Full title: Method comparison studies of telomere length measurement using qPCR approaches: a critical appraisal of the literature

Short title: A review of telomere length measurement method comparisons using qPCR

Alyssa R. Lindrose[1]*, Lauren W. Y. McLester-Davis[2], Renee I. Tristano[3], Leila Kataria[4], Shahinaz M. Gadalla[5], Dan T. A. Eisenberg[6], Simon Verhulst[7], Stacy Drury[1,2,8]*

[1] Department of Psychiatry and Behavioral Sciences, School of Medicine, Tulane University, New Orleans, LA, USA

[2] Tulane Brain Institute, Tulane University, New Orleans, LA, USA

[3] School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA

[4] School of Science and Engineering, Tulane University, New Orleans, LA, USA

[5] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

[6] Department of Anthropology, Department of Biology, Center for Studies in Demography and Ecology, University of Washington, Seattle, WA, USA

[7] Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, the Netherlands

[8] Department of Pediatrics, School of Medicine, Tulane University, New Orleans, LA, USA

*Corresponding authors

Email: alindrose@tulane.edu (ARL); sdrury@tulane.edu (SD)

Abstract

Use of telomere length (TL) as a biomarker for various environmental exposures and diseases has increased in recent years. Various methods have been developed to measure telomere length. PCR-based methods remain wide-spread for population-based studies due to the high-throughput capability. While several studies have evaluated TL measurement methods, the results have been variable. We conducted a literature review of TL measurement cross-method comparison studies that included a PCR-based method published between January 1, 2002 and May 25, 2020. A total of 25 articles were found that matched the inclusion criteria. Papers were reviewed for quality of methodologic reporting of sample and DNA quality, PCR assay characteristics, sample blinding, and analytic approaches to determine final TL. Overall, methodologic reporting was low as assessed by two different reporting guidelines for qPCR-based TL measurement. There was a wide range in the reported correlation between methods (as assessed by Pearson's r) and few studies utilized the recommended intra-class correlation coefficient (ICC) for assessment of assay repeatability and methodologic comparisons. The sample size for nearly all studies was less than 100, raising concerns about statistical power. Overall, this review found that the current literature on the relation between TL measurement methods is lacking in validity and scientific rigor. In light of these findings, we present reporting guidelines for PCR-based TL measurement methods and results of analyses of the effect of assay repeatability (ICC) on statistical power of cross-sectional and longitudinal studies. Additional cross-laboratory studies with rigorous methodologic and statistical reporting, adequate sample size, and blinding are essential to accurately determine assay repeatability and replicability as well as the relation between TL measurement methods.

22 # Introduction:

23 Telomeres, the protective nucleic acid and protein cap found at the end of all eukaryotic
24 chromosomes, have captured the attention of scientists, medical and public health professionals,
25 biotechnology companies, and the media over the last two decades. In 1973, Olovnikov proposed his
26 theory of marginotomy, which reasoned that during DNA replication, DNA polymerase would not be
27 able to completely copy the first DNA segment and, to prevent the loss of critical DNA sequences in
28 genes, a noncoding set of DNA nucleotides would be required to act as a buffer protecting the loss of
29 important, gene-encoding, sequences [1]. Subsequently in 1978, Blackburn et al. first reported the
30 actual DNA sequences of telomeres in yeast [2], followed by the first sequencing of the human telomere
31 in 1988 [3]. The sequencing of telomeric DNA paved the way for the development of methods that
32 measured the length of telomeres, beginning with the first report of telomere length measurement
33 using the Southern blot method for mammalian chromosomes in 1988 [3]. Since then, thousands of
34 papers assessing telomere length (TL) in human cells have been published across a myriad of different
35 scientific fields (Fig 1). As a result of the broad scientific interest in both the role of TL in disease
36 processes and the influence of environmental factors on TL dynamics, the number of studies evaluating
37 TL in human cells continues to increase, in part facilitated by the regular development of new methods
38 and modifications of existing assays.

