Int. J. Epidemiol. Advance Access published August 30, 2016

IEA

International Journal of Epidemiology, 2016, 1–4
doi: 10.1093/ije/dyw191
Letter to the Editor

Letter to the Editor

Telomere length measurement validity: the coefficient of variation is invalid and cannot be used to compare quantitative polymerase chain reaction and Southern blot telomere length measurement techniques

From Dan T.A. Eisenberg

Department of Anthropology, and Center for Studies in Demography and Ecology, University of Washington,Campus Box 353100, Seattle, WA 98195, USA. E-mail: dtae@dtae.net

The International Journal of Epidemiology continues to be an important forum for discussions of both methodological[1–7] and theoretical[8,9] issues in telomere epidemiology. The recent commentary by Verhulst and colleagues[2] helped me realize some critical issues that should be a part of this continuing dialogue.

Coefficient of variation

Verhulst and colleagues[2] note that the coefficient of variation (CV) may be a problematic measure of precision because it assumes that measurement error is proportional to the mean. There is, however, an even more fundamental problem with the CV measure. The CV statistic is an appropriate measure only for ratio variables, that is continuous variables with natural zero points (i.e. where zero clearly signifies having none of the thing being measured).[10] In theory, the zero point of quantitative polymerase chain reaction (qPCR)-measured telomere length (TL) represents the point of essentially no telomere sequence in a sample. For TL measured using Southern blot (SB), the zero point represents having no telomere and none of the sub-telomeric portions which are also measured by SB.[11] As shown below, these zero points between SB and qPCR measures do not correspond with each other and thereby create highly misleading estimates of measurement error. Further, within the same type of TL measure (SB or qPCR), the zero points often do not correspond across different implementations of these measurement techniques in different laboratories. When these zero points do not correspond, comparisons of CV are invalid.

To illustrate the problem of calculating CVs for data where measures do not correspond in their zero points, consider temperature measured three times with a thermometer which has both degrees Celsius (°C) and degrees Fahrenheit (°F) scales on it (Table 1, columns 1 and 2). In this case, for °C the CV is 10%, but for °F the CV is 3.6%. This is because the zero points of both °C and °F are arbitrary and do not correspond with each other. As shown below, the same problem of differing zero points is evident in measuring TL.

The problem of comparing CVs between qPCR- and SB-measured TL can be shown with an example similar to the temperature illustration above. This can be done by using the results of a previous regression analysis of the same DNA samples measured with qPCR and SB.[12] This regression analysis with SB TL as the dependent variable (Y) and qPCR TL as the independent variable (X) yielded a β estimate of 3.33 kilobase pairs (kbp) and y-intercept of 3.73 kbp. This y-intercept value implies that when qPCR TL is at zero, SB TL would be measured as 3.73 kbp. Similarly large y-intercept values are found in other

Table 1. Dependence of CV values on zero points for temperature (columns 1 & 2) and telomere length measurement (columns 3 & 4)

	1 °C	2 °F[1]	3 qPCR	4 SB[2]
Measure 1	9.0	48.2	0.92	6.79
Measure 2	10.0	50.0	1.02	7.13
Measure 3	11.0	51.8	1.04	7.19
CV	10.00	3.60	6.47	3.04

[1]F calculated using equation °C*9/5 + 32;
[2]SB calculated by using equation from Cawthon 2009: qPCR*3.33 + 3.73

© The Author 2016; all rights reserved. Published by Oxford University Press on behalf of the International Epidemiological Association

Downloaded from http://ije.oxfordjournals.org/ by guest on August 31, 2016

comparisons between qPCR and SB measures of the same samples.[13–15] These y-intercept values might reflect the size of the subtelomeric portions which are measured by SB, but not qPCR (although extreme caution is warranted in assuming this since no samples actually have zero telomere lengths and these y-intercept values are thus interpolated well beyond where data exist).

