23. Cunningham JM, Johnson RA, Litzelman K *et al*. Telomere length varies by DNA extraction method: implications for epidemiologic research. *Cancer Epidemiol Biomarkers Prev* 2013;**22**: 2047-54.
24. Hofmann JN, Hutchinson AA, Cawthon R *et al*. Telomere length varies by DNA extraction method: implications for epidemiologic research - letter. *Cancer Epidemiol Biomarkers Prev* 2014;**23**:1129-30.
25. Denham J, Marques F, Charchar F. Leukocyte telomere length variation due to DNA extraction method. *BMC Res Notes* 2014;**7**:877.
26. Tolios A, Teupser D, Holdt LM. Preanalytical conditions and DNA isolation methods affect telomere length quantification in whole blood. *PLoS One* 2015;**10**:e0143889.
27. Eisenberg DTA, Salpea KD, Kuzawa CW, Hayes MG, Humphries SE; on behalf of the European Atherosclerosis Research Study I. I. Group. Substantial variation in qPCR measured mean blood telomere lengths in young men from eleven European countries. *Am J Hum Biol* 2011;**23**:228-31.
28. Gardner M, Bann D, Wiley L *et al*. Gender and telomere length: systematic review and meta-analysis. *Exp Gerontol* 2014;**51**: 15–27.
29. Eisenberg DT. Inconsistent inheritance of telomere length (TL): is offspring TL more strongly correlated with maternal or paternal TL? *Eur J Hum Genet* 2014;**22**:8-9.
30. Haycock PC, Heydon EE, Kaptoge S, Butterworth AS, Thompson A, Willeit P. Leucocyte telomere length and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ* 2014;**349**:g4227.
31. Ridout KK, Ridout SJ, Price LH, Sen S, Tyrka AR. Depression and telomere length: A meta-analysis. *J Affect Disord* 2016;**191**: 237-47.
32. Forero DA, González-Giraldo Y, López-Quintero C, Castro-Vega LJ, Barreto GE, Perry G. Telomere length in Parkinson's disease: A meta-analysis. *Exp Gerontol* 2016;**75**:53-55.
33. Lin P-Y, Huang Y-C, Hung C-F. Shortened telomere length in patients with depression: a meta-analytic study. *J Psychiatr Res* 2016;**76**:84-93.
34. 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* 2016;**54**:159-69.
35. Zhu X, Han W, Xue W *et al*. The association between telomere length and cancer risk in population studies. *Sci Rep* 2016;**6**: 22243.
36. Aviv A, Valdes AM, Spector TD. Human telomere biology: pitfalls of moving from the laboratory to epidemiology. *Int J Epidemiol* 2006;**35**:1424-29.
37. Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull* 1980;**87**:245.

Response to: Reliability and validity of telomere length measurements

International Journal of Epidemiology, 2016, 1298–1301
doi: 10.1093/ije/dyw194

From Simon Verhulst,[1] Ezra Susser,[2] Pam R Factor-Litvak,[3] Mirre Simons,[4] Athanase Benetos,[5] Troels Steenstrup,[6] Jeremy D Kark[7] and Abraham Aviv[8,*]

[1]Groningen Institute for Evolutionary Life Sciences University of Groningen, Groningen, The Netherland, [2]Imprints Center for Genetic and Environmental Lifecourse Studies, Columbia University Mailman School of Public Health, and New York State Psychiatric Institute, New York, NY, USA, [3]Department of Epidemiology, Columbia University Mailman School of Public Health New York, NY, USA, [4]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK, [5]Département de Médecine Gériatrique, CHU de Nancy, and INSERM, U1116, Vandoeuvre-les-Nancy, France, [6]Copenhagen, Denmark, [7]Hebrew University-Hadassah School of Public Health and Community Medicine, Jerusalem, Israel, and [8]Center of Human Development and Aging, State University of New Jersey, Newark, NJ, USA

*Corresponding author. Center of Human Development and Aging, Rutgers, State University of New Jersey, New Jersey Medical School, 185 South Orange Ave, Newark, NJ 07103, USA. E-mail: avivab@njms.rutgers.edu