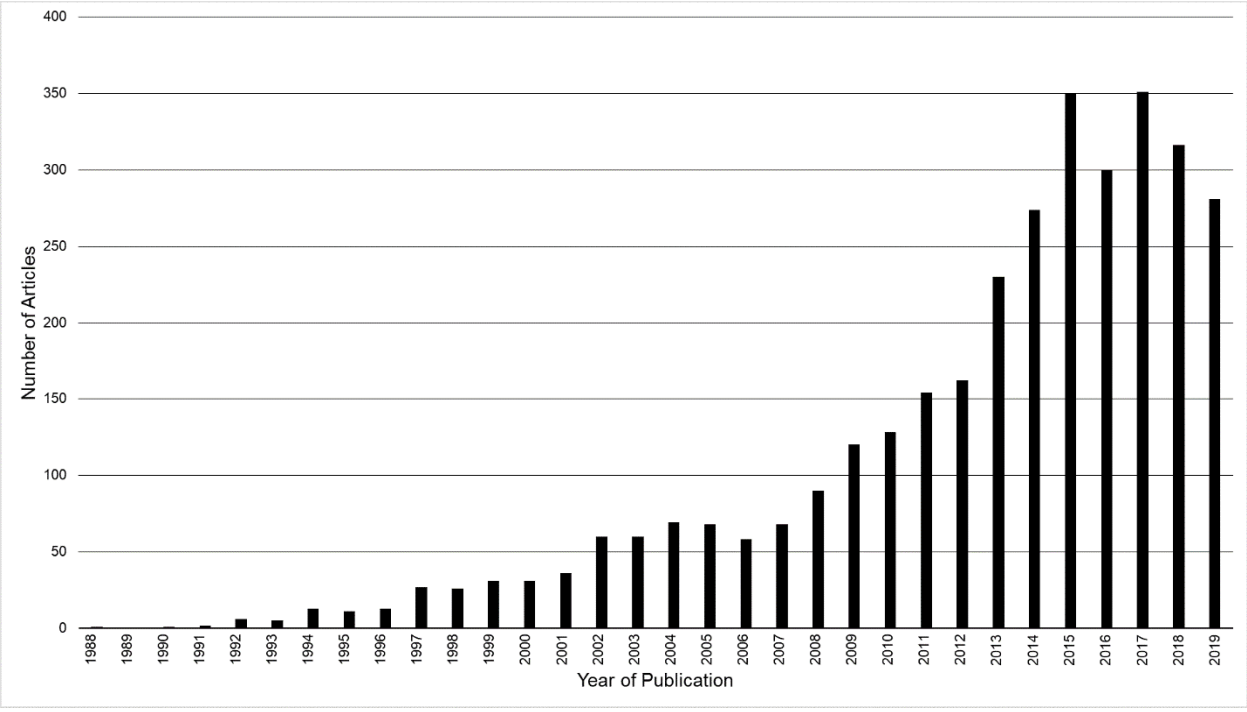
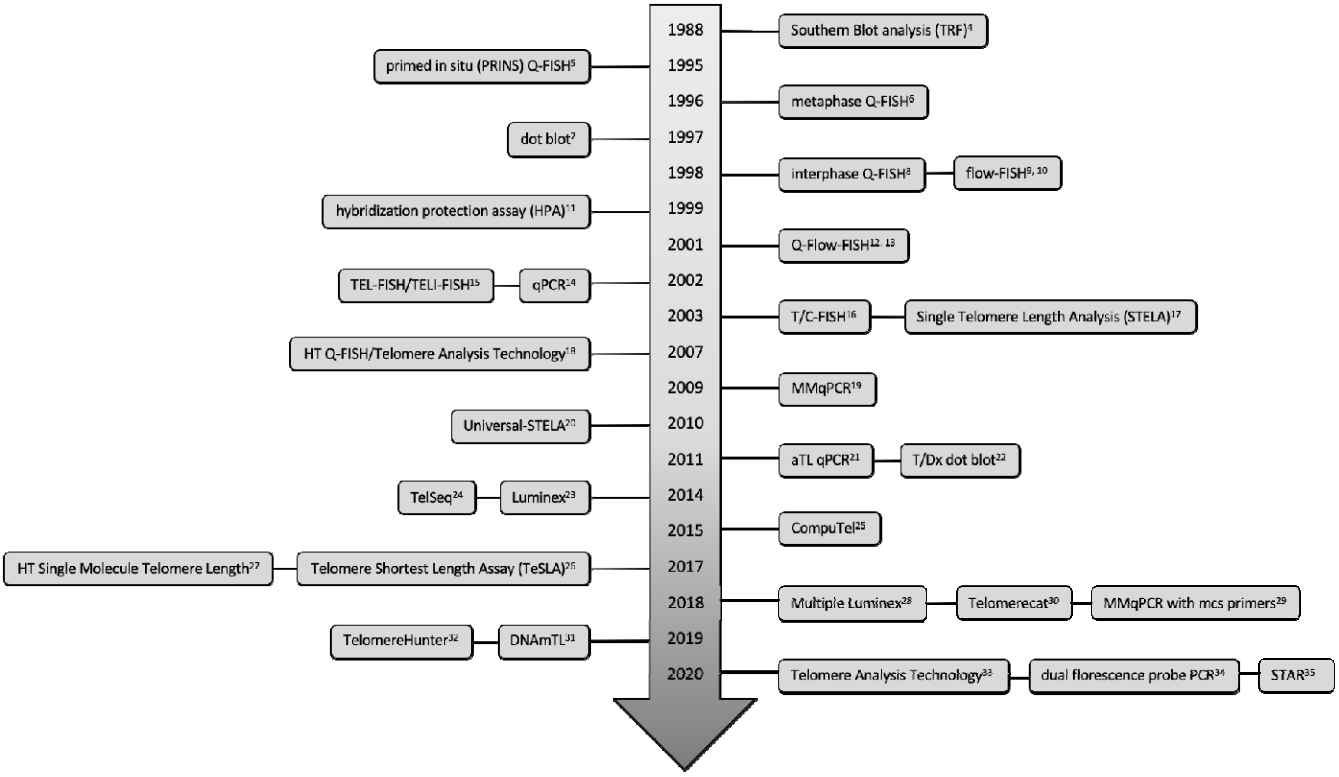


Fig 1. Telomere publications 1988-2019. All publications mentioning telomere length of human DNA from 1988 to 2019 obtained by searching "telomere," "length," and "human" in PubMed. Search was completed on May 25, 2020.

