. Telomere shortening in formerly abused and never abused women. *Biol Res Nurs*. 2012 Apr;**14**(2):115–123.

52. Ala-Mursula L, Buxton JL, Ek E, et al. Long-term unemployment is associated with short telomeres in 31-year-old men: An observational study in the Northern Finland Birth Cohort 1966. *PLoS One*. 2013;**8**(11).

53. Litzelman K, Witt WP, Gangnon RE, et al. Association between informal caregiving and cellular aging in the survey of the health of wisconsin: The role of caregiving characteristics, stress, and strain. *Am J Epidemiol*. 2014;**179**(11):1340–1352.

54. Ahola K, Sirén I, Kivimäki M, et al. Work-related exhaustion and telomere length: A population-based study. *PLoS One*. 2012;**7**(7):13–15.

55. Puterman E, Gemmill A, Karasek D, et al. Lifespan adversity and later adulthood telomere length in the nationally representative US Health and Retirement Study. *Proc Natl Acad Sci*. 2016;**113**(42):E6335–E6342.

56. Verhoeven JE, Revesz D, Oppen P van, Epel ES, Wolkowitz OM, Penninx BWJH. Anxiety disorders and accelerated cellular ageing. *Br J Psychiatry*. 2015;**206**:371–378.

57. Park M, Verhoeven JE, Cuijpers P, Reynolds CF, Penninx BWJH. Where you live may make you old: The association between perceived poor neighborhood quality and leukocyte telomere length. *PLoS One*. 2015;**10**(6):1–13.

58. Hjelmborg B, Nzietchueng R, Kimura M, et al. Leukocyte telomere length is inversely correlated with plasma Von Willebrand factor. *Thromb Res*. 2010;**125**(6):e339–e342.

59. Gardner MP, Martin-Ruiz C, Cooper R, et al. Telomere length and physical performance at older ages: an individual participant meta-analysis. *PLoS One*. 2013;**8**(7).

60. Harris SE, Marioni RE, Martin-Ruiz C, et al. Longitudinal telomere length shortening and cognitive and physical decline in later life: The Lothian Birth Cohorts 1936 and 1921. *Mech Ageing Dev*. Elsevier Ireland Ltd; 2016;**154**:43–48.

Table 1: The four statistical models compared.

No.	Model		
	Outcome variable	**Fixed predictor variable(s)[1]**	**Equivalent statistical test**
1	mΔLTL	Smoking	Two-sample t-test (or multiple regression[2])
2	mΔLTL	$mLTL_b$ + Smoking	Analysis of covariance or multiple regression[2]
3	$mLTL_{fu}$	$mLTL_b$ + Smoking	Analysis of covariance or multiple regression[2]
4	mLTL	Timepoint + Smoking + Timepoint×Smoking[3]	Repeated-measures analysis of variance or mixed-effects model

Notes: [1]Smoking and timepoint are categorical variables with two levels each (smoker/non-smoker and baseline/follow-up respectively) and $mLTL_b$ is a continuous variable.
[2]Multiple regression is appropriate if additional control variables are included (e.g. age, sex, race etc.)
[3]Model 4 additionally contains a random effect (intercept) of participant to account for repeated measures on individuals.

Table 2. Parameter values used in the simulations.

		Scenario			
		A	**B**	**C**	**D**
		No diff. in LTL_b		**True diff. in LTL_b**	
	Parameter	**No diff. in ΔLTL**	**True diff. in ΔLTL**	**No diff. in ΔLTL**	**True diff. in ΔLTL**
Non-smokers	**LTL_b (bp; mean±sd*)**	7430±777	7430±777	7500±777	7500±777
	ΔLTL ($bp.year^{-1}$; mean±sd*)	-40.7±46	-40±46	-40.7±46	-40±46
Smokers	**LTL_b (bp; mean±sd*)**	7430±777	7430±777	7359±777	7359±777
	ΔLTL ($bp.year^{-1}$; mean±sd*)	-40.7±46	-42±46	-40.7±46	-42±46

*Note that these standard deviations of LTL_b and annual attrition are likely to be overestimates of the true values, since both true variation and measurement error contribute to the measured values. However, in the absence of error-free measurements we used these published standard deviations as the best estimates available.