In his letter[1] on our recent commentary,[2] Dan Eisenberg argues that Southern blot (SB) and quantitative polymerase chain reaction (qPCR)-based methods have similar reliability of leukocyte telomere length (LTL) measurement, a suggestion that runs counter to the literature going back more than a decade. This literature shows that the SB-based method better captures associations of LTL with a host of variables than does the qPCR-based method. For example, a recent meta-analysis[3] concluded that the influence of sex on LTL was apparent when using SB, but absent when using qPCR, and we see a difference in measurement reliability as the most parsimonious explanation for this finding. We consider this meta-analysis important because it compares how these two methods perform in actual epidemiological studies. Such a comparison cannot be derived from the results achieved in small samples designed to

assess reliability in laboratory settings (a point elaborated later). Likewise, with respect to LTL dynamics (LTL and its age-dependent change), lower reliability is also the most likely explanation we can see for the much larger proportions of individuals with counter-intuitive elongation of LTL as they age (i.e. at the second measurement time point) in qPCR-based than in SB- based studies.[4] Moreover, lower reliability primarily explains the difficulties in detecting the longer LTL in African Americans than in Whites of European ancestry when using the qPCR-based method.[5] Thus, the epidemiological telomere literature strongly suggests higher reliability of SB compared with qPCR.

That said, we welcome Dan Eisenberg's letter since it provides a forum to further address highly relevant issues related to LTL measurements by the SB-based method, which measures LTL in kilo-bases (kb),[6] and the qPCR-based method, which measures telomere DNA content as the ratio of telomeric PCR product to single copy gene PCR product (T/S).[7,8] Below we respond to Eisenberg's points on: (i) the coefficient of variation (CV) as a tool to compare measurement reliability between LTL data generated by SB versus by qPCR; and (ii) the need for external validation of LTL measured by the two methods. We also address other validity-related matters.

The coefficient of variation

In our commentary[2] we suggested that it is not appropriate to use the CV to measure reliability for qPCR, nor to use the CV to compare reliability of qPCR and SB. We pointed out that the heteroscedasticity assumption of the CV does not hold for qPCR measurements. This reason alone is sufficient to require an alternative method to quantify and compare measurement reliability, for which we proposed considering the intraclass correlation coefficient (ICC), a well-established method in other research fields.[9]

However, Eisenberg makes the additional compelling point that when using the CV to compare two methods, one assumes that both measures have the same 'natural zero points'. He goes on to argue that this assumption does not hold for SB, because the terminal restriction fragments (TRFs) include sub-telomeric regions, and we agree. In making this assertion about the dependence of CV on equal zero points, Eisenberg also makes incorrect assertions, and these are unfortunately conflated because they are illustrated together in his Table 1.[1] He suggests that the higher CVs reported for the qPCR than the SB are entirely due to the different measurement scales used. To equate the two measurement scales, he (i) converted T/S from a ratio to an absolute measure of LTL, and (ii) estimated the sub-telomeric length, which is included in the terminal restriction fragments that are measured by the SB. Both of these procedures are fraught with problems and we believe that his conversions are erroneous. Moreover, Eisenberg's assertion that qPCR-based TL measurements have a natural zero point may not generally be correct, due to the way it is calculated. We discuss these issues in the Supplementary material (available as Supplementary data at *IJE* online) as it would require us to digress from the main points by including details and complexities that many readers may not find germane.

External validation of LTL measured by the two methods

Validity of measurement depends upon measurement reliability, but reliable measurement does not ensure validity. Eisenberg suggests, therefore, a way to gauge the validity of LTL measures. He proposes that 'well established correlates of LTL can be utilized as markers of external validity'. We agree that when LTL is measured by multiple methods that differ only in random measurement error (reliability), one does generally expect stronger and more valid correlations for the more reliable method with 'well-established correlates'. Nevertheless, we do not think that his proposal, a comparison of reported correlations, should serve to assess either validity or reliability. First, Eisenberg's approach cannot be used to assess validity because both methods could be biased; an independent benchmark