Currently, over two dozen assays have been developed to measure TL (Fig 2) [4-34]. These assays are classifiable into four broad categories: hybridization-based, PCR-based, sequence-based, and mixed methods (e.g. hybridization/PCR combination). These assays vary in the information they yield on TL. While most focus on the measurement of the average TL within the sample, assays also measure chromosome-specific TL [5, 16], the complete distribution of TL in a cell population [32], or the shortest TL [25]. The shortest TL has received considerable attention, given evidence from in vitro and preclinical models suggesting that the shortest TL is most predictive of cellular senescence [35, 36]. Several recent reviews have discussed the overall advantages and disadvantages of each method [37-39] focusing on cost, scalability, constraints of starting biological samples (e.g. living cells, amount of DNA, etc) and, to some extent, inter and intra-lab precision as specific challenges facing the field, including the use of

53 coefficient of variation (CV) compared to intraclass correlation coefficients (ICC) [40]. While studies of
54 basic telomere biology continue to explore the complex role that telomeres play in cellular and
55 organismal function, studies testing TL as markers of disease risk or environmental exposure must
56 balance biological relevance, methodologic precision, and experimental practicality, similar to other
57 epigenetic markers such as DNA methylation [41].



58

59 **Fig 2. Novel telomere methods developed from 1988 to 2020.** Data were obtained by literature search
60 and through references of telomere review papers and method comparison papers.

61 Over the last decade, debates have arisen over the utility and measurement of TL, particularly
62 with regards to qPCR-based methods. This debate is partially fueled by concerns related to the
63 reproducibility and replicability of TL measurements across studies, methods, and laboratories, and is
64 accentuated by new method development and adaptations of existing protocols without sufficient
65 consensus on the required quality control as more laboratories begin to perform TL assays

independently. In response to this debate, several studies have attempted to compare TL measured across different assay methods or laboratories. Some of these method comparison studies examined the direct correlation of TL measurement in the same sample using different assay methods and/or tested the repeatability of TL with the same method (e.g. the amount of within assay variation) [42, 43]. Others tested the relative correlation of the TL measured by different assays with an expected phenotype (e.g. aging, parent-offspring correlation) [44], or examined the relative ability of different assays of TL to predict a specific disease or health outcome [45, 46]. Each of these approaches requires a different analytic strategy and study design and comparison of outcomes is not straightforward. To date, the existing evidence remains insufficient to answer key methodologic questions related to differences in reproducibility and replicability across measurement assays and laboratories, and how/whether these differences affect the ability of TL to serve as a biological indicator of exposure or a predictor of disease or health risk [47]. Beyond these concerns, there remains a lack of consensus as to which, if any, methodology is the "gold standard," as even the classic Southern blot method is challenged by its inability to capture potentially critical metrics (e.g. full distribution, shortest telomere length, inclusion of the subtelomeric region). To ensure reliability in the widespread utilization of TL as a biomarker of environmental exposure and/or a predictor of a disease, measured by any method that is applicable to population studies, it is critical to systematically test fundamental issues related to assay reproducibility and replicability [48].

As part of a joint National Institute of Aging and National Institute of Environmental Health Science initiative that funded a U24 cooperative award and four separate U01 awards, a Telomere Research Network (TRN) was established in 2019 (trn.tulane.edu). The TRN is coordinating cross-method comparison studies with the long-term goal of developing methodological guidelines and recommendations for telomere research applicable to population-based studies. As a first step towards the goals of this network, we undertook a systematic literature review of published studies that directly

90 compared TL measured using at least one PCR-based method and another approach to determine how
91 these studies might inform the field, with particular attention to assay precision and accuracy of
92 different measurement assays and what research gaps remain.

93 As defined by the Committee on Reproducibility and Replicability in Science, precision is the
94 closeness of agreement between measured quantities obtained by replicate measurements, while
95 accuracy is the closeness of agreement between a measured quantity and a true value [48].
96 Reproducibility is defined as precision in measurement under conditions that involve different locations
97 or different measurement procedures, while repeatability is defined as precision in measurements that
98 include the same procedures/locations. Beginning from this perspective, this systematic review
99 evaluated the existing literature related to cross-method comparisons. This review focuses on PCR-
100 based methodologies due to their increasing use in population-based studies, their central role in the
101 debate related to assay precision, and the existence of two reporting guidelines - one created through
102 the TRN (Supplemental Table 1), and a second one created by a separate group in a recently published
103 manuscript [49]. The majority of PCR-based methods are derived from two seminal methodologic
104 papers by Richard Cawthon, the first describing a monoplex based assay (qPCR) and the second
105 describing a multiplex assay (MMqPCR) [13, 18]. Our review focuses specifically on the
106 comprehensiveness of methodologic reporting, correlation between TL measured by different assays,
107 assay repeatability and reproducibility, and overall scientific design of methodological comparisons.
108 Finally, we suggest areas of needed scientific examination and provide some guidance related to study
109 design, necessary sample size, and analytic approach, to address key remaining questions: (1) What is
110 known about the relationship between TL measured using PCR-based methods and other assays? (2)
111 What is known about the reproducibility and repeatability of PCR-based methods and how does this
112 relate to other TL measurement techniques? (3) What are the implications of methodologic precision for
113 sample size and power? (4) What are appropriate guidelines to systematically evaluate the precision of

both existing and future TL assays? Addressing these important questions is a requisite step in advancing our understanding of the ability of TL, measured by any approach, to serve as a sentinel of psychosocial and environmental exposures and a predictor of future disease.