Table 3. Summary of the datasets analysed.

Cohort (acronym)	Country	Mean age at baseline (years)	Mean follow-up interval (years)	LTL measurement method	Number of participants by baseline smoking status[a]		Diff. in LTL_b between smokers and never-smokers (Cohen's d)[b]	Signatures of LTL measurement error (data from smokers and never-smokers pooled)		Diff. in $\Delta LTL.year^{-1}$ between smokers and never-smokers (standardised β [s.e.])[d]		Reference for cohort
					Current smokers	Never-smokers		Correlation between LTL_b and LTL_{fu} (r)	Correlation between LTL_b and ΔLTL (r)[c]	Model 1[e]	Model 2	
ADELAHYDE (ADE)	France	68.1	8.3	Southern blot	5	42	-0.99	0.93	-0.09	0.49 [0.47]	0.49 [0.50]	[58]
Caerphilly Cohort Study (CCS)	Wales, UK	64.2	8.0	qPCR	207	169	-0.12	0.03	-0.81	0.22 [0.10]	0.12 [0.06]	[59]
Evolution de la Rigidité Artérielle (ERA)	France	58.6	9.5	Southern blot	27	86	0.19	0.96	-0.32	-0.30 [0.22]	-0.24 [0.21]	[27]
Hertfordshire Ageing Study (HAS)	England, UK	67.0	9.2	qPCR	29	93	-0.19	-0.10	-0.75	-0.12 [0.21]	-0.27 [0.14]	[59]
Lothian Birth Cohort 1921 (LBC1921)	Scotland, UK	80.2	9.2	qPCR	3	78	-0.40	0.35	-0.23	0.10 [0.59]	0.06 [0.59]	[59]
Lothian Birth Cohort 1936 (LBC1936)	Scotland, UK	69.6	6.0	qPCR	75	415	-0.16	0.54	-0.31	-0.10 [0.13]	-0.15 [0.12]	[60]
MRC National Survey of Health and Development (NSHD)	England, UK	53.4	9.3	qPCR	204	335	-0.06	0.08	-0.80	0.03 [0.09]	-0.02 [0.05]	[59]

[a]These numbers are smaller than the numbers given in the original reference for the cohort because we only included participants for whom there was telomere length and age at both baseline and follow-up and smoking status at baseline; furthermore, participants who had quit smoking prior to baseline were excluded. [b]Negative numbers indicate that LTL_b is shorter in smokers. [c]Negative numbers indicate that longer LTL_b is associated with greater telomere loss. [d]Negative numbers indicate greater telomere loss in smokers. [e]Models 1 and 2 correspond to models 1 and 2 in Table 1.

