

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

Melissa Bateson<sup>1</sup> Dan T. A. Eisenberg<sup>2</sup> and Daniel Nettle<sup>1</sup>

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

2 Department of Anthropology, University of Washington, USA.

Email for correspondence: [melissa.bateson@ncl.ac.uk](mailto:melissa.bateson@ncl.ac.uk)

**In analyses of longitudinal changes in leukocyte telomere length (LTL) it is common practice to control statistically for baseline LTL. However, theoretical considerations arising from collider bias suggest that this practice could lead to overestimation of the difference in LTL attrition between groups that have experienced different exposures. We used simulated LTL data to explore whether adjusting for baseline LTL results in biased estimates of the true difference in LTL attrition between individuals with different exposures using smokers and non-smokers as an example. We show that if baseline LTL is shorter in smokers than non-smokers and LTL measurement error is non-zero, then adjusting for baseline LTL results in overestimating the true difference in telomere attrition between smokers and non-smokers. The size of this latter bias increases with increasing LTL measurement error. Since it is a robust finding that smokers have shorter baseline LTL than non-smokers and LTL measurement error is substantial, we conclude that the type 1 error rate for reports of effects of smoking on telomere attrition is likely to be above 5%. Using real data from seven longitudinal cohorts we show that in line with our simulation results, the estimated difference in attrition between smokers and non-smokers is greater in models controlling for baseline LTL. Furthermore, as predicted by our simulations, the size of this latter difference is positively associated with signatures of LTL measurement error. On the basis of our analyses we recommend that models of LTL attrition should not control for baseline LTL. Although we have couched our analysis in terms of the effects of smoking, our findings are likely to have general relevance to other factors studied in relation to telomere attrition. Many claims of accelerated LTL attrition in individuals exposed to disease, stress or adversity will need to be re-assessed.**

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

## 1 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 stress and adversity (Pepper, Bateson and Nettle, 2018). Recent meta-analyses show that LTL tends to be shorter in individuals who are smokers (Astuti *et al.*, 2017), are more sedentary (Mundstock, Zatti, *et al.*, 2015; Denham, O'Brien and Charchar, 2016), are obese (Mundstock, Sarria, *et al.*, 2015), were subjected to childhood trauma (Z. Li *et al.*, 2017) or psychosocial stress (Hanssen *et al.*, 2017), suffer from schizophrenia (Polho *et al.*, 2015; Rao *et al.*, 2016), post-traumatic stress disorder (X. Li *et al.*, 2017), anxiety or depression (Schutte and Malouff, 2015;

Ridout *et al.*, 2016) or have higher perceived stress (Mathur *et al.*, 2015). These latter studies have been widely assumed to support the hypothesis that the exposures studied increase 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 (Bateson and Nettle, 2018). To test the hypothesis that an exposure causes an increase in telomere attrition it is necessary to demonstrate that the same exposures associated with shorter LTL in cross-sectional studies also increase the rate of LTL attrition within individuals over time. To do this, telomere attrition is estimated from longitudinal datasets in which LTL is measured twice in each individual, first at baseline ( $LTL_b$ ) and again at follow-up ( $LTL_{fu}$ ). 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 ( $\Delta LTL$ ). In the current paper we express this difference such that negative values indicate telomere attrition and positive values telomere elongation; dividing by the follow-up interval provides an estimate of the annual rate of attrition ( $\Delta LTL \cdot \text{year}^{-1}$ ). Multiple regression approaches can then be used to estimate the rates of telomere attrition associated with different exposures during the follow-up period (Ehrlich *et al.*, 2009; e.g. Bendix *et al.*, 2014; Huzen *et al.*, 2014; Weischer, Bojesen and Nordestgaard, 2014; Müezziner *et al.*, 2015; Puterman *et al.*, 2015; Révész *et al.*, 2016). In the current paper we address the question of how these statistical models should be constructed in order to obtain unbiased estimates of the effects of a given exposure on the rate of telomere attrition. While our discussion is relevant to all of the exposures listed above, we have chosen to use smoking behaviour to illustrate the impact of different analytic strategies. Our rationale for choosing smoking is that it is one of the most commonly reported exposures in studies of telomere dynamics, meaning that many datasets are available for meta-analysis.

## 1.1 Controlling for baseline LTL

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 would imply including baseline LTL as a covariate in analyses of the effect of smoking on the rate of LTL attrition. Indeed, we have found nine studies that report the effect of smoking on LTL attrition and all of these control for baseline LTL in their multiple regression models (Aviv *et al.*, 2009; Ehrlich *et al.*, 2009; Farzaneh-Far *et al.*, 2010; Bendix *et al.*, 2014; Huzen *et al.*, 2014; Weischer, Bojesen and Nordestgaard, 2014; Müezziner *et al.*, 2015; Révész *et al.*, 2016; Toupance *et al.*, 2017). What are the arguments in favour of controlling for baseline telomere length?

In a highly-cited paper, Vickers and Altman (2001) considered the best analysis approach to adopt in controlled trials of an intervention with baseline and follow-up measurement. They concluded that analysis of covariance (which controls for baseline measurement in an analysis of change) will yield unbiased estimates of the effect of the intervention on the measured outcome variable of interest. Furthermore, they argue that this is generally the most powerful analytic approach, and the efficiency gains over analysing a simple change score will be greatest when the correlation between baseline and follow-up measurements is low. This latter paper is cited as the justification for controlling for baseline telomere length in at least one study of the factors affecting telomere attrition that we have found (Van Ockenburg *et al.*, 2015). In studies of telomere dynamics, the correlation between baseline and follow-up telomere measurements is often low due to measurement error; in a recent meta-analysis of 18 longitudinal telomere datasets we reported Pearson correlation coefficients with a range of 0-0.66 for measurements made with qPCR and 0.92-0.97 for those made with Southern blotting (Bateson *et al.*, no date). This might seem like a strong argument for controlling for baseline telomere length in analyses of change.

However, there is an important difference between controlled trials (the subject of Vickers and Altman, 2001) and epidemiological studies aimed at identifying the factors affecting the rate of telomere attrition. In controlled trials, subjects are randomly allocated to treatment groups at the start of the study, meaning that it is logically impossible for any baseline differences in the outcome variable being measured to be caused by the intervention applied between the baseline and follow-up measurements. In contrast, in epidemiological studies, there is no random allocation of subjects to exposures and exposures will often be present prior to the baseline measurement, meaning that it is quite possible that a baseline difference in telomere length is caused by the exposure or exposures under investigation. This difference between controlled trials and epidemiological studies is critically important, because the causal relationships between a set of variables determine the correct analytic strategy and inappropriate statistical control can introduce biases in estimates of effects on the outcome of interest (Greenland, 2003). Thus, in the current context, we need to understand whether controlling statistically for baseline telomere length biases estimates of the effect of smoking (or any other exposure) on the rate of telomere attrition.

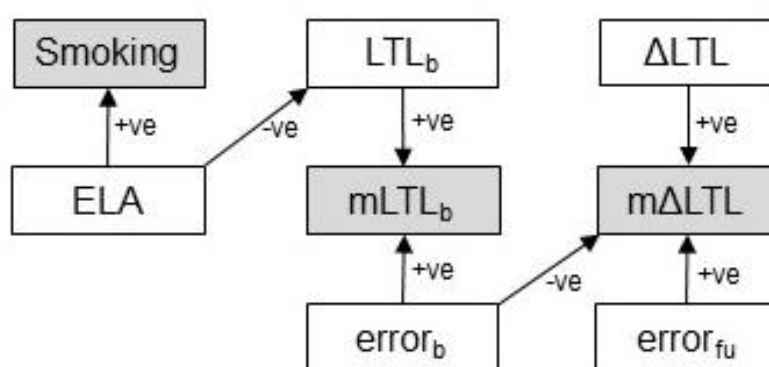
Glymour et al. (2005) examined the consequences of baseline control in asking whether educational attainment affects change in cognitive function in old age. They showed that in many plausible scenarios, controlling for baseline cognitive function induces a spurious statistical association between education and change in cognitive function. More generally, they concluded that when exposures are associated with baseline health status, an estimation bias arises if the change in health status precedes the baseline measurement or if there is measurement error in health status. Given that many exposures of interest—smoking being a prime example—will be present prior to baseline telomere measurement and that telomere length is known to be measured with error, there is strong reason to suspect that controlling for baseline telomere length could be a source of bias in analyses of telomere attrition.

## 1.2 Directed acyclic graphs

Directed acyclic graphs (DAGs), also known as causal diagrams, are recommended as a method for representing the causal relationships among a set of variables and for identifying the correct analytic strategy (Greenland, Pearl and Robins, 1999; Glymour *et al.*, 2005; Glymour and Greenland, 2008). In Figure 1 we use a DAG to present one possible hypothesis for the causal relationships among smoking, baseline telomere length and telomere attrition. The DAG in Figure 1 represents the null hypothesis that smoking does not affect the rate of telomere attrition; it assumes that the association between smoking and baseline telomere length is brought about by both variables being caused by exposure to early-life adversity (ELA). To reflect the presence of error in the measurement of LTL we distinguish between true and measured values of LTL and  $\Delta$ LTL; measured values are indicated with a prefix of m. Although we are ultimately interested in true LTL and  $\Delta$ LTL, these are latent variables that are not directly accessible to us, and any analyses must therefore use mLTL and m $\Delta$ LTL, explaining the presence of both true and measured values in the DAG. We assume that mLTL<sub>b</sub> is positively related to true LTL<sub>b</sub> and baseline measurement error (error<sub>b</sub>), and that m $\Delta$ LTL is positively related to  $\Delta$ LTL and follow-up measurement error (error<sub>fu</sub>). However, m $\Delta$ LTL must also be negatively related to baseline measurement error (see Appendix 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 measured baseline LTL and measured LTL attrition that is commonly observed in longitudinal telomere datasets (Verhulst *et al.*, 2013).