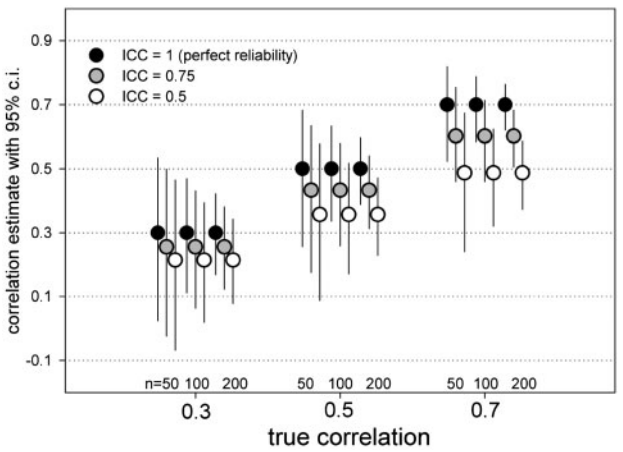

Figure 1. Estimated correlations r and 95% confidence intervals (Y axis) depending on the true value of the correlation being estimated (X axis) for different levels of measurement reliability (ICC values 0.5, 0.75 and 1). The estimates of r deviate more from the true value when the ICC is lower. Increasing sample size (50, 100 or 200) decreases the confidence interval of the estimate, but does not affect the average point estimate. Note that confidence intervals are large relative to the effect of the ICC differences on the r estimates, and hence comparing correlation coefficients will have low statistical power to detect effects of measurement reliability on r. Note further that correlations in epidemiological telomere studies will reach values as high as the upper limit (0.7) in this example only in exceptional circumstances.

criterion is needed to gauge validity. We do not yet have such a benchmark. Thus, one correlation may be stronger than another, but how much closer the best correlation is to the (unknown) benchmark value remains unknown. Second, Eisenberg's approach is in practice not informative with respect to measurement reliability. Typically, the sample sizes in studies that compared different methods to measure TL are too small to detect relevant differences in measurement reliability using this approach.

We illustrate this concept in Figure 1, where it can be seen that even very large differences in measurement reliability are difficult to detect with sample sizes of 50-200. Thus, the estimates that Eisenberg reports for TL associations with age for sample sizes of 50-190 (Table 2 in his letter)[1] may be viewed as random samples from the broad distributions in our Figure 1 (albeit they are not entirely independent because they are based on the same set of samples within each study). Differences in the distribution of age (frequently non-normal) will also skew correlations between studies, also arguing against the use of age as a comparative measure for validity between studies. Due to the limited sample sizes, possible differences in age distribution between studies and consequently large sampling variation, we attribute little value to these point estimates or Eisenberg's conclusion that there was on average no difference between SB- and qPCR-based studies.

Validity-related matters

Eisenberg notes that the qPCR assay is considerably influenced by DNA extraction techniques and sample handling. In our commentary,[2] based on theoretical considerations,[13] we also suggested that the integrity of the amplifiable sequences by the qPCR might be critical for obtaining accurate results. This point has recently been confirmed in a study showing that DNA integrity has a considerable effect on TL measurements by qPCR.[14] The requirement to test DNA integrity in each sample, which is a routine for SB, is typically not done in qPCR-based studies of LTL and reduces the high-throughput advantage of the qPCR-based method.

Finally, there is more to understanding the role of telomeres in health and disease than just knowing that telomeres are longer (or shorter) in one individual versus another. Consequently, many studies have transformed their qPCR data, expressed in T/S units, to kb, because the information in T/S data is limited to relative differences that cannot be compared between studies. The pitfalls of transforming the T/S to kb become apparent in Eisenberg's own work.[15] His transformation of T/S to kb produced an implausible range of LTL across European populations, i.e. from 5.1 kb in Naples to 18.6 kb in Ghent. In his letter, Eisenberg attributes this to the high inter-laboratory variation in generating T/S, which can be considerable according to the paper that was the subject of our commentary.[12] However, if that was the only reason, the range of variation in LTL resulting from transforming qPCR measures generated in a single laboratory to LTL data should be similar to measures generated by SB. This is hardly the case, since the range of transformed qPCR data[16] is considerably larger than the range generated by SB.[17,18] A likely explanation is that transforming qPCR to SB data is based on linear models, whereas the relation between data generated by qPCR and SB is often curvilinear (see Supplementary material, available as Supplementary data at *IJE* online).

Conclusions

The SB- and the qPCR-based methods of LTL measurements have been important partners in gaining insight into the role of telomeres in human health and disease. Together they have been instrumental, for instance, in deciphering the LTL-SNPs[19–23] that opened the door to remarkable advances in understanding the role of telomeres in a host of human diseases. The lower measurement reliability of the qPCR-based method renders it less suited to studies with small sample-sizes, but the high throughput of the method makes it a powerful tool in large epidemiological studies where increasing sample size can offset higher measurement error. Note, however, that lower measurement reliability and the attendant non-differential misclassification also attenuates effect size estimates, and the attenuation effect is independent of sample size (Figure 1). Thus, as we stated before, we see a role for both methods, depending on characteristics of the study design. However, in the light of the epidemiological literature, we consider attempts to show that the qPCR-based method is equal to the SB method in reliability to be misleading.