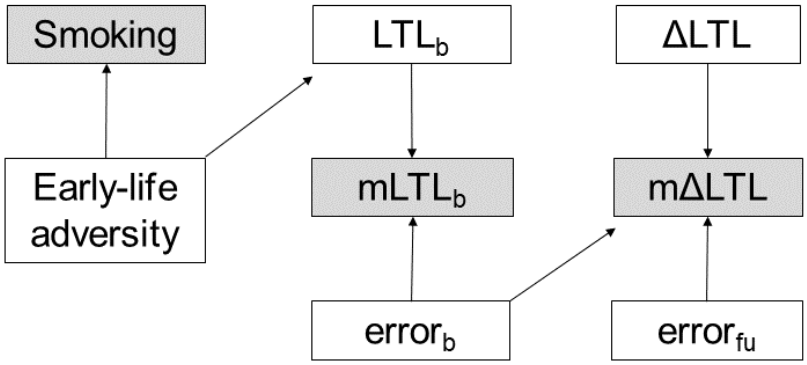

Figure 1. Directed acyclic graph summarising the assumed causal relations between smoking, mLTL$_b$ and mΔLTL. The graph additionally includes the following unmeasured/latent variables: exposure to early-life adversity, true LTL$_b$, baseline measurement error (error$_b$), true telomere change (ΔLTL) and follow-up measurement error (error$_{fu}$). Error$_b$ and error$_{fu}$ are uncorrelated and independent of LTL and ΔLTL. Causal relationships are indicated by arrows. This diagram is analogous to that presented in Glymour et al (29; Figure 3) and Glymour and Greenland (30; Figure 12-14) and can thus be subjected to an identical analysis. See text for further details.