In Figure 1, a path connects smoking with mLTL<sub>b</sub> via ELA and LTL<sub>b</sub>. ELA is assumed to cause both smoking and LTL<sub>b</sub>. Thus, as long as ELA is not controlled for, a negative association will be present

between smoking and  $mLTL_b$ . A path also connects smoking with  $m\Delta LTL$  via  $ELA$ ,  $LTL_b$ ,  $mLTL_b$  and  $error_b$  (in DAGs, a path is a series of lines connecting two variables, regardless of arrow direction). On this path,  $mLTL_b$  is caused by both  $LTL_b$  and  $error_b$  and is therefore what is termed a ‘collider’. In the parlance of DAGs, a collider blocks a path, meaning that smoking is independent of  $m\Delta LTL$  (under our null hypothesis). However, controlling statistically for  $mLTL_b$  unblocks the path between smoking and  $m\Delta LTL$  and hence introduces a spurious association between smoking and  $m\Delta LTL$ . This latter phenomenon is known as ‘collider bias’ (Greenland, 2003; Cole *et al.*, 2010). In summary, it follows from the assumptions embodied in Figure 1 that controlling for  $mLTL_b$  should inflate estimates of the effect of smoking on LTL attrition via collider bias. The size of this bias should depend on both the strength of the association between smoking and  $LTL_b$  and the size of the measurement error.



**Figure 1.** Directed acyclic graph (DAG) summarising the assumed causal relations between smoking, measured baseline LTL ( $mLTL_b$ ) and measured LTL attrition ( $m\Delta LTL$ ; shaded boxes) and various unmeasured/latent variables including: exposure to early-life adversity ( $ELA$ ), true baseline telomere length ( $LTL_b$ ), baseline measurement error ( $error_b$ ), true telomere change ( $\Delta LTL$ ) and follow-up measurement error ( $error_{fu}$ ).  $error_b$  and  $error_{fu}$  are uncorrelated and independent of LTL and attrition. Causal relationships are indicated by arrows with the direction of the effect given next to the arrow. The absence of an arrow from smoking to  $\Delta LTL$  shows the assumption that smoking does not affect the rate of LTL attrition (the null hypothesis). This DAG is analogous to the DAGs presented in Glymour *et al.* (2005; Figure 3) and Glymour and Greenland (2008; Figure 12-14) and can thus be subjected to an identical analysis. See Section 1.2 for further details.

### 1.3 Aims

Since most of the longitudinal studies of the effects of smoking on telomere attrition also report an effect of smoking on baseline LTL, the above analysis raises the question of whether the results reported for the effects of smoking on LTL attrition are biased. Since measurement error affects baseline LTL, and baseline LTL affects apparent telomere attrition (via regression to the mean), we can ask how the extent of measurement error affects the size of this bias. In the remainder of this paper we address these questions via a combination of simulation and re-analysis of existing datasets. In section 2 we use a simulation model to ask whether, as predicted above, controlling for baseline LTL biases estimates of the association between smoking and telomere attrition. In section 3, we re-analyse empirical datasets to test the predictions arising from our simulation. Specifically, we explore the variation between datasets in signatures of measurement error and ask whether controlling for baseline LTL increases estimates of the association between smoking and LTL attrition.

## 2 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  $\Delta LTL$ ) are known. We can then verify how adding different magnitudes of measurement error and using different statistical analysis approaches affect our ability to estimate the values of these latent variables correctly. We simulated longitudinal LTL datasets in which we set the true differences between smokers and non-smokers in  $LTL_b$ ,  $\Delta LTL$  and the LTL measurement error ( $error_{r_b}$  and  $error_{r_u}$ ) based on realistic values obtained from the literature. We then used these simulated datasets to calculate the size of biases in the parameter estimates for the difference in  $\Delta LTL$  between smokers and non-smokers obtained from different statistical models in which we varied whether, and if so how, we controlled for  $LTL_b$ .

### 2.1 Methods

We simulated four different scenarios to describe the true differences in  $LTL_b$  and  $\Delta LTL$  between smokers and non-smokers: (A) No difference in  $LTL_b$  and no difference in  $\Delta LTL$ ; (B) No difference in  $LTL_b$ , but a true difference in  $\Delta LTL$ ; (C) A true difference in  $LTL_b$ , but no difference in  $\Delta LTL$ ; and (D) A true difference in  $LTL_b$  and a true difference in  $\Delta LTL$ . Since scenarios C and D both assumed a true difference in baseline LTL between smokers and non-smokers (with shorter LTL in smokers), we predicted that estimates of the difference in telomere attrition between smokers and non-smokers would be biased in these scenarios if measured baseline LTL ( $mLTL_b$ ) was included in the statistical models as a control variable. The parameter values used in each scenario were taken from Aviv et al. (2009), who reported a significant baseline difference in LTL between smokers and non-smokers of 141 bp and a non-significant  $m\Delta LTL$  between smokers and non-smokers of  $-2 \text{ bp}\cdot\text{year}^{-1}$  (Table 1).

**Table 1: 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 $\Delta LTL$	True diff. in $\Delta LTL$	No diff. in $\Delta LTL$	True diff. in $\Delta LTL$
Non-smokers	$LTL_b$ (bp; mean $\pm$ sd*)	7451 $\pm$ 777	7451 $\pm$ 777	7481 $\pm$ 777	7481 $\pm$ 777
	$\Delta LTL$ (bp $\cdot$ year $^{-1}$ ; mean $\pm$ sd*)	-40.7 $\pm$ 46	-40 $\pm$ 46	-40.7 $\pm$ 46	-40 $\pm$ 46
Smokers	$LTL_b$ (bp; mean $\pm$ sd*)	7451 $\pm$ 777	7451 $\pm$ 777	7392 $\pm$ 777	7392 $\pm$ 777
	$\Delta LTL$ (bp $\cdot$ year $^{-1}$ ; mean $\pm$ sd*)	-40.7 $\pm$ 46	-42 $\pm$ 46	-40.7 $\pm$ 46	-42 $\pm$ 46

\*Note that the standard deviations of baseline LTL and annual attrition in Table 1 are likely to be overestimates of the standard deviations of the true variables, 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.

The simulation of LTL values was based on one previously described by Bateson and Nettle (Bateson and Nettle, 2016) and was implemented in the statistical computing language R. The script for the simulation is available as supplementary information: "R\_script\_TL\_Bias\_2018\_final.R". 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 as given in Table 1. Each participant was then assigned a value of  $\Delta\text{LTL}\cdot\text{year}^{-1}$  by again drawing an independent random sample from normal distributions for attrition with means and standard deviations given in Table 1. This rate of attrition was applied for 10 years starting with the true  $\text{LTL}_b$  to yield a true  $\text{LTL}_{fu}$  for each participant. We assumed that each participant experienced a consistent rate of attrition over the follow-up interval (equivalent to setting  $r = 1$  in Bateson and Nettle's (2016) simulation). Measurement error was introduced into both baseline and follow-up LTL values by assuming that  $m\text{LTL}$  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  $\text{LTL}\cdot\text{CV}/100$  where CV is the coefficient of variation of the measurement error. Measured  $\Delta\text{LTL}$  for each participant was calculated as the difference between  $m\text{LTL}_b$  and  $m\text{LTL}_{fu}$ . We assumed values of CV of 0, 2, 4, 8, 10, 12 and 14%, and generated 1000 replicate data sets for each value of CV in each of the four scenarios (A, B, C and D).

Next, we modelled the dataset from each replicate with four different statistical models that represent alternative approaches found in the literature for modelling the difference in telomere attrition rates between individuals with different exposures (see Table 2). In these models,  $m\Delta\text{LTL}$ ,  $m\text{LTL}_b$  and  $m\text{LTL}_{fu}$  were continuous variables and smoking was categorical (smokers versus non-smokers). Model 1 is the basic model in which  $m\Delta\text{LTL}$  is predicted by smoking status with no statistical control for  $m\text{LTL}_b$  (e.g. Shlush *et al.*, 2011). Model 2 includes control for  $m\text{LTL}_b$  by including  $m\text{LTL}_b$  as a covariate; model 2 represents the approach most commonly adopted in the current telomere dynamics literature (e.g. Puterman *et al.*, 2015; Révész *et al.*, 2016; Toupance *et al.*, 2017). Model 3 is a variant of model 2 in which the response variable is  $m\text{LTL}_{fu}$  as opposed to  $m\Delta\text{LTL}$  (e.g. Shalev *et al.*, 2013; Carlson *et al.*, 2015). Model 4 is a repeated-measures equivalent of model 1 in which the response variable is  $m\text{LTL}$  and timepoint (baseline versus follow-up) is entered as a predictor (e.g. Shin *et al.*, 2008); in this model inclusion of the interaction between timepoint and smoking is necessary to test the hypothesis that attrition differs between smokers and non-smokers. Note that models 1 and 4 contain no control for  $m\text{LTL}_b$ , whereas models 2 and 3 control for  $m\text{LTL}_b$  by including it as a covariate.

