Supplementary Material

Assessing the effect of measurement error on studies of telomere length

The correlation (r) between the error-prone Luminex method and the Southern blot method ranged from 0.65 to 0.75 (**Figure 2**). However, the correlation between the Luminex measure and the "true" average telomere content of a DNA sample will be higher than the values we observe, because the Southern blot method is not free of error. We conducted simulations to assess the effect of measurement error on estimation of associations between Luminex-based TL measurements and disease phenotypes. We considered a range of values for the correlation between the Luminex and the true value: 0.6, 0.7, 0.8, and 0.9.

Simulation methods

For each simulated scenario, we generated 1,000 datasets consisting of 5,000 observations and three variables: the true value of telomere length (X), an error-prone measurement of telomere length (X*, representing the Luminex-based measure), and a continuous outcome (Y) influenced by X. X was a randomly generated standard normal variable. Y was modeled as a random number from a standard normal distribution plus a linear effect of X:

$$y_i = \beta_{xy} x_i + \rho_i \text{ with } \rho_i \sim N(0,1). \quad (1)$$

β_{xy} was set to either 0.0, 0.025, 0.05, 0.1, or 0.2. The error-prone measure of X, X*, was generated by adding a normally-distributed error component to X, as follows:

$$x_i^* = x_i + \tau_i \text{ with } \tau_i \sim N(0, \delta_x^*). \quad (2)$$

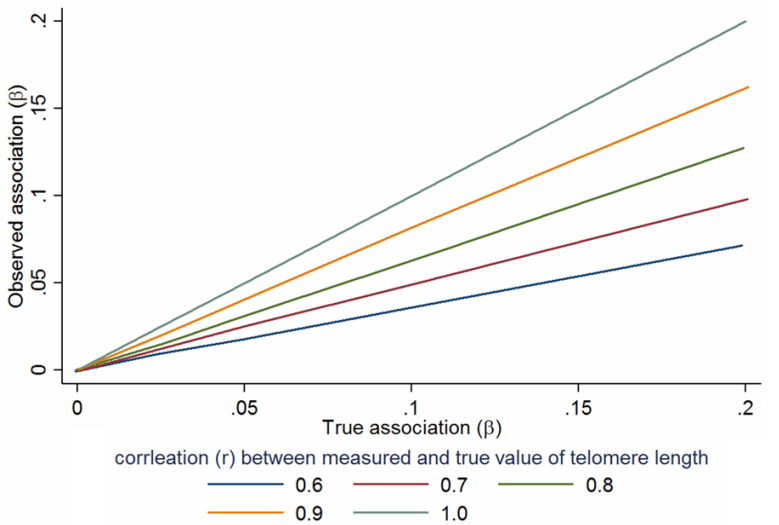
δ_{x^*} was chosen to produce a specific R^2 values corresponding to the square of the r values listed above (0.6, 0.7, 0.8, and 0.9) for the regression of X* on X, using the following equation:

$$R^2_{xx^*} = \frac{Var(\beta_{xx}^* X)}{Var(\beta_{xx}^* X) + Var(\tau)}. \quad (3)$$

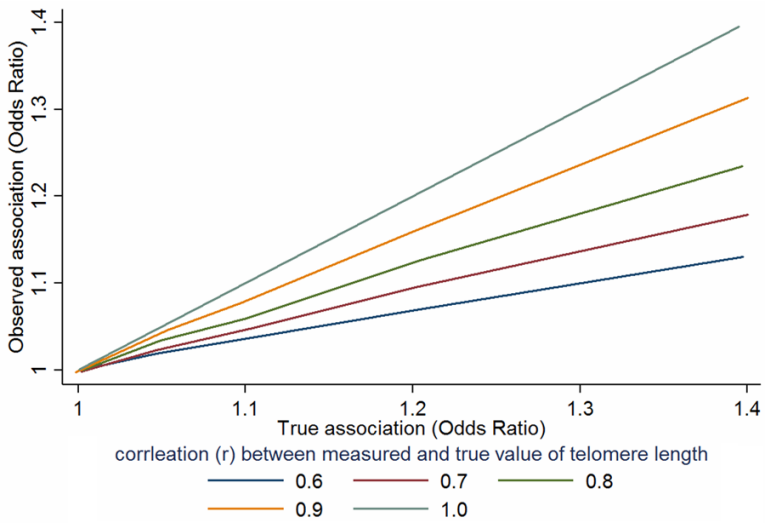
To examine the effect of measurement error for binary outcomes, data on X and X* were generated as above, but Y was generated as a binary outcome using a logistic model:

$$P(Y_i = 1/X_i) = 1/(1 + e^{-(\beta_0 + \beta_{xy} x_i)}). \quad (4)$$

β_{xy} was chosen to produce specific odds ratios for the true effect of X on Y (OR=1.0, 1.05, 1.1, 1.2, and 1.4), and β_0 was chosen to produce an average population risk of 0.10. In each simulated dataset, the association between X* and Y was estimated using linear regression (for continuous outcomes) and logistic regression (for binary outcomes). The median beta coefficients are ORs from these regressions are displayed in Supplementary Figures 1 and 2.



Supplementary Figure 1. The median estimate for the association between telomere length and a continuous outcome in the presence of measurement error for telomere length (based on 1,000 simulations).



Supplementary Figure 2. The median estimate for the association between telomere length and a binary outcome in the presence of measurement error for telomere length (based on 1,000 simulations).

Supplementary Table 1. Precision of the Luminex assay and the Southern Blot method for measuring telomere length/content

	Per-sample inter-assay CVs (based on duplicate samples)			
	Geometric mean	Arithmetic mean	Median	Overall/Pooled CV
Luminex assay	5.45	7.60	7.13	9.1
Southern Blot	0.97	1.36	1.31	1.7

CV, coefficient of variation (%).