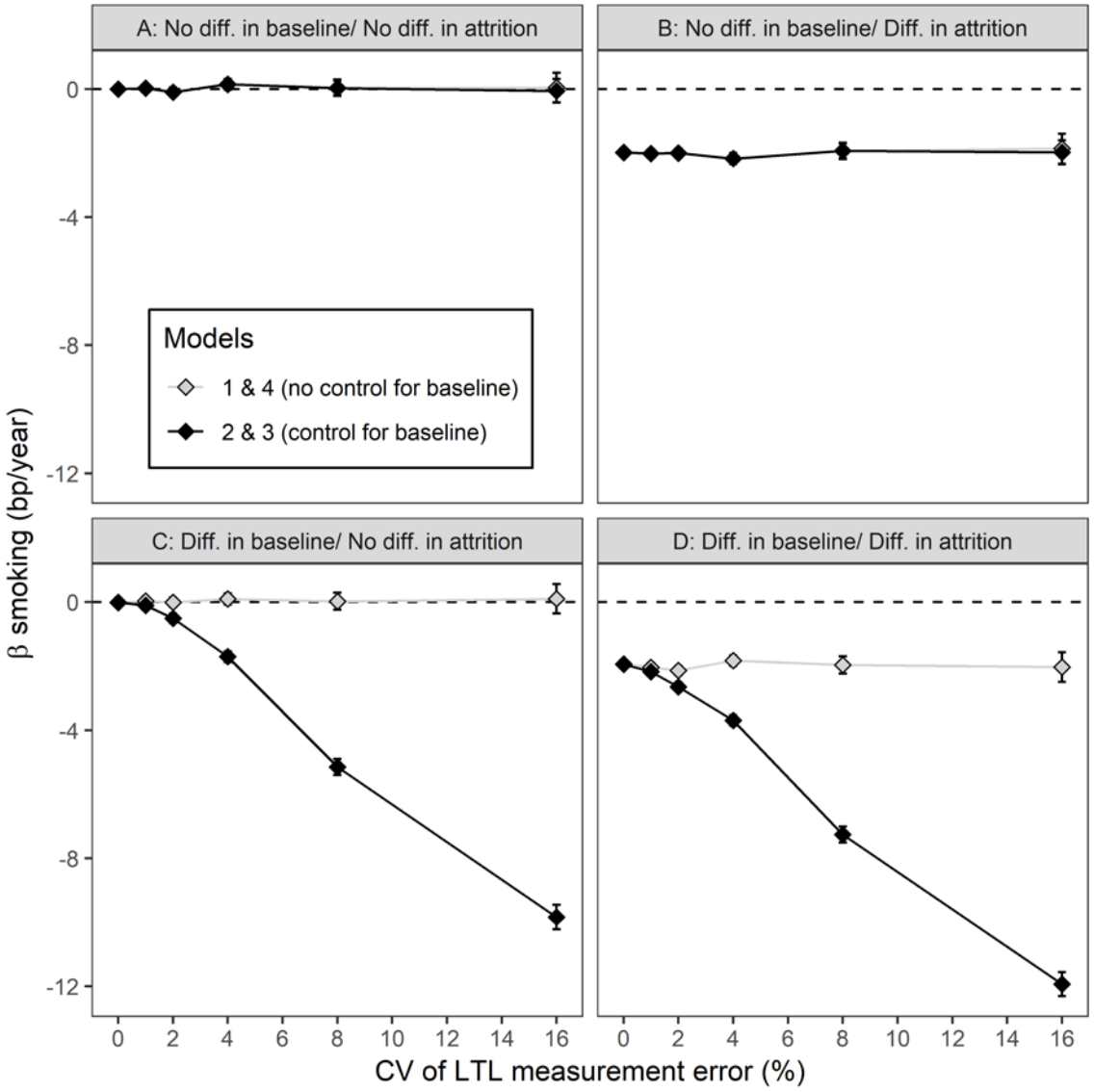

Figure 2. Controlling for LTL_b exaggerates estimates of the difference in ΔLTL between smokers and non-smokers when there is a difference in LTL_b. The estimated difference in mΔLTL between smokers and non-smokers as a function of measurement error (CV). The β estimates were obtained by fitting four alternative models to data simulated given four sets of assumptions regarding the true differences between smokers and non-smokers (scenarios A-D in Table 2). The dashed lines indicate no difference in mΔLTL between smokers and non-smokers. Data points are the mean ± 95% confidence intervals obtained from modelling the data from 1000 replicate simulations. The four scenarios were as follows: (A) no difference in LTL_b and no difference in ΔLTL; (B) no difference in LTL_b but a true difference in ΔLTL; (C) a true difference in LTL_b but no difference in ΔLTL; and (D) A true