**Table 2: The four statistical models compared.**

No.	Model	
	Response variable	Fixed predictor variable(s)
1	$m\Delta\text{LTL}$	Smoking
2	$m\Delta\text{LTL}$	$m\text{LTL}_b$ + Smoking
3	$m\text{LTL}_{fu}$	$m\text{LTL}_b$ + Smoking
4	$m\text{LTL}$	Timepoint + Smoking + Timepoint*Smoking <sup>1</sup>

<sup>1</sup>Model 4 additionally contains a random effect of participant to account for repeated measures.

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 'lme4' package (Bates *et al.*, 2015). To compare the estimates of the difference in  $m\Delta\text{LTL}\cdot\text{year}^{-1}$  between smokers and non-smokers produced by the different models we recorded the  $\beta$  coefficients for the 'Smoking' variable produced by models 1, 2 and 3 and the 'Timepoint\*Smoking' variable for model 4. To analyse the frequency of type 1 error rates (i.e. incorrectly rejecting the null hypothesis of no difference in attrition between smokers and non-smokers in the scenarios where there was no true difference in  $\Delta\text{LTL}$ ) we additionally recorded whether the latter  $\beta$  coefficients were significantly different from zero (at  $p < 0.05$ ). Summarised

output data from one run of the simulation are available as “Supplementary\_dataset.CSV”. These data were used to create Figures 2 and 3.

## 2.2 Results

### 2.2.1 Parameter estimates

Figure 2 shows the estimated difference in  $m\Delta LTL \cdot year^{-1}$  between smokers and non-smokers derived from the four models in each of the four different scenarios that we simulated. Scenarios A and C simulate datasets in which the null hypothesis of no difference in  $\Delta LTL$  between smokers and non-smokers is true. In scenario A, in which there is also no difference in  $LTL_b$  between smokers and non-smokers, all models correctly estimate the true difference in rates of attrition between smokers and non-smokers as zero. However, in scenario C, in which there is a true difference in  $LTL_b$ , while models 1 and 4 correctly estimate the difference in rates of attrition between smokers and non-smokers as zero, models 2 and 3 overestimate this difference at non-zero values of measurement error, and the overestimation increases as  $LTL$  measurement error increases.

Scenarios B and D simulate datasets in which there is a true difference in  $\Delta LTL$  between smokers and non-smokers of  $-2 \text{ bp} \cdot year^{-1}$  in smokers. In scenario B, in which there is no true difference in  $LTL_b$ , all models correctly estimate the difference in attrition between smokers and non-smokers at  $\sim -2 \text{ bp} \cdot year^{-1}$ . However, in scenario D, in which there is a true difference in  $LTL_b$  between smokers and non-smokers, while models 1 and 4 correctly estimate the difference in attrition between smokers and non-smokers, models 2 and 3 overestimate the difference at non-zero values of measurement error, and this overestimation increases as measurement error increases.

We also experimented with a more rarely-used statistical approach: correcting  $m\Delta LTL \cdot year^{-1}$  for expected regression to the mean using the formula supplied by Verhulst et al. (2013), and using this as the response variable (e.g. Nettle et al. 2015), either with or without control for  $mLTL_b$ . The results for these models exactly mirror models 1 and 2: the difference in  $\Delta LTL$  is correctly estimated without control for  $mLTL_b$ , but overestimated in proportion to measurement error if  $mLTL_b$  is controlled for (data not shown).

In summary, it appears that when there is a true difference in  $LTL_b$  between smokers and non-smokers (scenarios C and D), controlling for  $LTL_b$  by including it as a covariate (models 2 and 3) results in estimating an exaggerated difference (i.e. a bias) in the rate of attrition in smokers compared to non-smokers. This bias occurs whenever measurement error is greater than zero.

### 2.2.2 Type 1 errors

Figure 3 shows the probability of type 1 errors in scenarios A and C (where there is no true difference in  $\Delta LTL$  between smokers and non-smokers). In scenario A, the probability of type 1 errors with all models is around 0.05, as would be expected. However, in scenario C (where there is a difference in  $LTL_b$  between smokers and non-smokers) the type 1 error rates with models 2 and 3 reflect the exaggerated estimates of difference in  $\Delta LTL$  seen in Figure 2C, rising as measurement error increases.

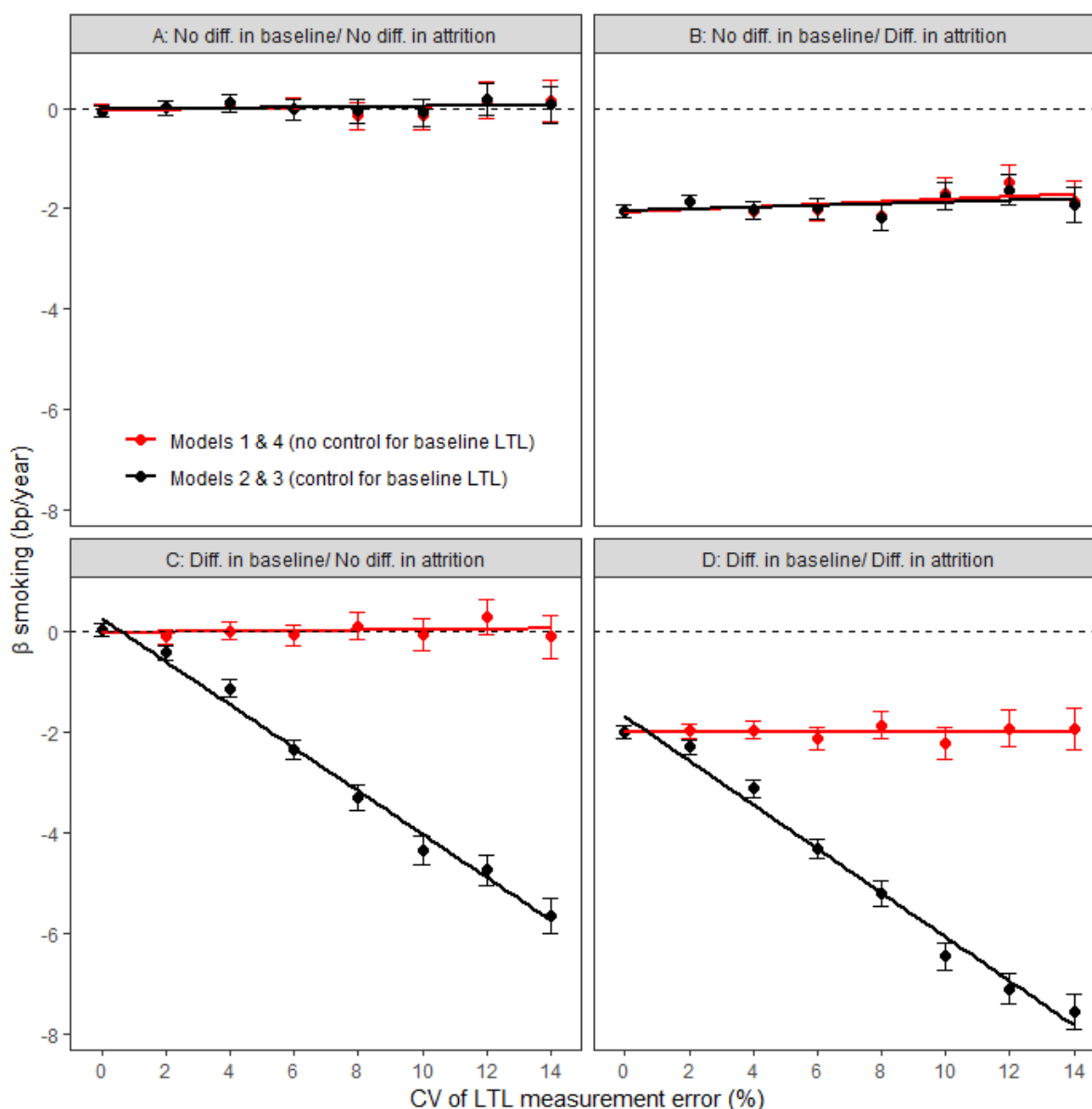
## 2.3 Discussion

Our results show that as long as there is no true difference in baseline  $LTL$  between smokers and non-smokers (scenarios A and B in our simulations), then all of the statistical modelling approaches that we have considered accurately estimate the difference in  $LTL$  attrition between smokers and non-smokers. However, if there is a true difference between smokers and non-smokers in baseline  $LTL$  (scenarios C and D) *and*  $LTL$  measurement error is non-zero, then controlling for baseline  $LTL$  biases estimates of the difference in attrition between smokers and non-smokers. Specifically, the



difference in attrition is overestimated and the size of this overestimation increases as LTL measurement error increases, for realistic values of measurement error. This overestimation of the difference in attrition translates into a type 1 error rate of above the usually-accepted 5% level in scenario D.