Methods

Manuscript search

To identify relevant papers that reported on cross method comparisons of any qPCR-based method (qPCR, absolute TL (aTL), and MMqPCR) and another method of TL assessment (PCR-based or otherwise) or the same PCR-based method conducted in separate laboratories, we conducted a critical review beginning with a literature search (Fig 3). The following key terms "telomere," "telomere length," and "human" were searched in PubMed and Web of Science. From these initial results, a second search included the keyword "PCR" to identify the initial titles for screening. Search criteria included papers published since January 1, 2002 (the year the first method to measure TL by qPCR was published) through May 10, 2020. The references of selected papers were also reviewed to identify any additional papers. A list of identified papers was presented to the TRN Steering Committee, who also suggested additional papers. Initial review of papers for inclusion was accomplished through evaluation of both the abstract and methods section, as some manuscripts were not directly focused on methodological comparison and instead only reported the cross-method comparison on a subset of samples. Final inclusion in this review met the following baseline requirements:

1. Article published in a peer-reviewed journal (abstracts and pre-prints not included).
2. Article was not presenting the initial development of a new method or a substantial refinement of an existing methodology. This type of study was excluded due to the intrinsic bias in reporting novel method development.

136 3. Included a direct comparison of TL using the same biological sample measured with two distinct
137 TL assays or the same assay in two or more separate laboratories.
138 4. At least one of the methods used to measure TL was solely qPCR-based. TeSLA and STELA were
139 not considered due to the additional hybridization component of the assay.

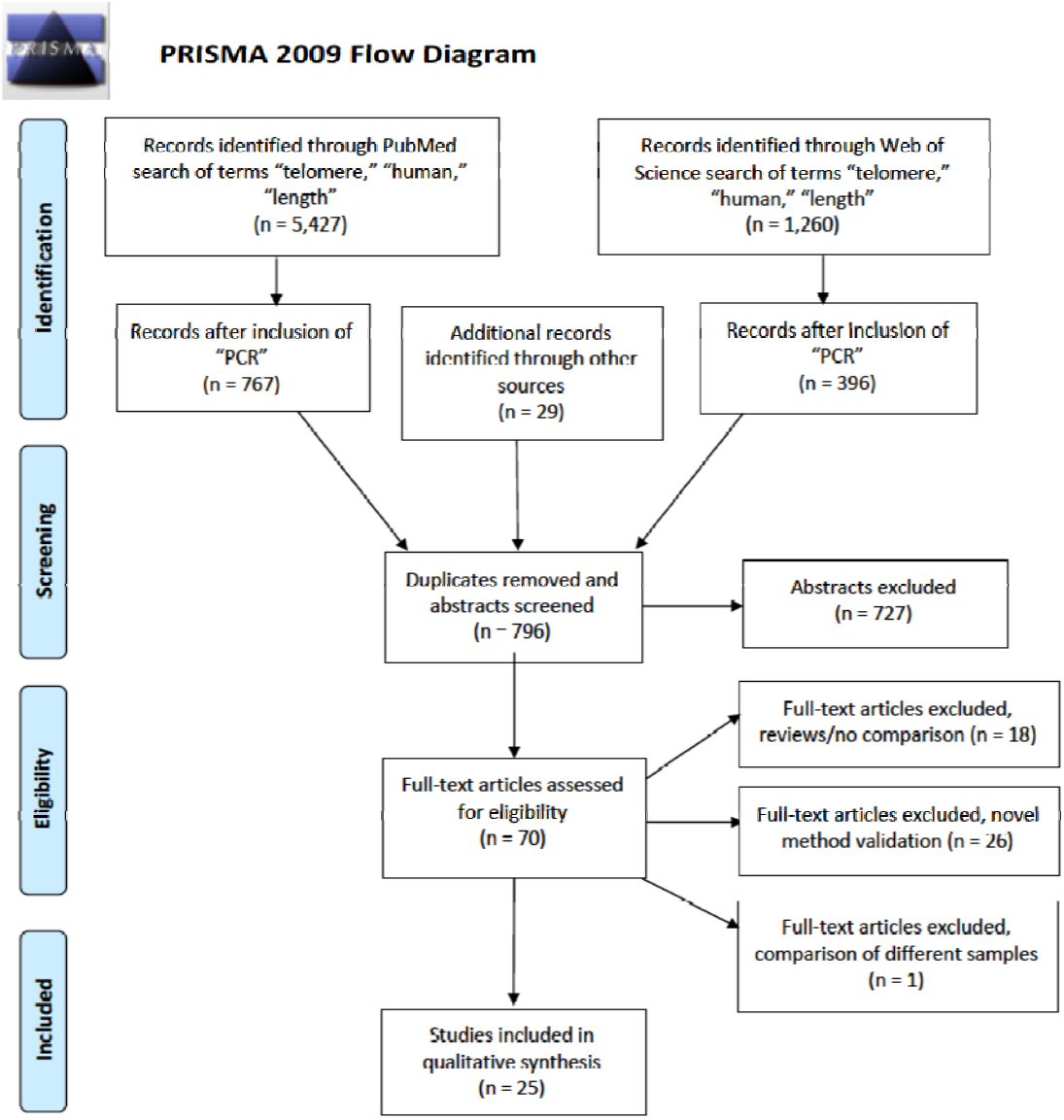


Fig 3. PRISMA flow diagram of literature search and article selection for inclusion. Date range searched was between Jan 1 2002 to May 10 2020. Other sources of article identification included suggestions from participants of the TRN and review of reference lists of selected papers.