difference in LTL_b and a true difference in ΔLTL. The true difference in LTL_b between smokers and non-smokers in scenarios C and D was LTL_b 141 bp shorter in smokers. The true difference in ΔLTL between smokers and non-smokers in scenarios B and D was ΔLTL -2 $bp.year^{-1}$ greater in smokers.

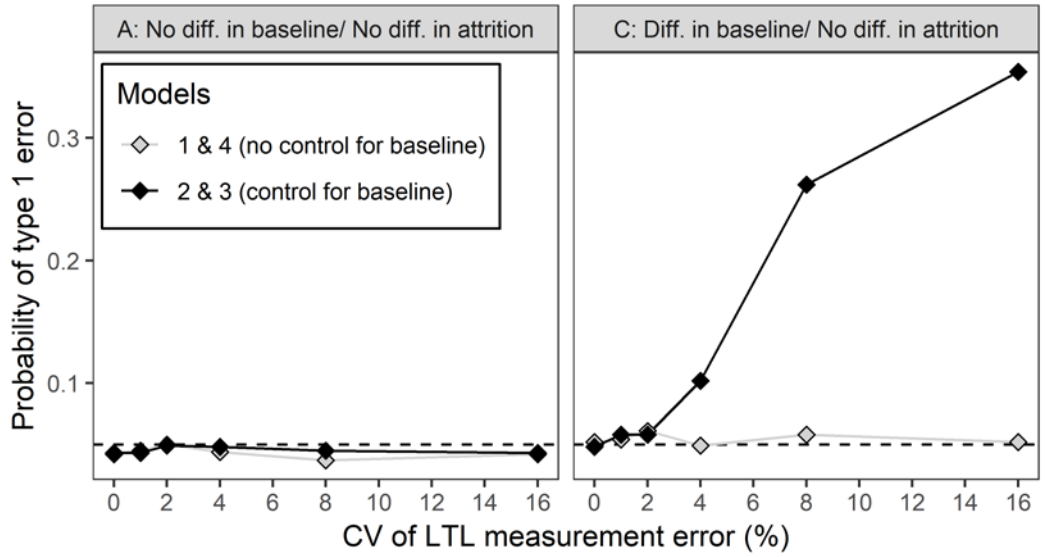

Figure 3. Controlling for LTL_b increases the probability of false-positive errors when there is a difference in LTL_b. Probability of a type 1 error as a function of measurement error (CV) for the four models under consideration. Data points represent the proportion of simulations yielding a p-value below 0.05 in 1000 replicate simulations. The left and right panels show the probability of type 1 errors in scenarios A and C respectively (corresponding with Figure 2). The difference in LTL_b between smokers and non-smokers in scenario C was LTL_b 141 bp shorter in smokers.