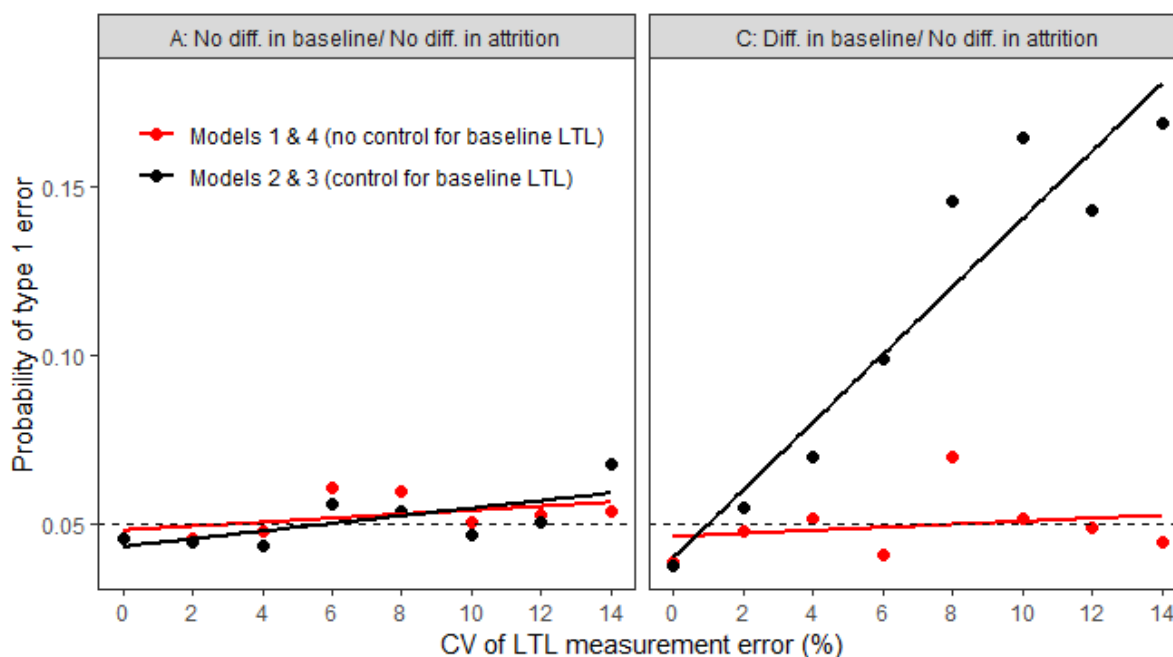
It is worth pointing out that scenario B is unlikely to be very common, unless the baseline LTL measurement is taken early in life, before the participants have started smoking. Likewise, scenario A is not typical, given the abundant cross-sectional evidence that smokers have shorter telomeres than non-smokers (Astuti *et al.*, 2017). Thus, the scenarios likely to be empirically widespread are exactly those (C and D) where bias will occur if baseline LTL is controlled for.



**Figure 2. Controlling for baseline LTL exaggerates estimates of the difference in telomere attrition between smokers and non-smokers.** The estimated effects of smoking on LTL attrition rate obtained from fitting four alternative models to data simulated given four sets of assumptions regarding the true differences between smokers and non-smokers (scenarios A-D). The dashed lines indicate no difference between smokers and non-smokers in LTL attrition. Data points are the mean  $\pm$  95% confidence intervals obtained from modelling the



data from 1000 replicate simulations and lines are simple linear fits. The four scenarios are as follows: (A) No difference in attrition and no difference in baseline LTL; (B) A true difference in attrition, but no difference in baseline LTL; (C) No difference in attrition, but a true difference in baseline LTL; and (D) A true difference in attrition and a true difference in baseline LTL. The true effect of smoking on LTL attrition rate in scenarios B and D was an additional  $-2 \text{ bp}\cdot\text{year}^{-1}$  in smokers. The true effect of smoking on baseline LTL in scenarios C and D was that smoker's baseline LTL were 141 bp shorter.



**Figure 3. Controlling for baseline LTL increases the probability of type 1 errors for detecting a difference in attrition between smokers and non-smokers.** Type 1 error rates for the four different different models under consideration. Data points represent the proportion of simulations yielding a p-value below 0.05 in 1000 replicate simulations, and lines are simple linear fits. Panels A and C show the probability of type 1 errors occurring in scenarios A and C where there is no true effect of smoking on the LTL attrition.

### 3 Does controlling for baseline LTL inflate estimates of the difference in attrition between smokers and non-smokers in real datasets?

In section 2 we showed, using a simulation model, that if there is a true difference in  $LTL_b$  between smokers and non-smokers and LTL measurement error is non-zero, then controlling for  $mLTL_b$  in models of  $m\Delta LTL$  results in overestimating the true difference in  $\Delta LTL$  between smokers and non-smokers. Furthermore, the size of this latter bias increases as LTL measurement error increases. Meta-analyses of real LTL data show that baseline LTL is shorter in smokers than non-smokers and that there is known to be substantial variation between studies in the magnitude of LTL measurement error. We therefore predict that in real longitudinal datasets, estimates of the difference in  $\Delta LTL$  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  $\Delta LTL$  between smokers and non-smokers should be larger when they are derived from models controlling for  $mLTL_b$ , and that the size of this effect of modelling strategy should increase as measurement error increases.

The aim of the current section is to test these predictions using real data from seven cohorts in which LTL was measured longitudinally. 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. This is important, because the existence of variation in measurement error is a pre-requisite for the subsequent analysis. Second, we tested whether the estimated association between smoking and  $m\Delta LTL$  is greater when the association is derived from a model controlling for  $LTL_b$  (model 2) compared a model without control for  $LTL_b$  (model 1), and whether any difference is explained by differences in LTL measurement error among cohorts. Finally, we asked whether the effects of modelling strategy that we have identified generalise from smoking behaviour to other putative influences on the rate of telomere attrition, namely sex and body mass index (BMI) (e.g. Farzaneh-Far *et al.*, 2010). Meta-analyses show that LTL is shorter in adult males (Gardner *et al.*, 2014) and in individuals with higher BMI (Mundstock, Sarria, *et al.*, 2015). We therefore predict that just as for smoking, estimates of the statistical effects of sex and BMI on  $m\Delta LTL$  could be exaggerated in models controlling for  $mLTL_b$  and that the size of this bias should be related to LTL measurement error.

### 3.1 Methods

We used data from participants in seven longitudinal cohorts whose LTL had been measured at least twice (Table 3). For the analyses of effects of smoking, we restricted the dataset to those participants who were either current or never smokers at the time of the baseline LTL measurement (designated ‘smokers’ and ‘non-smokers’ respectively); for clarity, those who had quit smoking at some point prior to the baseline measurement were excluded. For the analyses of sex and BMI we used the full dataset for which longitudinal LTL data were available. As our estimate of BMI, we used the mean of BMI at baseline and follow-up where this was available; and otherwise either BMI at baseline or BMI at follow-up, whichever was available. We were unable to analyse the effect of sex for the Caerphilly Cohort Study (CCS), since this cohort was restricted to male participants.

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  $\Delta LTL \cdot year^{-1}$  was calculated as  $(mLTL_{fu} - mLTL_b) / (age_{fu} - age_b)$  so that negative values indicate telomere attrition over the follow-up interval.

To characterise the measurement error present in each cohort we did not use the CVs reported for the cohorts, because there is considerable variation between labs in exactly which CVs are calculated and reported and consistent measures were not available for all cohorts. 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  $m\Delta LTL$ . All else being equal, the correlation between  $mLTL_b$  and  $mLTL_{fu}$  will be less positive the higher the measurement error, and the correlation between  $mLTL_b$  and  $m\Delta LTL$  will be more negative the higher the measurement error (Steenstrup *et al.*, 2013; Verhulst *et al.*, 2013).

For each cohort, we modelled the difference in  $m\Delta LTL \cdot year^{-1}$  between smokers and non-smokers using models 1 and 2 (Table 2). These models yielded estimates of the standardised  $\beta$  coefficient for the association between smoking and  $m\Delta LTL \cdot year^{-1}$ . 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 stronger negative association between smoking and  $m\Delta LTL \cdot year^{-1}$  in model 2 compared to model 1 will therefore be indicated by a more negative value of  $\Delta\beta$ .

**Table 3. Summary of the datasets analysed for smoking**

Cohort (acronym)	Country	Mean age at baseline (years)	Mean follow-up interval (years)	LTL measurement method	Number of participants by baseline smoking status <sup>1</sup>		Diff. in LTL <sub>b</sub> between smokers and never-smokers (Cohen's d) <sup>2</sup>	Signatures of LTL measurement error (data from smokers and never-smokers pooled)		Diff. in $\Delta$ LTL <sub>b</sub> .year <sup>-1</sup> between smokers and never-smokers (standardised $\beta$ [s.e.]) <sup>4</sup>		Reference for cohort
					Smokers	Never-smokers		Correlation between baseline LTL and follow-up LTL (r)	Correlation between baseline LTL and $\Delta$ LTL (r) <sup>3</sup>	Model 1	Model 2	
ADELAHYDE (ADE)	France	68.1	8.3	TRF	5	42	-0.99	0.93	-0.09	0.49 [0.47]	0.49 [0.50]	(Hjelmborg <i>et al.</i> , 2010)
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]	(Gardner <i>et al.</i> , 2013)
Evolution de la Rigidité Artérielle (ERA)	France	58.6	9.5	TRF	27	86	0.19	0.96	-0.32	-0.30 [0.22]	-0.24 [0.21]	(Toupance <i>et al.</i> , 2017)
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]	(Gardner <i>et al.</i> , 2013)
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]	(Gardner <i>et al.</i> , 2013)
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]	(Harris <i>et al.</i> , 2016)
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]	(Gardner <i>et al.</i> , 2013)

<sup>1</sup>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. <sup>2</sup>Negative numbers indicate that LTL<sub>b</sub> is shorter in smokers. <sup>3</sup>Negative numbers indicate that longer LTL<sub>b</sub> is associated with greater telomere loss. <sup>4</sup>Negative numbers indicate greater telomere loss in smokers.