Reporting review:

Included papers were evaluated for quality of methodologic reporting using two different indices of reporting guidelines for PCR-based telomere studies. The first was created through consensus of the initial participants in the TRN (Supplemental Table 1). The second was derived from recommendations published by Morinha et al. 2020 (Supplemental Table 2) [49]. We included both guidelines for two reasons. First, there is not empirical data to distinguish between the two guidelines in terms of ensuring rigor and reproducibility for the field. Second, as several of the papers reviewed were authored by participants involved in the creation of the TRN guidelines, the inclusion of both guidelines provided some degree of impartiality. Both guidelines contain overlap with the MIQE guidelines and include characterization of the importance of each recommendation [50]. In terms of specific differences, the Morinha recommendations included several pre-analytic considerations not included in the TRN guidelines (e.g. volume of sample processed, robotic instrumentation vs manual), while storage buffer and the percentage of samples tested for DNA integrity were included in the TRN guidelines but not in the Morinha guidelines. The latter also required greater detail for qPCR validation such as the standard curve and calibration samples, as well as a requirement for the melt curve and Ct of the negative control, which were not included in the TRN guidelines. Both guidelines assess the comprehensiveness of the information describing the PCR assay itself as well as analytic considerations for final TL determination. A grading rubric for each set of guidelines was developed to reduce subjective reviewer interpretation (Supplemental Table 3). A composite assessment for each index was divided into three subcategories for the TRN guidelines and five subcategories for the Morinha guidelines. These broadly encompassed sample collection and processing, DNA quality metrics and storage; PCR assay components and quality control; and data analysis. Two of three reviewers

independently assessed each article for fulfillment of reporting guidelines (ARL, LWYM, and SSD). The scores for each individual item were compared and discrepancies resolved by the third reviewer.

Additional characteristics assessed included sample blinding prior to analyses, single lab or multi-lab testing, conversion/transformation of raw TL measurement prior to comparison, and whether the study design evaluated repeatability and/or reproducibility. Lastly, when available, sample size, means and standard deviation of TL measurement are included to assess study power. Although several studies included means and SD of the entire sample, only a subset reported the means and SD of the samples utilized in the method comparison analyses. This review only included method comparisons that involved at least one PCR-based method as currently reporting guidelines are only available for PCR-based methods. As the majority of PCR-based measurements of TL are relative, it was not possible to address accuracy.

Correlation between methods was assessed by using Pearson's r or r^2 values where provided. Weighted average correlation coefficients were determined for each type of comparison by converting reported Pearson's r values (or the square root of reported r^2 values) to Fisher's z values, and weighting by sample size. A forest plot was generated from the weighted r average, total sample size for that correlation, and 95% confidence interval (CI) range using DistillerSR Forest Plot Generator from Evidence Partners (https://www.evidencepartners.com/resources/forest-plot-generator/).

ICC calculation:

Given established analytic shortcomings related to the use of the CV as a metric of testing the repeatability of TL, or the correlation of TL measurement between assays, raw data from cross method comparison studies was used to calculate ICCs for comparison between methods where available [40, 51, 52]. ICCs for one study were also calculated using a two-way, single measurement, absolute

188 agreement, random effects model, known as ICC(A,1) and for average measurements ICC(A,k) in
189 McGraw & Wong's (1996) terminology [53]. The R script used for calculating ICC and associated
190 instructions can be found in the Supplemental Methods.

191 To provide guidance for future study design, we present several different power analyses
192 outlining the relation between sample size, ICC, and ability to detect group differences. All calculations
193 assumed a realistic (true) standard deviation of 650 base pairs (bp), and N is the combined n of the two
194 groups to be compared and was assumed to be equally distributed among the two groups. Power
195 analysis for cross-sectional comparisons was done using G*Power[54], while power of longitudinal
196 comparisons was estimated through simulations. To examine the impact of variation in ICC on
197 longitudinal TL studies, the statistical power to detect a significant change (paired-t-test) in telomere
198 length of 25 bp/year for sample sizes of 25, 50 and 100 individuals, and an interval of 8 years between
199 baseline and follow-up (i.e. on average 200 bp in total), as a function of measurement repeatability (e.g.
200 reliability) expressed as the ICC. Measurement error was simulated by adding a random number from a
201 normal distribution to the true TL, with the error set at different levels to generate variation in ICC
202 between simulations. Population SD of telomere length was assumed to be 650 bp at both time points
203 and telomere shortening was simulated assuming a Poisson distribution with mean/variance of
204 25bp/year. This is close to the mean shortening rate typically observed in adults in studies where the
205 age-dependent SD is estimated to be close to 650 bp, and thus the scaling of shortening rate to the
206 overall variance is realistic. Furthermore, power of comparisons using data with another SD can be read
207 from the graphs after rescaling the data to have an SD of 650.

208 Results

209 The initial search revealed 5427 articles and the inclusion of "PCR" as a search term limited the
210 results to 767 articles, whose abstracts and methods were read (Figure 3). An additional 30 articles were

211 identified through assessment of the references of method validation papers and other included cross-
212 method validation studies. A review of these 797 abstracts identified 70 articles for assessment of the
213 full text and supplemental information to determine inclusion in this review. 26 articles were
214 determined to be novel method validation and excluded. 18 articles were excluded as they were either
215 reviews or did not include direct method comparisons. One article was excluded due to the
216 determination that the DNA samples used for cross-method comparison were obtained at different time
217 points. We also included nine papers that, while not specifically designed as a cross method comparison,
218 included sufficient details comparing TL measurement using different assays. This resulted in a total of
219 25 articles included in this review (Table 1).