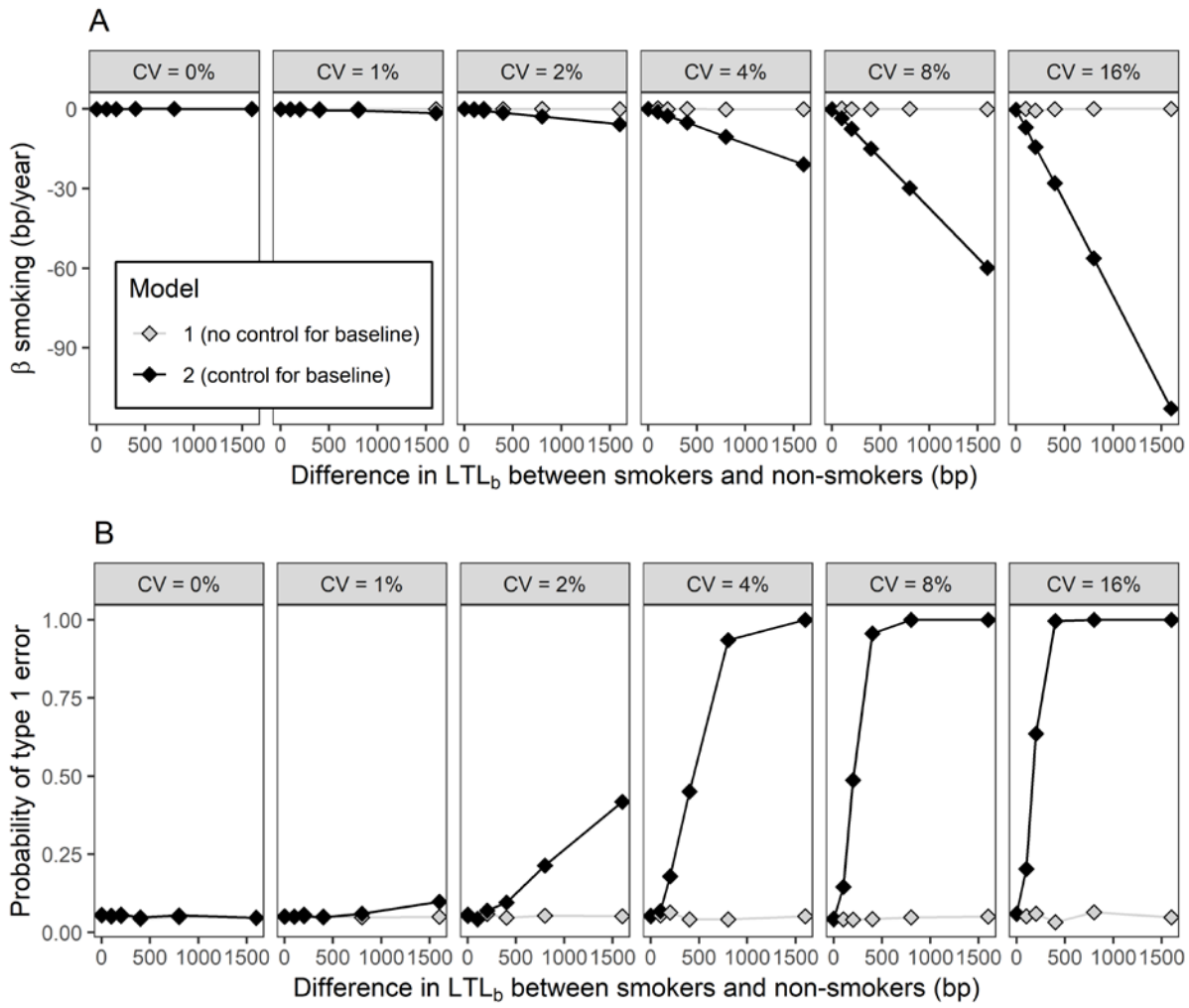

Figure 4. The bias caused by controlling for LTL_b is a synergistic interaction between difference in LTL_b and measurement error. The data in this figure come from a simulation of scenario C only (a true difference in LTL_b, but no difference in ΔLTL). Panel A shows the estimated difference in mΔLTL between smokers and non-smokers as a function of the difference in LTL_b and CV for models 1 and 2. Data points are the mean ± 95% confidence intervals obtained from modelling the data from 1000 replicate simulations. Panel B shows the probability of a type 1 error as a function of the difference in LTL_b and CV for models 1 and 2. Data points represent the proportion of simulations yielding a p-value below 0.05 in 1000 replicate simulations.