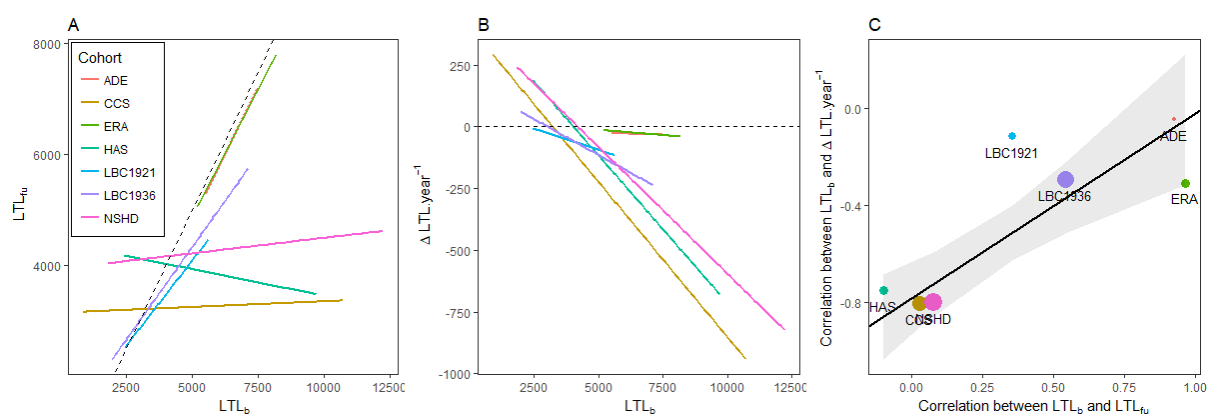
To compare the results obtained across the seven cohorts we used meta-regression, fitting linear regression models to the values obtained for each cohort. To account for the different numbers of participants per cohort, we weighted data points by the number of participants in each cohort.

## 3.2 Results

### 3.2.1 Descriptive statistics

We analysed data from the seven longitudinal cohorts detailed in Table 3. The combined dataset included data from 1,768 adults, comprising 550 current smokers and 1,218 never-smokers at the time of the baseline measurement. The mean age at baseline of the cohorts was  $65.9 \pm 8.5$  years (mean $\pm$ sd; range: 53.4-80.2) and the mean follow-up interval was  $8.5 \pm 1.2$  years (mean $\pm$ sd; range: 6.0-9.5) Signatures of measurement error

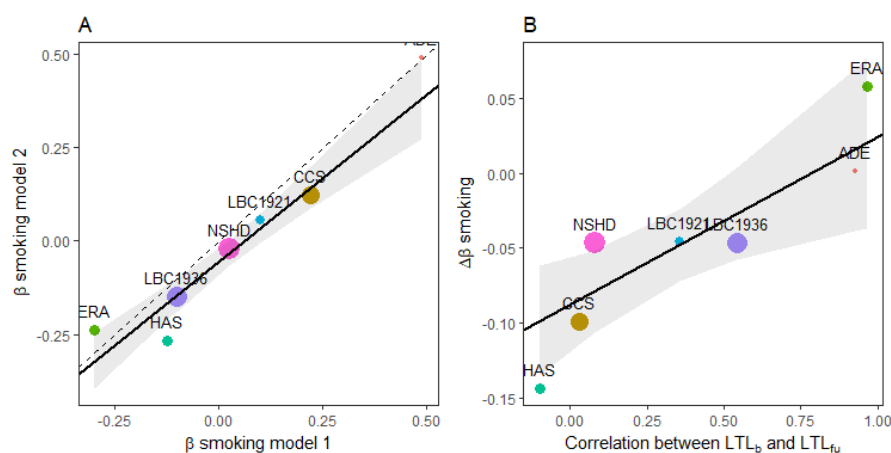
Five cohorts measured LTL using the quantitative polymerase chain reaction (qPCR) method and four used the arguably more precise terminal restriction fragment method (TRF; Aviv *et al.*, 2011). Figure 4A shows the correlation between  $mLTL_b$  and  $mLTL_{fu}$  for each cohort. For all cohorts, the correlation is less than one, as would be expected if there is independent measurement error present at both time points. However, the strength of the correlation differs markedly between cohorts, with Pearson correlation coefficients ranging from -0.01 to 0.97 (Table 3). As expected, the two cohorts measured with TRF (ADE and ERA) have the highest correlation coefficients. Figure 4B shows the correlations between  $mLTL_b$  and  $m\Delta LTL \cdot year^{-1}$ . For all cohorts, the correlation is negative, as would be expected given regression to the mean arising from measurement error. As expected, the two cohorts measured with TRF have the smallest negative correlation coefficients. Figure 4C shows that there is a significant negative association between the  $mLTL_b$ - $mLTL_{fu}$  correlation coefficient and the  $mLTL_b$ - $m\Delta LTL \cdot year^{-1}$  correlation coefficient for each cohort (weighted linear regression:  $\beta \pm se = -0.76 \pm 0.18$ ,  $t = -4.17$ ,  $p = 0.0088$ ). This latter relationship is expected if the relationships present in Figures 4A and 4B are both attributable to a common cause, in this case measurement error. Thus, predicted signatures of LTL measurement error are present and the magnitude of the signature varies among the seven cohorts.



**Figure 4. 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 least squares 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. B: The relationship between  $mLTL_b$  and  $m\Delta LTL \cdot year^{-1}$  for each of the seven cohorts. The lines were obtained from simple linear regression. C: Meta-regression between the  $mLTL_b$ - $mLTL_{fu}$  correlation coefficient and the  $mLTL_b$ - $m\Delta LTL \cdot year^{-1}$  correlation coefficient across cohorts. 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.

### 3.2.2 Effects of modelling strategy

We compared estimates of the difference in  $m\Delta LTL \cdot year^{-1}$  between smokers and non-smokers derived from simple models that did not control for  $mLTL_b$  (model 1) with estimates derived from models that controlled for  $mLTL_b$  (model 2). Table 3 reports the standardised  $\beta$  coefficients for the difference in  $m\Delta LTL \cdot year^{-1}$  between smokers and non-smokers derived from models 1 and 2 for each of the seven datasets. Coefficients from models 1 and 2 are strongly positively correlated, but not identical (Figure 5A; weighted linear regression:  $\beta \pm se = 0.89 \pm 0.11$ ,  $t = 8.15$ ,  $p = 0.0005$ ). There is a non-significant tendency for the coefficients from model 2 to be more negative, indicating a bigger estimated difference in  $m\Delta LTL \cdot year^{-1}$  between smokers and non-smokers compared to model 1 (paired t-test:  $t(6) = 1.87$ ,  $p = 0.1106$ ). If the latter comparison is restricted to the five cohorts measured with qPCR, then the difference between models 1 and 2 is significant (paired t-test:  $t(4) = 3.87$ ,  $p = 0.0180$ ). Further supporting the hypothesis that the difference in estimates derived from models 1 and 2 is caused by differences in measurement error, there is a significant positive relationship between the  $mLTL_b$ - $mLTL_{fu}$  correlation coefficient (a proxy for measurement error) for each cohort and  $\Delta\beta$  (Figure 5B; weighted linear regression  $\beta \pm se = 0.11 \pm 0.04$ ,  $t = 2.91$ ,  $p = 0.0336$ ).



**Figure 5. The biasing effect of controlling for baseline LTL.** A: The relationship between the beta coefficients for smoking derived from models 1 and 2. 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  $LTL_{fu}$ ) and the difference between the  $\beta$  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.

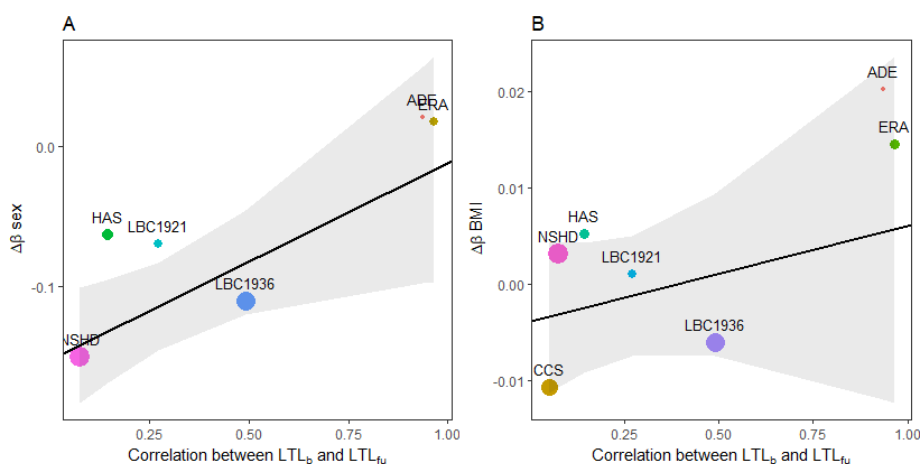
### 3.2.3 Generalisation to sex and BMI

The combined dataset available for analysing effects of sex and BMI included data from 3,313 adults, comprising 2,077 males and 1,236 females. Table 4 reports the results of repeating the analysis described in section 3.2.3, first with sex, and second with BMI in place of smoking status. As observed for smoking, there is a positive relationship between the  $LTL_b$ - $LTL_{fu}$  correlation coefficient (a proxy for measurement error) and  $\Delta\beta$  (the difference between the estimates derived from models 1 and 2) for both sex and BMI (Figures 6A and B). This relationship is marginally non-significant in the case of sex (weighted linear regression  $\beta \pm se = 0.14 \pm 0.06$ ,  $t = 2.43$ ,  $p = 0.0722$ ) and non-significant in the case of BMI (weighted linear regression  $\beta \pm se = 0.01 \pm 0.01$ ,  $t = 0.82$ ,  $p = 0.4480$ ).