Using the above β and y-intercept estimates, qPCR measures can be linearly transformed into interpolated SB measures (Table 1, columns 3 and 4). I note that the specific conversion factor obtained from the particular set of samples and laboratory protocols referenced here should not be used to convert qPCR measures to SB more broadly.[16–18] By calculating CV values for the qPCR- and SB-interpolated TL measures, the results are values of 6.47% and 2.30%, respectively. Comparisons of the inter-assay CVs of qPCR and SB measures in recent studies show size differences similar to these (6.45% and 1.74%, respectively, in one study[14] and 5.8% and 1.5%, respectively, in another[19]). Comparisons using conversion factors from other analyses examining qPCR- and SB-measured TL in the same samples yield similar results: a qPCR CV of 6.47% being equivalent to SB CVs of 2.30% and 2.03%.[13,15]

CV statistics have frequently been used to compare the precision of different TL measurement techniques—usually with the suggestion that these show much greater precision for SB than qPCR techniques.[2,11,17,18,20,21] As illustrated above, concluding this based on CV values makes as much sense as saying that the Fahrenheit scale has lower measurement error than the Celsius scale. This means that the statistical power estimates given by Verhulst and colleagues[2] cannot be used to compare power between qPCR and SB methods if raw CV values are used to estimate measurement error. This problem of differing y-intercept values is also evident when comparing one set of qPCR measurements with another and one set of SB with other SB measures. Regression models from the qPCR data from laboratories 5 through 9 in Martin-Ruiz *et al.*[1] show substantial deviations from null expectations of a y-intercept of zero (*P*- values for 10 different permutations of comparisons between laboratories of: 0.001, 0.001, 0.003, 0.014, 0.027, 0.129, 0.223, 0.283, 0.316 and 0.392). Similarly, SB methods using different restriction enzymes result in differing y-intercepts,[11] and variation in subtelomeric length across samples, populations and/or species may also influence the estimated y-intercept value.

Alternatives to assessing TL measurement quality

How can the field of telomere biology better quantify measurement quality? Verhulst and colleagues[2] suggest, using the intra-class correlation coefficient (ICC) instead of the CV. The ICC is not susceptible to the same set of problems pointed out above. However, the ICC quantifies the internal replicability (precision) of measures, but does not tell us about how well each method is measuring the latent construct of telomere length (i.e. accuracy). External validity of TL measures is an approach which should be focused on more to help show measurement accuracy. To accomplish this, I suggest that well-established correlates of TL can be used as markers of external validity.[15] For example, in humans TL is highly heritable and declines with age. Random measurement error attenuates correlations, so if in the same samples putative measures of TL show weaker correlations with family relatedness or with age, than it can be inferred that these measures are likely worse. One possible problem with using heritability as a measure of external validity is that SB measures subtelomeric portions; to the extent that these subtelomeric portions vary by relatedness, this may inflate or deflate TL heritability estimates from SB measures.

We are aware of two previous comparisons that have used age as a measure of external validity of qPCR versus SB in the same set of samples.[14,15] This approach is applied again here using data from the methods paper which debuted the multiplex qPCR assay[12] together with subject age data (both generously shared by Richard Cawthon). These new and past data are found in Table 2 and show little difference in correlations with age between SB and qPCR measures ($P = 0.84$), suggesting that there is no discernable difference in external validity between these qPCR and SB measures.

Moving forward

The CV is not a valid precision statistic of TL measurements and our above calculations suggest previously observed differences in CV between SB and qPCR may be more representative of improper statistics than actual differences in precision. In the absence of appropriate measures, we cannot say that the precision or accuracy of the qPCR technique is better or worse than SB. The qPCR technique has the clear benefits of requiring far less DNA

Table 2. External validity of qPCR and SB measures by examining correlations of each with age

qPCR-age (r)	SB-age (r)	*P* of diff[1]	*N*	ref
−0.42	−0.54	0.07	50	14
−0.51	−0.46	0.46	190	15
−0.70	−0.71	0.80	95	12
−0.56	−0.55	0.84	335	combined[2]

[1]*P* value calculated difference in considering correlation between qPCR and SB measures (Steiger, 1980)[37];

[2]Fisher's r to Z-transformation was used to combine weighted r-values.

Downloaded from http://ije.oxfordjournals.org/ by guest on August 31, 2016

and being substantially cheaper than the SB technique. However, the qPCR assay is deceptively difficult to implement, and reasons for concern remain about poor quality uses of the qPCR assay in the literature. For example, the qPCR assay can be substantially influenced by DNA extraction technique and sample handling.[22–26] A previous cross-population analysis of TL using qPCR (which I was first author of) showed dramatic population variation[27] which probably reflects similar analytical artefacts rather than actual population differences. qPCR techniques do not seem to allow reliable comparison of TL across cohorts or populations unless samples are collected and handled identically. Recent meta-analyses of TL associations tend to reveal high levels of heterogeneity[28–35]—perhaps due to frequently poor quality of qPCR data.