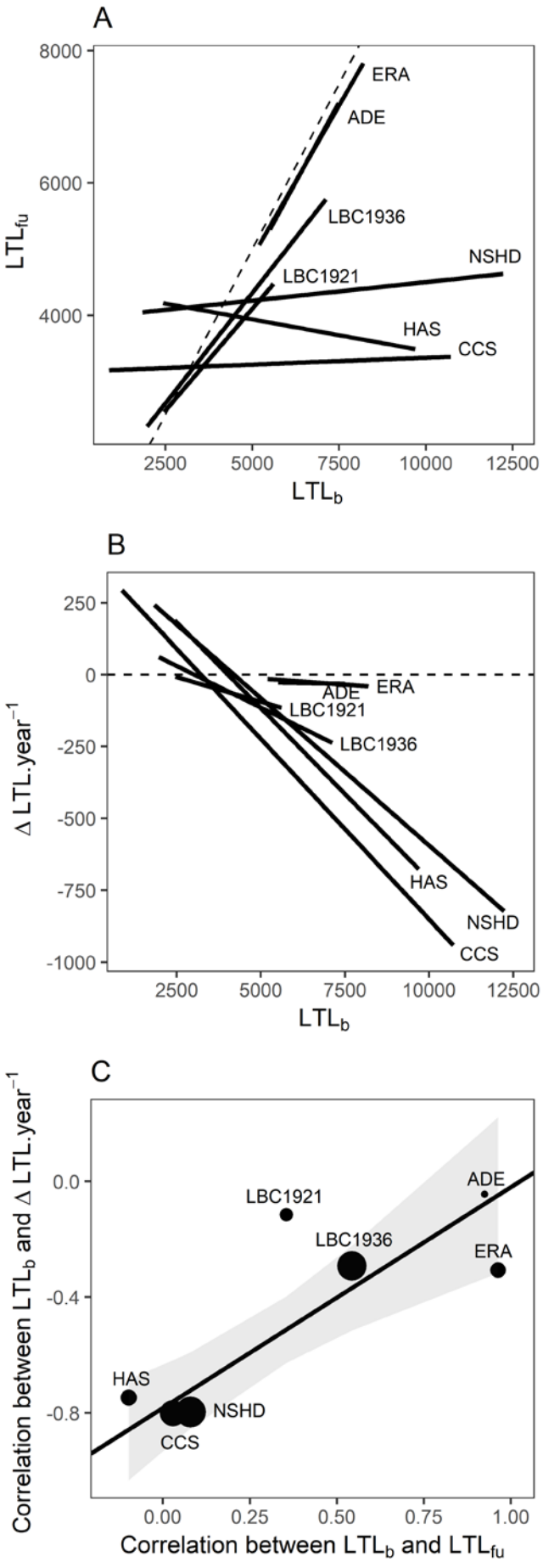

Figure 5. Signatures of measurement error differ between cohorts. A: The relationship between $mLTL_b$ and $mLTL_{fu}$ for each of the seven cohorts. The lines were obtained from simple linear regression. The dashed line shows the expectation if there is no change in mLTL between baseline and follow-up. Most of the data fall below the dashed line, indicating that in most participants, mLTL shortened between baseline and follow-up. Slopes closer to one indicate lower measurement error. B: The relationship between $mLTL_b$ and $m\Delta LTL.year^{-1}$ for each of the seven cohorts. The lines were obtained from simple linear regression. The dashed line shows the expectation if there is no measurement error. Flatter slopes indicate lower measurement error. C: Meta-regression between the correlation coefficients derived from the associations shown in panels A and B. The size of the point representing each cohort is proportional the number of participants. The solid black line was derived from a linear regression in which the points were weighted by the number of participants in each cohort and the grey ribbon shows the 95% confidence interval for this line. More positive values on both axes correspond to lower measurement error.

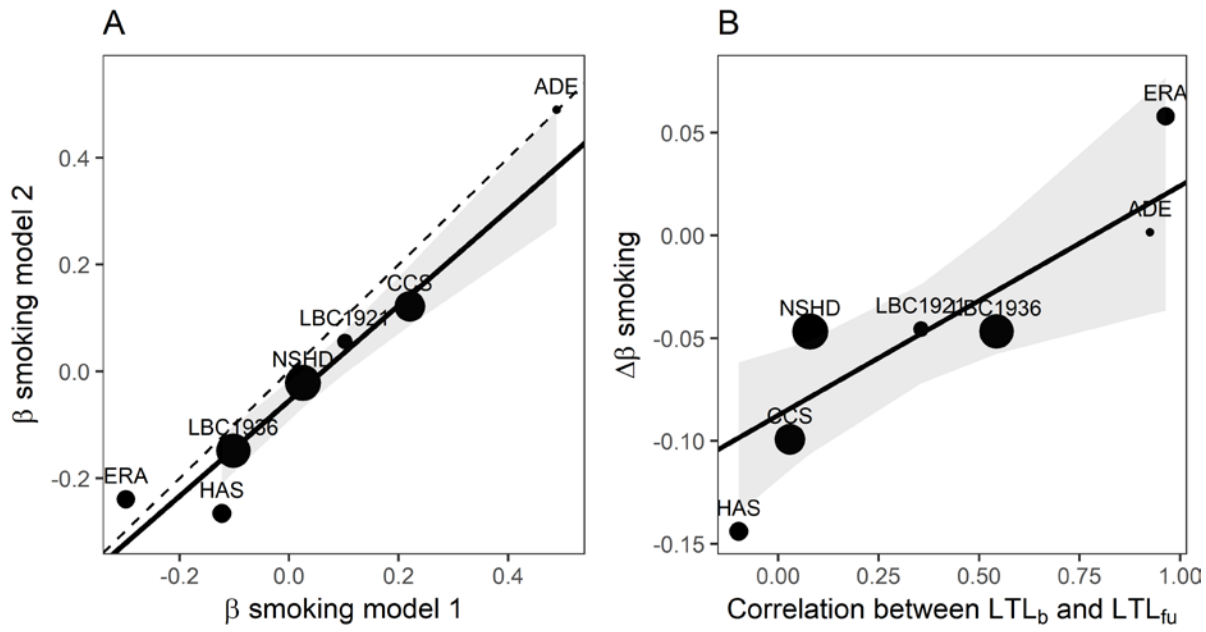

Figure 6. The biasing effect of controlling for LTL$_b$. A: The relationship between the β coefficients for smoking derived from models 1 (no control for LTL$_b$) and 2 (control for LTL$_b$). The dotted line shows the expectation if the coefficients were identical. B: The correlation between a signature of LTL measurement error (the correlation between LTL$_b$ and LTL$_{fu}$; larger values indicate lower measurement error) and the difference between the β coefficients derived from models 1 and 2. In both panels, the solid black line was derived from a linear regression in which the points were weighted by the number of participants in each cohort and the grey ribbon shows the 95% confidence interval for this line.