**Table 4. Summary of the datasets analysed for sex and BMI.**

Cohort	Number of participants		Correlation between LTL <sub>b</sub> and LTL <sub>fu</sub>	Differences between sexes in telomere length/attrition <sup>1</sup> (standardised $\beta$ [s.e.])			Associations between BMI and telomere length/attrition <sup>2</sup> (standardised $\beta$ [s.e.])		
	Male	Female		LTL <sub>b</sub>	$\Delta$ LTL <sub>year</sub> <sup>-1</sup>		LTL <sub>b</sub>	$\Delta$ LTL <sub>year</sub> <sup>-1</sup>	
				Model 1	Model 2	Model 1	Model 2		
ADE	33	35	0.94	-0.55 [0.23]	-0.18 [0.24]	-0.16 [0.26]	-0.22 [0.12]	0.14 [0.12]	0.16 [0.13]
CCS	756	0	0.05	NA	NA	NA	0.01 [0.04]	0.023 [0.04]	0.012 [0.02]
ERA	108	54	0.96	-0.065 [0.17]	-0.01 [0.17]	0.01 [0.16]	-0.06 [0.08]	-0.05 [0.08]	-0.04 [0.08]
HAS	158	95	0.15	0.09 [0.13]	0.13 [0.13]	0.06 [0.09]	-0.01 [0.06]	0.06 [0.06]	0.07 [0.04]
LBC1921	78	81	0.27	0.37 [0.16]	-0.12 [0.16]	-0.19 [0.16]	-0.01 [0.08]	-0.01 [0.10]	-0.01 [0.08]
LBC1936	444	414	0.49	0.38 [0.07]	0.21 [0.07]	0.09 [0.07]	0.02 [0.03]	0.00 [0.03]	-0.00 [0.03]
NSHD	500	557	0.08	0.19 [0.06]	0.25 [0.06]	0.10 [0.04]	-0.00 [0.03]	-0.01 [0.03]	-0.01 [0.02]

Notes: <sup>1</sup>For sex, positive standardised  $\beta$ s indicate that males have longer LTL<sub>b</sub> and greater  $m\Delta$ LTL<sub>year</sub><sup>-1</sup>. <sup>2</sup>For BMI, positive standardised  $\beta$ s indicate that participants with higher BMI have longer LTL<sub>b</sub> and greater  $m\Delta$ LTL<sub>year</sub><sup>-1</sup>.



**Figure 6. Measurement error predicts biases for sex and BMI.** A: The correlation between a signature of LTL measurement error (the correlation between LTL<sub>b</sub> and LTL<sub>fu</sub>) and the difference between the  $\beta$  coefficients for sex derived from models 1 and 2. B: As panel A but the  $\beta$  coefficients are for BMI. 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.

### 3.3 Discussion

We compared the difference in telomere attrition between smokers and non-smokers estimated using two different statistical modelling strategies applied to telomere data from seven longitudinal cohorts. Model 1 contained no control for baseline telomere length, whereas model 2 controlled for baseline telomere length by including it as a covariate. Models 1 and 2 produced different estimates of the difference in the rate of telomere attrition between smokers and non-smokers. Specifically, we observed a tendency for the estimates derived from model 2 to suggest a more negative effect of smoking on telomere attrition than those derived from model 1. Since there can only be one true difference in the rate of telomere attrition between smokers and non-smokers, the  $\beta$  coefficients for smoking derived from either model 1 or model 2 (or both) must be incorrect. The fact that controlling for baseline LTL increases estimates of the effect of smoking rather than decreasing them suggests that baseline LTL is not a proxy for positive confounders of the difference in telomere attrition between smokers and non-smokers, but instead introduces a bias. Indeed, the DAG analysis in section 1.2 and our simulation results in section 2.2 both argue that controlling for baseline LTL (model 2) yields biased estimates, and that the correct estimates come from the model without control for baseline (model 1). Thus, it seems likely that model 2 is producing biased estimates of the effect of smoking on telomere attrition. This conclusion is strengthened by our finding that the discrepancy between the coefficients for smoking derived from models 1 and 2 is predicted by a proxy for the magnitude of the telomere measurement error present in a cohort. This latter relationship was predicted by our simulation that produces an association between the magnitude of bias and measurement error. Our results for smoking generalise to two other factors known to be associated with LTL and argued to influence attrition, namely sex and BMI. The weaker correlations for sex and particularly BMI are likely to reflect weaker evidence for true associations between sex and BMI and LTL<sub>b</sub> compared with the more robust association established for smoking.

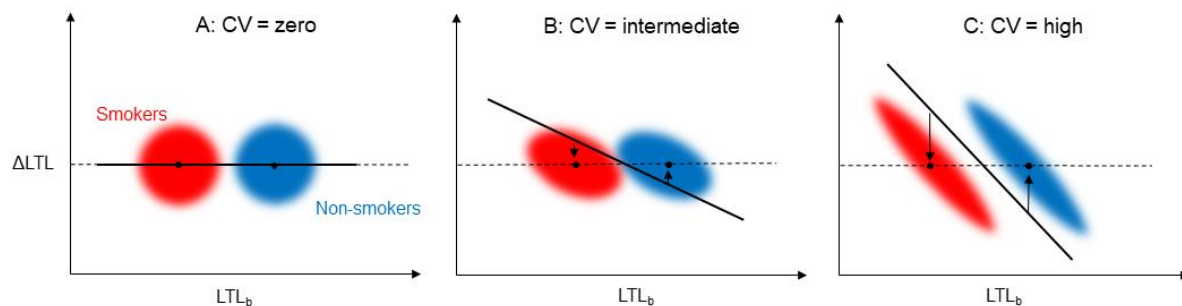
## 4 General Discussion

We have used three separate lines of evidence to argue that controlling for baseline telomere length in analyses of telomere attrition is likely to cause biases in estimates of the effects of exposures such as smoking. First, we used an analysis of directed acyclic graphs to show that under a realistic set of assumptions, baseline TL is likely to be a collider on the path linking smoking and telomere attrition. Controlling for baseline TL is therefore predicted to introduce collider bias in the form of an overestimation of the true difference in telomere attrition between smokers and non-smokers. Second, we used a simple simulation model to confirm, again under a realistic set of assumptions, that controlling for baseline TL does indeed inflate estimates of the true difference in telomere attrition between smokers and non-smokers, but only when a true difference in TL 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 baseline TL was included in statistical models as a covariate. Furthermore, the magnitude of this latter difference was predicted by TL 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 baseline TL is problematic. Figure 7 is our attempt to provide a graphical explanation for why the above bias occurs. The red and blue clouds indicate distributions of LTL<sub>b</sub> and  $\Delta$ LTL measurements for smokers and non-smokers respectively. The three panels depict cartoon measured LTL data from a scenario in which there is no true difference in  $\Delta$ LTL between smokers and non-smokers (the centres of the clouds for smokers and non-smokers lie on the same horizontal dotted line indicating an identical



level of attrition), but there is a true baseline LTL difference, with smokers having shorter LTL<sub>b</sub> than non-smokers (the centre of the cloud for smokers is to the left of the centre of the cloud for non-smokers). This scenario is equivalent to that simulated in scenario C in section 2. All three panels of Figure 7 show the relationship between baseline LTL<sub>b</sub> and  $\Delta$ LTL as a solid black regression line (note that this is the same relationship shown in Figure 4B). As LTL measurement error increases above zero, a negative relationship between LTL<sub>b</sub> and  $\Delta$ LTL is introduced as a result of regression to the mean. The slope of the regression line rotates about the mean values of LTL<sub>b</sub> and  $\Delta$ LTL, becoming more negative as the CV increases (panels B and C). Controlling for baseline LTL in an analysis of  $\Delta$ LTL is conceptually equivalent to analysing the residuals from the regression of  $\Delta$ LTL on LTL<sub>b</sub>. The effect of the rotation of the regression line that occurs as measurement error increases is to create mean negative residuals of the data from the regression line for smokers and mean positive residuals for non-smokers (indicated by the vertical arrows between the centre of each cloud and the regression line). Thus, a spurious difference in the residual  $\Delta$ LTL between smokers and non-smokers is created, despite the fact that no true difference in  $\Delta$ LTL exists. This bias only occurs because the smokers have a mean LTL<sub>b</sub> that is lower than that of non-smokers; it would not occur if there was no true difference in LTL<sub>b</sub>.



**Figure 7. Cartoon illustrating the biasing effect of controlling for LTL<sub>b</sub> in analyses of  $\Delta$ LTL.** See text for explanation.

Given first, that there are robust differences in baseline TL between smokers and non-smokers (Aviv *et al.*, 2009; Huzen *et al.*, 2014; Weischer, Bojesen and Nordestgaard, 2014; Müezziner *et al.*, 2015; Révész *et al.*, 2016), second, that TL measurement error is substantial (Aviv *et al.*, 2011) and third, that most published analyses of the effect of smoking on telomere attrition control for baseline TL, we predict that the difference in LTL attrition between smokers and non-smokers is likely to have been overestimated. Reports of significantly accelerated LTL attrition in smokers compared to non-smokers should therefore be interpreted with caution (e.g. Bendix *et al.*, 2014; Huzen *et al.*, 2014). In a recent meta-analysis in which we re-analysed LTL data from 18 longitudinal cohorts without control for baseline LTL, we found no evidence to support accelerated LTL attrition in adult smokers (Bateson *et al.*, no date).