Eisenberg and we agree that the CV is a suboptimal indicator of measurement reliability, supporting the use, for example, of the ICC to evaluate measurement reliability in future telomere studies. Notably, the ICC can be calculated in different ways that provide different information,[9] and it is important therefore to report exactly how the ICC was calculated. Should inter-assay ICCs or other substitutes for CV become standard in telomere epidemiology, the data will be available to make a more quantitative assessment of the reliability of the different techniques as they are applied in the field.

Supplementary Data

Supplementary data are available at *IJE* online.

References

1. Eisenberg D. 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;**45**:1293–96.

2. Verhulst S, Susser E, Factor-Litvak PR *et al*. Commentary: The reliability of telomere length measurements. *Int J Epidemiol* 2015;**44**:1683–86.
3. Gardner M, Bann D, Wiley L *et al*. Gender and telomere length: systematic review and meta-analysis. *Exp Gerontol* 2014;**51**: 15–27.
4. Steenstrup T, Hjelmborg JV, Kark JD, Christensen K, Aviv A. The telomere lengthening conundrum - artifact or biology? *Nucleic Acids Res* 2013;**41**:e131.
5. Elbers CC, Garcia ME, Kimura M *et al*. Comparison between southern blots and qPCR analysis of leukocyte telomere length in the health ABC study. *J Gerontol A Biol Sci Med Sci* 2014;**69**:527–31.
6. Kimura M, Stone RC, Hunt SC *et al*. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat Protoc* 2010;**5**(9):1596–607.
7. Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 2009;**37**:e21.
8. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res* 2002;**30**:e47.
9. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–28.
10. 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**:e134.
11. Eisenberg DT, Kuzawa CW, Hayes MG. Improving qPCR telomere length assays: Controlling for well position effects increases statistical power. *Am J Hum Biol* 2015;**27:**570–75.
12. Martin-Ruiz CM, Baird D, Roger L *et al*. Reproducibility of telomere length assessment: an international collaborative study. *Int J Epidemiol* 2014;**44**:1673–83.
13. Brisco MJ, Latham S, Bartley PA, Morley AA. Incorporation of measurement of DNA integrity into qPCR assays. *Biotechniques* 2010;**49**:893-97.
14. Tolios A, Teupser D, Holdt LM. Preanalytical conditions and DNA isolation methods affect telomere length quantification in whole blood. *PLoS One* 2015;**10**:e0143889.
15. Eisenberg DT, Salpea KD, Kuzawa, CW, Hayes MG,, Humphries SE; European Atherosclerosis Research Study II Group. Substantial variation in qPCR measured mean blood telomere lengths in young men from eleven European countries. *Am J Hum Biol* 2011;**23**:228–31.
16. Njajou OT, Cawthon RM, Damcott CM *et al*. Telomere length is paternally inherited and is associated with parental lifespan. *Proc Natl Acad Sci U S A* 2007;**104:**12135–39.
17. Factor-Litvak P, Susser E, Kezios K *et al*. Leukocyte telomere length in newborns: implications for the role of telomeres in human disease. *Pediatrics* 2016, Mar 11. pii: peds.2015-3927. [Epub ahead of print.]
18. Hunt SC, Chen W, Gardner JP *et al*. Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. *Aging Cell* 2008;**7**:451–58.
19. Codd V, Mangino M, van der Harst P *et al*. Common variants near TERC are associated with mean telomere length. *Nat Genet* 2010;**42**:197–99.
20. Codd V, Nelson CP, Albrecht E *et al*. *Nat Genet* 2013; **45**:422–27.
21. Mangino M, Christiansen L, Stone R *et al*. DCAF4, a novel gene associated with leucocyte telomere length. *J Med Genet* 2015;**52:**157–62.
22. Levy D, Neuhausen SL, Hunt SC *et al*. Genome-wide association identifies OBFC1 as a locus involved in human leukocyte telomere biology. *Proc Natl Acad Sci U S A* 2010; **107**:9293-98.
23. Mangino M, Hwang SJ, Spector TD *et al*. *Hum Mol Genet* 2012;**21**:5385–94.