220 **Table 1: Characteristics of included papers.**

Study	Year	Methods	Multi-lab?	Sample Type	Sample Size	Age Range (years)	TL Mean ± SD
Martin-Ruiz*[55]	2004	qPCR x TRF		cell lines	22	N/A	NR
Gardner*[56]	2007	qPCR x TRF		kidney	32	0.1-71.4	NR
Hunt*[57]	2008	qPCR x TRF		blood	72, 162	19-93	T/S = 2.71 ± 0.08 (n=36); 2.71 ± 0.05 (n=81); 3.01 ± 0.10 (n=36); 3.05 ± 0.06 (n=81)
Salpea*[58]	2008	qPCR x TRF	Yes	blood	32	24-54	T/S = 1.38 (1.31-1.44) (n=765)
Ehrlenbach*†[59]	2009	qPCR x TRF		blood	56	51-81	T/S = 1.31 (0.95-1.90) (n=669)
Pavesi*[60]	2009	qPCR x TRF		blood, cord blood	28	0-97	NR
Aviv[42]	2011	qPCR x TRF	Yes	blood	50	41-70	NR
Imam*[61]	2012	qPCR x FlowFISH		blood, cord blood, DBS	29	0	T/S = 6.23 ± 1.54 (n=35); 5.64 ± 1.54 (n=51)
Zanet[62]	2013	qPCR x MMqPCR		blood	32	2-59	NR
Gutierrez-Rodrigues[46]	2014	qPCR x FlowFISH x TRF		blood, cord blood	70, 51	0-88	NR
Martin-Ruiz[43]	2015	qPCR x MMqPCR x TRF x STELA	Yes	cell lines, placenta	12	N/A	NR
Eisenberg[44]	2015	MMqPCR x TRF	Yes	blood	190	NR	NR
Jodcyzyk*[63]	2015	qPCR x TRF		blood	20	0-35/50	T/S = 1.184 ± 0.371 (n=677); 1.104 ± 0.153 (n=351)
Panero*[64]	2015	aTL x TRF		bone marrow cells	102	30-87	7.2 ± 0.46 kb (n=48); 6.15 ± 0.41 kb (n=54)
Hsieh[65]	2016	qpCR x MMqPCR x TRF x flowFISH		blood, cord blood, DBS, placenta, cell lines	33 - 84	NR	NR
Lynch*†[66]	2016	qPCR x TRF	Yes	blood	101, 111	53-63, 26-43	6.4 kb (4.3-8.3) (n=101); 6.3 kb (5.1-7.9) (n=111)
Wand[67]	2016	qPCR x TRF x FlowFISH		blood	11	24-69	NR
Behrens[68]	2017	qPCR x FlowFISH x T/C-FISH x TRF		blood, cord blood	83 - 99	0-81	NR

Khincha†[69]	2017	qPCR x TRF x FlowFISH		blood	35, 53	3-69	T/S = 0.50 (0.21-1.48) (n=35); 0.91 (0.45-1.27) (n=53)
Lee[70]	2017	qPCR x TRF x WGS		cell lines	20	N/A	NR
Tarik[71]	2018	MMqPCR x TRF		blood	94	NR	T/S = 1.02 ± 0.32
Wang[72]	2018	qPCR x FlowFISH		blood	181	19-53	NR
Gadalla*[45]	2018	qPCR x qPCR	Yes	blood	100	NR	NR
Ventura Ferreira[73]	2019	MMqPCR x FlowFISH		blood	105	24-66	NR
Ropio[74]	2020	qPCR x TRF x aTL		cell lines	20	N/A	4.32 ± 0.143 kb

221 Sample size reflects the *n* used for method comparisons. If the TL mean and standard deviation (SD)
222 reflects reported values for the entire sample size, the n is included in parentheses.
223 * Papers not specifically intended to compare correlation of methods.
224 † TL median and range were provided in lieu of mean and SD.
225 N/A: not applicable to this sample type.
226 NR: not reported
227 DBS: dried blood spot

228 Paper Characteristics

229 The most common methods comparison among the 25 papers evaluated in this review was
230 monoplex quantitative PCR (qPCR) and the telomere restriction fragment (TRF) method by Southern blot
231 (n=17). Four studies compared multiplex qPCR (MMqPCR) with TRF. Seven studies compared qPCR with
232 the flow-FISH method, and two studies compared MMqPCR with flow-FISH. Three studies involved
233 comparisons of MMqPCR and qPCR, and two studies involved direct comparisons of qPCR methods or
234 separate laboratories. Other methods examined included STELA (compared with qPCR, MMqPCR, and
235 TRF in one study) [43], and absolute TL (two studies, one comparing aTL to TRF and one comparing to
236 aTL to qPCR) [64, 75]. T/C-FISH was also examined in one study as it related to qPCR [68]. Additionally,
237 one paper compared the correlation of several whole genome sequencing (WGS) platforms to qPCR-
238 based measurement[70]. Note some studies compared more than two methods [43, 46, 65, 67-70, 72,
239 75].

240 Whole blood was the most common sample type used (n=19), but cord blood (n=5), peripheral
241 blood mononuclear cells (PBMCs) (n=4), and cell lines (n=5) were also utilized as well as a range of other
242 sample types. Several studies reported on more than one sample type.

243 The reported sample size for the cross-method comparisons ranged from 12 to 181 and only 7
244 papers had a total sample size greater than 100. Five studies reported the means and standard
245 deviations and two reported the median and range of TL measurements for the study. Two studies
246 provided the raw values of the PCR-based TL measurements.