Our findings are likely to have much broader impact than the specific cases of the effects of smoking, BMI and sex. Our findings are relevant to estimating the effect of any factor that is associated with a true difference in LTL at the time of baseline measurement on the rate of subsequent LTL attrition. The same considerations apply to the study of the association of LTL attrition with any disease, stress or adversity that has been linked with shorter baseline LTL in cross-sectional studies (Pepper, Bateson and Nettle, 2018). There is a growing literature claiming that exposure to various forms of stress and adversity accelerates LTL attrition, based predominantly on cross-sectional data (Epel *et al.*, 2004; Damjanovic *et al.*, 2007; Ahola *et al.*, 2012; Humphreys *et al.*, 2012; Ala-Mursula *et al.*,

2013; Litzelman *et al.*, 2014; Verhoeven *et al.*, 2014, 2015; Park *et al.*, 2015; Puterman *et al.*, 2016). While cross-sectional associations between exposure to stress and short LTL do not prove that stress causes LTL attrition (Bateson and Nettle, 2018), longitudinal studies have started to emerge that appear to support a causal relationship (e.g. Shalev *et al.*, 2013; Puterman *et al.*, 2015; Révész *et al.*, 2016). Unfortunately, just as in the literature on effects of smoking, it is typical for authors to control for baseline LTL in these latter studies, meaning that the results should be treated with caution. Re-analyses of these latter datasets is required to establish whether the claimed differences in LTL attrition are in fact biases introduced by incorrect control for baseline LTL. Specifically, we predict that removing baseline LTL 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 baseline LTL were an innocuous incidental variable that needed to be controlled for to increase power). Instead, it will systematically shift the parameter estimates for the effect of the exposure on LTL attrition towards zero.

#### 4.1 Conclusions

We have shown that controlling statistically for baseline LTL incorrectly inflates estimates of the difference in LTL attrition between smokers and non-smokers, and that the size of this bias is positively related to the size of LTL measurement error. Furthermore, we have argued that this bias is not restricted to smoking and will occur for any factor that, like smoking, is associated with shorter LTL at the time of the baseline LTL measurement. On the basis of our analyses we recommend that models of LTL attrition should not control for baseline LTL. Given that the majority of previous analyses of factors affecting LTL attrition control for baseline LTL, many claims of accelerated LTL attrition in individuals exposed to disease, stress or adversity need to be re-assessed.

## 5 Acknowledgements

DN is funded by an Advanced Grant from the European Research Council (AdG 666669, COMSTAR). 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 and Rachel Cooper (MRC National Survey of Health and Development: doi: 10.5522/NSHD/Q101; doi: 10.5522/NSHD/Q102), Ian Deary (Lothian Birth Cohorts 1921 and 1936). The MRC NSHD is funded by the UK Medical Research Council. NSHD data used in this publication ( ) are available to bona fide researchers upon request to the NSHD Data Sharing Committee via a standard application procedure. Further details can be found at <http://www.nshd.mrc.ac.uk/data>.

## 6 Supplementary information

- “R\_script\_TL\_Bias\_2018\_final.R”: Script for the simulations presented in Section 2.
- “Supplementary\_dataset.CSV”: Simulated dataset used to produce Figures 2 and 3.

## 7 Appendix 1

$\Delta$ LTL is estimated from longitudinal datasets in which LTL is measured twice, at baseline (mLTL<sub>b</sub>) and follow up (mLTL<sub>fu</sub>). The measured  $\Delta$ LTL per year (m $\Delta$ LTL·year<sup>-1</sup>) for the *i*th individual is calculated via the following formula:

$$\text{m}\Delta\text{LTL}\cdot\text{year}^{-1}_i = (\text{mLTL}_{\text{fu},i} - \text{mLTL}_{\text{b},i}) / (\text{age}_{\text{b},i} - \text{age}_{\text{fu},i}) \quad (\text{Equation 1})$$

Thus, a negative value of m $\Delta$ LTL indicates telomere attrition and a positive value telomere elongation. An individual's measured LTL can be written as the sum of their true LTL and a measurement error:

$$\text{mLTL}_{\text{b},i} = \text{LTL}_{\text{b},i} + \text{error}_{\text{b},i} \quad (\text{Equation 2})$$

$$\text{mLTL}_{\text{fu},i} = \text{LTL}_{\text{fu},i} + \text{error}_{\text{fu},i} \quad (\text{Equation 3})$$

Here, error<sub>b,i</sub> and error<sub>fu,i</sub> are the errors introduced by measurement for that individual at baseline and follow-up respectively. We assume that error<sub>b,i</sub> and error<sub>fu,i</sub> are drawn from independent distributions. Equation 1 can now be expressed in terms of equations 2 and 3:

$$\begin{aligned} \text{m}\Delta\text{LTL}_i &= \text{LTL}_{\text{fu},i} + \text{error}_{\text{fu},i} - (\text{LTL}_{\text{b},i} + \text{error}_{\text{b},i}) \\ &= \Delta\text{LTL}_i + \text{error}_{\text{fu},i} - \text{error}_{\text{b},i} \end{aligned} \quad (\text{Equation 4})$$

From equation 4 it is evident that there is an inverse relationship between m $\Delta$ LTL<sub>i</sub> and error<sub>b,i</sub>. In other words, a larger positive baseline measurement error for an individual results in a more negative m $\Delta$ LTL, which implies greater measured telomere attrition, for that individual. This is an example of so-called regression to the mean: baseline values are negatively correlated with measures of change because individuals with high mLTL<sub>b</sub> generally have smaller mLTL<sub>fu</sub> and vice versa.

## 8 References

- Ahola, K. *et al.* (2012) 'Work-related exhaustion and telomere length: A population-based study', *PLoS ONE*, 7(7), pp. 13–15. doi: 10.1371/journal.pone.0040186.
- Ala-Mursula, L. *et al.* (2013) '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*, 8(11). doi: 10.1371/journal.pone.0080094.
- Astuti, Y. *et al.* (2017) 'Cigarette smoking and telomere length: A systematic review of 84 studies and meta-analysis', *Environmental Research*. Elsevier Inc., 158, pp. 480–489. doi: 10.1016/j.envres.2017.06.038.
- Aviv, A. *et al.* (2009) 'Leukocyte telomere dynamics: Longitudinal findings among young adults in the Bogalusa Heart Study', *American Journal of Epidemiology*, 169(3), pp. 323–329. doi: 10.1093/aje/kwn338.
- Aviv, A. *et al.* (2011) 'Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR', *Nucleic Acids Research*, 39(20), p. e134. doi: 10.1093/nar/gkr634.
- Bates, D. *et al.* (2015) 'Fitting linear mixed-effects models using {lmer}', *Journal of Statistical*

*Software*, 67(1), pp. 1–48. doi: 10.1159/000323281.

Bateson, M. *et al.* (no date) 'Smoking does not accelerate leukocyte telomere attrition: a meta-analysis of 18 longitudinal cohorts', *Nicotine and Tobacco Research*.

Bateson, M. and Nettle, D. (2016) 'The telomere lengthening conundrum - it could be biology', *Aging Cell*, 1474–9726(2). doi: 10.1111/accel.12555.

Bateson, M. and Nettle, D. (2018) 'Why are there associations between telomere length and behaviour?', *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 373, p. 20160438. doi: 10.1098/rstb.2016.0438.

Bendix, L. *et al.* (2014) 'Longitudinal changes in leukocyte telomere length and mortality in humans', *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 69 A(2), pp. 231–239. doi: 10.1093/gerona/glt153.

Carlson, L. E. *et al.* (2015) 'Mindfulness-based cancer recovery and supportive-expressive therapy maintain telomere length relative to controls in distressed breast cancer survivors', *Cancer*, 121(3), pp. 476–484. doi: 10.1002/cncr.29063.

Cole, S. R. *et al.* (2010) 'Illustrating bias due to conditioning on a collider', *International Journal of Epidemiology*, 39(2), pp. 417–420. doi: 10.1093/ije/dyp334.

Damjanovic, A. K. *et al.* (2007) 'Accelerated Telomere Erosion Is Associated with a Declining Immune Function of Caregivers of Alzheimer's Disease Patients', *The Journal of Immunology*, 179(6), pp. 4249–4254. doi: 10.4049/jimmunol.179.6.4249.

Denham, J., O'Brien, B. J. and Charchar, F. J. (2016) 'Telomere length maintenance and cardio-metabolic disease prevention through exercise training', *Sports Medicine*. Springer International Publishing, 46(9), pp. 1213–1237. doi: 10.1007/s40279-016-0482-4.

Ehrlenbach, S. *et al.* (2009) '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', *International Journal of Epidemiology*, 38(6), pp. 1725–1734. doi: 10.1093/ije/dyp273.

Epel, E. S. *et al.* (2004) 'Accelerated telomere shortening in response to life stress.', *Proceedings of the National Academy of Sciences of the United States of America*, 101(49), pp. 17312–5. doi: 10.1073/pnas.0407162101.

Farzaneh-Far, R. *et al.* (2010) 'Telomere length trajectory and its determinants in persons with coronary artery disease: Longitudinal findings from the heart and soul study', *PLoS ONE*, 5(1), p. e8612. doi: 10.1371/journal.pone.0008612.

Gardner, M. *et al.* (2014) 'Gender and telomere length: systematic review and meta-analysis.', *Experimental gerontology*, 51, pp. 15–27. doi: 10.1016/j.exger.2013.12.004.