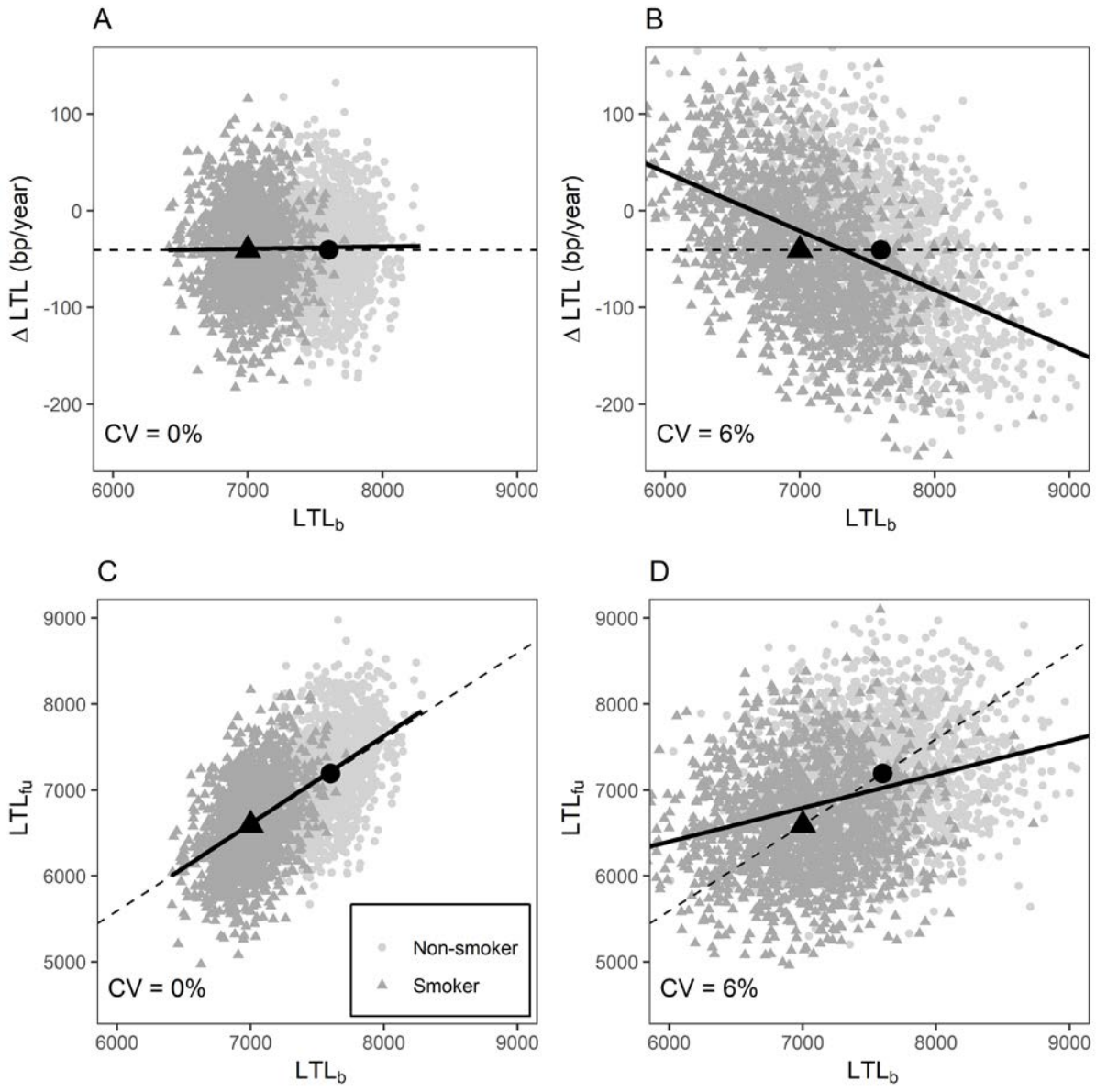

Figure 7. Graphical illustration of the biasing effect of controlling for LTL_b in analyses of ΔLTL and LTL_{fu}. This figure is based on simulated data and exaggerates the true difference in LTL between smokers and non-smokers. See text for explanation.