247 Overall quality of reporting of PCR assay methodology

248 Of the 25 studies included in this systematic review, the average completion score across both
249 reporting guidelines was 51%, with an average of 52% for the TRN guidelines and 50% for the Morinha
250 guidelines (Table 2). Overall, papers included between 26-75% and 29-78% of the recommended
251 reporting metrics for the TRN and Morinha guidelines, respectively. Some metrics were consistently
252 reported in nearly all papers, including the sample type, single copy gene name, and type of PCR method
253 utilized. However, only about 10% of the included papers reported on sample storage, PCR efficiencies,
254 or the number of samples excluded due to quality concerns with the assay.

255 **Table 2. Results of reporting guidelines analysis for TRN and Morinha Reporting Guidelines.**

	Category	**Average Score**
TRN Reporting Guidelines	Sample	37%
	qPCR assay	75%
	Data analysis	39%
	Total	**52%**
Morinha Reporting Guidelines	Sample	46%
	DNA extraction	32%
	qPCR validation	14%
	qPCR protocol	66%
	Data analysis	78%
	Total	**50%**
	Average of Both	**51%**

256

257 Average score reflects the average score across all papers for each reporting category. Total scores
258 reflect the average score across all papers from all reporting items.

259 DNA processing

260 For both the TRN guidelines and Morinha guidelines, reporting of sample type, storage, DNA extraction,
261 and DNA quality/integrity was poor, with an average of 37% for the TRN Sample/DNA category, 46% for
262 the Morinha Sample category, and 32% for the DNA category of the Morinha guidelines. Storage
263 conditions for both the biological samples and extracted DNA were poorly reported, with 24% or less of
264 studies providing this information (Table 3). Fewer than half of the studies reported on metrics related
265 to DNA integrity.

266 **Table 3. Item-specific reporting results of the TRN Reporting Guidelines.**

Category	Reporting Item	# Papers Reported
Sample type, storage, DNA extraction and integrity	Sample type	25
	Sample storage temperature	5
	Sample storage time before extraction	2
	Sample storage buffer	6
	DNA extraction method	20
	DNA storage conditions	4
	DNA freeze-thaw cycles	4
	Method of documenting DNA quality/integrity	8
	% of samples tested for DNA quality/integrity	6
qPCR assay	State type of PCR method	25
	PCR machine	21
	Source of master mix & reagents	22
	Final reaction volume	19
	Telomere primer sequences	21
	Telomere primer concentration	21
	Single copy gene name	25
	Single gene primer sequence	22
	Singe gene primer concentration	20
	Full PCR program	18

	PCR efficiency of both primers	3
	Source of control samples	14
	Concentration of DNA standard	13
Data analysis	Mean and standard deviation/median and range of TL	13
	Number of sample replicates	24
	Level of independence of replicates	14
	Analytic method to determine final TL	23
	Method of accounting for variation between replicates	11
	Method of accounting for well position effects	5
	Method of accounting for between-plate effects	12
	% of samples repeated as a result of failed QC	3
	% of samples excluded from analysis	3
	Acceptable range of PCR efficiency for primers	6
	ICCs of sample/study groups	2
	T/S ratio transformed to Z score prior to analysis	2

PCR assay

Reporting on PCR assay conditions and quality control varied. While many metrics of the PCR assay were well-reported, only 18 of 25 studies reported the full cycling conditions. The lowest reporting metric related to PCR was experimental efficiency, with only 12% reporting actual PCR efficiencies. Additionally, just over half (56%) of studies reported the source of their control samples.

Analytic approaches

Several key reporting gaps were noted in relation to assay quality control and analytic approaches to determining final TL. As with any biologic assay, there is the potential that a specific sample will fail quality control metrics and need to be repeated. Only six papers reported on the number of samples repeated and/or the number of samples that failed quality control. While all but one study reported the numbers of sample replicates, surprisingly, only 14 studies reported on the level of independence of sample replicates (e.g. replicates run on the same plate or on different plates/different

280 times) and only half of the studies reported the means and standard deviations (or median and range) of
281 the T/S ratio.

282 Cross-laboratory studies

283 Only six studies compared analyses across more than one laboratory (Table 1) [42-45]. Of these
284 studies, three described how samples were blinded before analyses. Further, of these cross-laboratory
285 studies, only three studies included the same assay performed in different laboratories [43, 45, 66].

286 Reproducibility

287 Reproducibility, a critical criterion for biologic assays, refers to the relation between
288 measurements using the same assay in different locations or the comparison of values generated using
289 different measurement procedures. This systematic review attempted to assess the relative
290 reproducibility of PCR-based measure of TL in different laboratories as well as the reproducibility
291 precision, e.g. the closeness of two or more measurements, in TL measurement using different methods.

292 Relative reproducibility

293 The current literature does not provide sufficient data to address the relative reproducibility, as,
294 to date, only three studies have tested this directly by performing the same assay in different
295 laboratories or settings. In one study that blinded comparison samples before they were sent to the
296 external laboratories, the median CV across laboratories for qPCR was 18.3%, while the median CV for
297 STELA/TRF based TL measurement was 9.2% [43]. However given the dependence of CVs on the y-
298 intercept, the interpretation of these CVs remains challenging [40]. In the second study, where samples
299 were not blinded before being assayed, the reported within-lab CVs for replicate qPCR measurements
300 were 2.5% and 8.6% [45]. As the laboratories involved utilized different PCR primers, and slightly
301 different methods, it was not possible to directly compare cross-laboratory reproducibility. The third

302 study found inter-assay CVs of 12.0 and 1.2% in two participating labs performing qPCR, but an
303 additional laboratory's results were excluded from analysis due to an extremely high CV of 27%.
304 Correlation between each laboratory and TRF results were calculated, but no correlation results were
305 provided for the two qPCR assays, and ICC estimates were not reported.