Gardner, M. P. *et al.* (2013) 'Telomere length and physical performance at older ages: an individual participant meta-analysis', *PLoS ONE*, 8(7). doi: 10.1371/journal.pone.0069526.

Glymour, M. M. *et al.* (2005) 'When is baseline adjustment useful in analyses of change? An example with education and cognitive change', *American Journal of Epidemiology*, 162(3), pp. 267–278. doi: 10.1093/aje/kwi187.

Glymour, M. M. and Greenland, S. (2008) 'Causal Diagrams', in Rothman, K. T., Sander, G., and Lash, T. L. (eds) *Modern Epidemiology*. 3rd. Philadelphia, USA., pp. 183–209.

Greenland, S. (2003) 'Quantifying biases in causal models: classical confounding vs collider-stratification bias', *Epidemiology*, 14(3), pp. 300–306. doi: 10.1097/01.EDE.0000042804.12056.6C.

- Greenland, S., Pearl, J. and Robins, J. M. (1999) 'Causal Diagrams for Epidemiologic Research', *Epidemiology*, 10(1), pp. 37–48. doi: 10.1097/00001648-199901000-00008.
- Hanssen, L. M. *et al.* (2017) 'The relationship between childhood psychosocial stressor level and telomere length: a meta-analysis', *Health Psychology Research*, 5(1). doi: 10.4081/hpr.2017.6378.
- Harris, S. E. *et al.* (2016) 'Longitudinal telomere length shortening and cognitive and physical decline in later life: The Lothian Birth Cohorts 1936 and 1921', *Mechanisms of Ageing and Development*. Elsevier Ireland Ltd, 154, pp. 43–48. doi: 10.1016/j.mad.2016.02.004.
- Hjelmborg, B. *et al.* (2010) 'Leukocyte telomere length is inversely correlated with plasma Von Willebrand factor', *Thrombosis Research*. Elsevier Ltd, 125(6), pp. e339–e342. doi: 10.1016/j.thromres.2010.03.006.
- Humphreys, J. *et al.* (2012) 'Telomere shortening in formerly abused and never abused women', *Biological research for nursing*, 14(2), pp. 115–123. doi: 10.1177/1099800411398479.
- Huzen, J. *et al.* (2014) 'Telomere length loss due to smoking and metabolic traits', *Journal of Internal Medicine*, 275, pp. 155–163. doi: 10.1111/joim.12149.
- Li, X. *et al.* (2017) 'The association between post-traumatic stress disorder and shorter telomere length: A systematic review and meta-analysis', *Journal of Affective Disorders*. Elsevier B.V., 218(March), pp. 322–326. doi: 10.1016/j.jad.2017.03.048.
- Li, Z. *et al.* (2017) 'Association between childhood trauma and accelerated telomere erosion in adulthood: A meta-analytic study', *Journal of Psychiatric Research*. Elsevier Ltd, 93, pp. 64–71. doi: 10.1016/j.jpsychires.2017.06.002.
- Litzelman, K. *et al.* (2014) 'Association between informal caregiving and cellular aging in the survey of the health of wisconsin: The role of caregiving characteristics, stress, and strain', *American Journal of Epidemiology*, 179(11), pp. 1340–1352. doi: 10.1093/aje/kwu066.
- Mathur, M. B. *et al.* (2015) 'Perceived stress and telomere length: A systematic review, meta-analysis, and methodologic considerations for advancing the field', *Brain, Behavior, and Immunity*. Elsevier Inc., 54, pp. 158–169. doi: 10.1016/j.bbi.2016.02.002.
- Müezzinler, A. *et al.* (2015) 'Smoking habits and leukocyte telomere length dynamics among older adults: results from the ESTHER cohort', *Experimental Gerontology*. Elsevier Inc., 70, pp. 18–25. doi: 10.1016/j.exger.2015.07.002.
- Mundstock, E., Sarria, E. E., *et al.* (2015) 'Effect of obesity on telomere length: systematic review and meta-analysis', *Obesity*, 23(11), pp. 2165–2174. doi: 10.1002/oby.21183.
- Mundstock, E., Zatti, H., *et al.* (2015) 'Effects of physical activity in telomere length: Systematic review and meta-analysis', *Ageing Research Reviews*. Elsevier B.V., 22, pp. 72–80. doi: 10.1016/j.arr.2015.02.004.
- Van Ockenburg, S. L. *et al.* (2015) 'Stressful life events and leukocyte telomere attrition in adulthood: a prospective population-based cohort study', *Psychological Medicine*, 45(July), pp. 2975–2984. doi: 10.1017/S0033291715000914.
- Park, M. *et al.* (2015) 'Where you live may make you old: The association between perceived poor neighborhood quality and leukocyte telomere length', *PLoS ONE*, 10(6), pp. 1–13. doi: 10.1371/journal.pone.0128460.
- Pepper, G. V., Bateson, M. and Nettle, D. (2018) 'Telomeres as integrative markers of exposure to stress and adversity: A systematic review and meta-analysis', *Royal Society Open Science*, in press. doi: <https://doi.org/10.1101/320150>.

Polho, G. B. *et al.* (2015) 'Leukocyte telomere length in patients with schizophrenia: A meta-analysis', *Schizophrenia Research*. Elsevier B.V., 165(2–3), pp. 195–200. doi: 10.1016/j.schres.2015.04.025.

Puterman, E. *et al.* (2015) 'Determinants of telomere attrition over 1 year in healthy older women: stress and health behaviors matter', *Molecular Psychiatry*. Nature Publishing Group, 20(4), pp. 529–535. doi: 10.1038/mp.2014.70.

Puterman, E. *et al.* (2016) 'Lifespan adversity and later adulthood telomere length in the nationally representative US Health and Retirement Study', *Proceedings of the National Academy of Sciences*, 113(42), pp. E6335–E6342. doi: 10.1073/pnas.1525602113.

Rao, S. *et al.* (2016) 'Accelerated leukocyte telomere erosion in schizophrenia: Evidence from the present study and a meta-analysis', *Journal of Psychiatric Research*, 79, pp. 50–56. doi: 10.1016/j.jpsychires.2016.04.010.

Révész, D. *et al.* (2016) 'Baseline biopsychosocial determinants of telomere length and 6-year attrition rate', *Psychoneuroendocrinology*. Elsevier Ltd, 67, pp. 153–162. doi: 10.1016/j.psychneuen.2016.02.007.

Ridout, K. K. *et al.* (2016) 'Depression and telomere length: A meta-analysis', *Journal of Affective Disorders*, 191, pp. 237–247. doi: 10.1016/j.jad.2015.11.052.

Schutte, N. S. and Malouff, J. M. (2015) 'The association between depression and leukocyte telomere length: a meta-analysis', *Depression and Anxiety*, 32, pp. 229–238. doi: 10.1002/da.22351.

Shalev, I. *et al.* (2013) 'Exposure to violence during childhood is associated with telomere erosion from 5 to 10 years of age: a longitudinal study.', *Molecular psychiatry*. Nature Publishing Group, 18, pp. 576–81. doi: 10.1038/mp.2012.32.

Shin, Y. A. *et al.* (2008) 'Exercise training improves the antioxidant enzyme activity with no changes of telomere length', *Mechanisms of Ageing and Development*, 129(5), pp. 254–260. doi: 10.1016/j.mad.2008.01.001.

Shlush, L. I. *et al.* (2011) 'Telomere elongation followed by telomere length reduction, in leukocytes from divers exposed to intense oxidative stress - Implications for tissue and organismal aging', *Mechanisms of Ageing and Development*. Elsevier Ireland Ltd, 132(3), pp. 123–130. doi: 10.1016/j.mad.2011.01.005.

Steenstrup, T. *et al.* (2013) 'The telomere lengthening conundrum - Artifact or biology?', *Nucleic Acids Research*, 41(13), pp. 1–7. doi: 10.1093/nar/gkt370.

Toupance, S. *et al.* (2017) 'Short telomeres, but not telomere attrition rates, are associated with carotid atherosclerosis', *Hypertension*, 70, pp. 420–426. doi: 10.1161/short.

Verhoeven, J. E. *et al.* (2014) 'Cellular aging in depression: Permanent imprint or reversible process?: An overview of the current evidence, mechanistic pathways, and targets for interventions Prospects & Overviews J. E. Verhoeven et al.', *BioEssays*, pp. 968–978. doi: 10.1002/bies.201400068.

Verhoeven, J. E. *et al.* (2015) 'Anxiety disorders and accelerated cellular ageing', *The British Journal of Psychiatry*, 206, pp. 371–378. doi: 10.1192/bjp.bp.114.151027.

Verhulst, S. *et al.* (2013) 'Do leukocyte telomere length dynamics depend on baseline telomere length? An analysis that corrects for "regression to the mean".', *European Journal of Epidemiology*, 28, pp. 859–66. doi: 10.1007/s10654-013-9845-4.

Vickers, A. J. and Altman, D. G. (2001) 'Statistics Notes: Analysing controlled trials with baseline and follow up measurements', *Bmj*, 323(7321), pp. 1123–1124. doi: 10.1136/bmj.323.7321.1123.

Weischer, M., Bojesen, S. E. and Nordestgaard, B. G. (2014) 'Telomere shortening unrelated to smoking, body weight, physical activity, and alcohol intake: 4,576 general population individuals with repeat measurements 10 years apart', *PLoS Genetics*, 10(3). doi: 10.1371/journal.pgen.1004191.