Abraham Aviv, Simon Verhulst and colleagues[2,36] have played an important role in stressing that the relative statistical power to detect associations using different TL measurement techniques should be carefully considered. Underpowered studies are more likely to result in false results, to waste resources and to create confusion in the literature. I look forward to seeing power analyses conducted based on valid measurements of precision and accuracy. However, power analyses are not enough. Power analyses should be coupled with cost-benefit analyses that include considerations of financial costs. That is even if SB is shown to be a more precise technique than qPCR, the much higher relative costs of SB may mean that, in cases where many samples are readily available, qPCR is still a more cost-effective technique to glean the same information and level of statistical power. Measurement precision can also be improved by increasing the number of replicates run per sample—something which is much cheaper to do with qPCR than SB. Owing to the complexities of implementation of TL measurement techniques, validation of a technique in one laboratory does not mean that measurements are reliable in another laboratory trying to deploy the technique. I join Abraham Aviv, Simon Verhulst and colleagues in calling for 'impartial and rigorous comparisons in large-scale epidemiological studies between the qPCR and Southern blot methods'.[2,7]

Acknowledgements

This letter was much improved by discussions and feedback from Richard Cawthon, Rosana Risques, Simon Verhulst, Abraham Aviv, Hilary Bethancourt and Rebecca Ortega. Richard Cawthon generously shared unpublished data which was used in Table 2.

References

1. Martin-Ruiz CM, Baird D, Roger L *et al*. Reproducibility of telomere length assessment: an international collaborative study. *Int J Epidemiol* 2015;**44:**1673-834.
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. Martin-Ruiz CM, Baird D, Roger L *et al*. Is Southern blotting necessary to measure telomere length reproducibly? Authors' Response to: Commentary: The reliability of telomere length measurements. *Int J Epidemiol* 2015;**44**:1686-87.
4. Martin-Ruiz CM, Baird D, Roger L *et al*. Reproducibility of telomere length assessment: Authors' Response to Damjan Krstajic and LjubomirButurovic. *Int J Epidemiol* 2015;**44:**1739-41.
5. Martin-Ruiz CM, Baird D, Roger L *et al*. Reproducibility of telomere length assessment - an international collaborative study. *Int J Epidemiol* 2015;**44:**1749-54.
6. Krstajic D, Buturovic L. Reproducibility of telomere length assessment. *Int J Epidemiol* 2015;**44:**1738-39.
7. Aviv A. Commentary: Raising the bar on telomere epidemiology. *Int J Epidemiol* 2009;**38:**1735-36.
8. Eisenberg DT, Kuzawa CW. Commentary: The evolutionary biology of the paternal age effect on telomere length. *Int J Epidemiol* 2013;**42**:462-65.
9. Aviv A, Susser E. Leukocyte telomere length and the father's age enigma: implications for population health and for life course. *Int J Epidemiol* 2013;**42**:457-62.
10. Dunn OJ, Clark V. *Basic Statistics : A Primer for the Biomedical Sciences*. 4th edn. Hoboken, NJ: Wiley, 2009.
11. Kimura M, Stone RC, Hunt SC *et al*. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat Protocols* 2010;**5**:1596-607.
12. Cawthon RM. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 2009;**37**:e21.
13. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res* 2002;**30**:e47.
14. 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.
15. 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.
16. Horn T, Robertson BC, Gemmell NJ. The use of telomere length in ecology and evolutionary biology. *Heredity* 2010;**105**:497-506.
17. Aubert G, Hills M, Lansdorp PM. Telomere length measurement - caveats and a critical assessment of the available technologies and tools. *Mutat Res* 2012;**730**:59-67.
18. Aviv A. The epidemiology of human telomeres: faults and promises. *J Gerontol A Biol Sci Med Sci* 2008;**63**:979-83.
19. 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* 2013;**69**:527-31.
20. Montpetit AJ, Alhareeri AA, Montpetit M *et al*. Telomere length: a review of methods for measurement. *Nurs Res* 2014;**63**:289-99.
21. Sanders JL, Newman AB. Telomere length in epidemiology: a biomarker of aging, age-related disease, both, or neither? *Epidemiol Rev* 2013;**35**:112-31.
22. Boardman LA, Skinner HG, Litzelman K. Telomere length varies by DNA extraction method: implications for epidemiologic research—response. *Cancer Epidemiol Biomarkers Prev* 2014;**23**:1131.

Downloaded from http://ije.oxfordjournals.org/ by guest on August 31, 2016

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.

Downloaded from http://ije.oxfordjournals.org/ by guest on August 31, 2016