306 Reproducibility precision

307 Reproducibility precision, e.g. the closeness of two or more measurements using different
308 techniques, was addressed to some extent in 19 of the 25 studies reviewed. However, only six studies
309 involved assays performed in different laboratories. The correlation of TL measurement with qPCR-
310 based assays to other assays was, for the most part, reported as linear regression and correlation
311 coefficients (Figure 4). Other papers reported Bland-Altman analyses or did not report a measure of
312 correlation at all. One paper reported mean LTL values for both TRF and qPCR, but did not report a
313 measure of correlation [66]. When examining these results, it should be kept in mind that these
314 methodological studies were generally done in laboratories with extensive experience in the focal
315 technique, and as such are unlikely to be representative of the field at large.

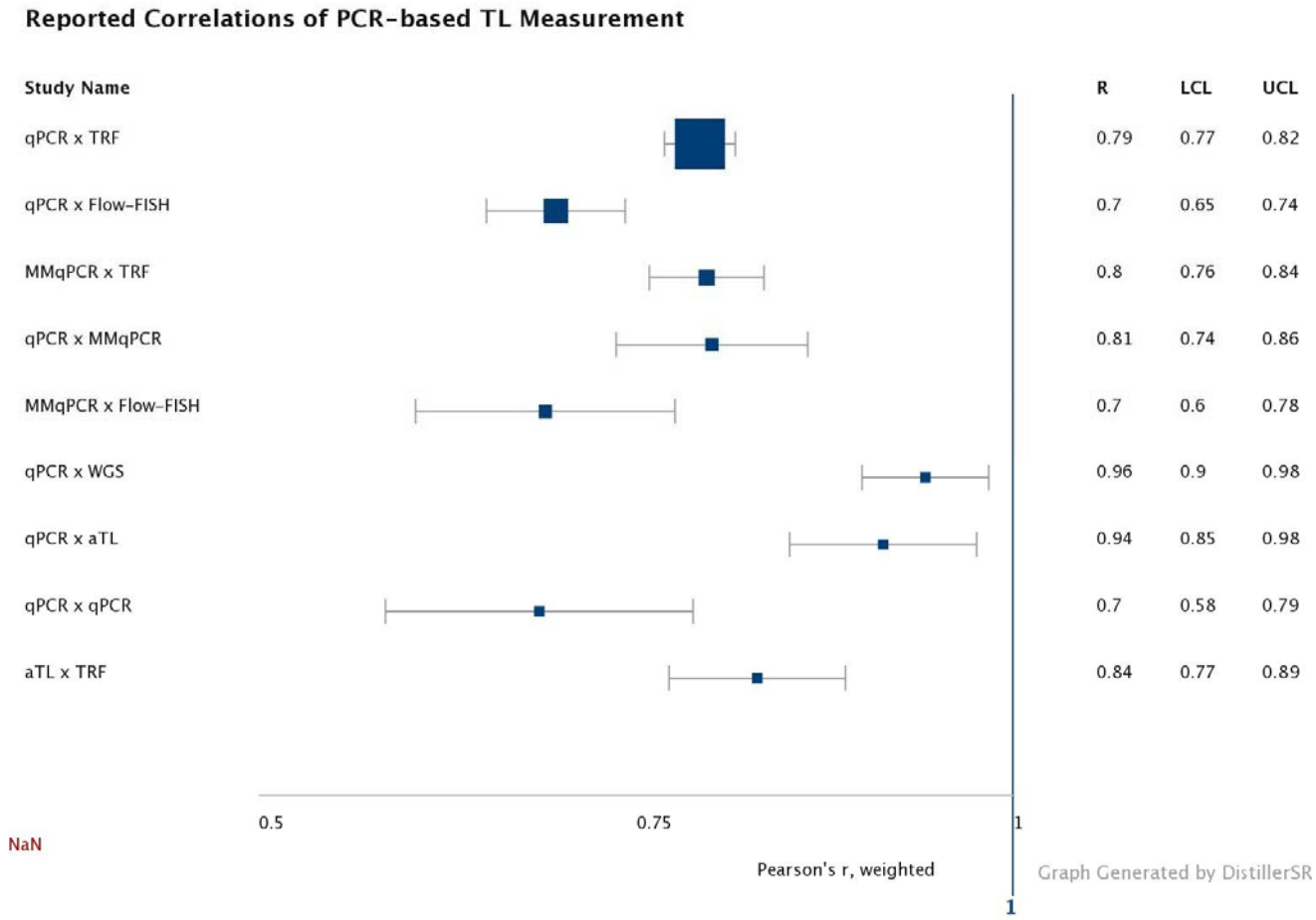


Figure 4. Weighted average correlation (Pearson's r) and 95% CI for each method compared with qPCR, MMqPCR, or aTL. LCL: lower confidence limit. UCL: upper confidence limit. The point sizes for each comparison are weighted by the number of individual papers with that comparison.

As qPCR and TRF were the most common methods compared, these studies typically reported high correlation, with a weighted correlation coefficient for all studies around 0.75. Correlation of other methods with qPCR or MMqPCR were more variable. No studies have compared MMqPCR to aTL, or whole genome sequencing (WGS). Only one paper each compared qPCR and qPCR (in separate labs), aTL and TRF, qPCR and aTL, or qPCR with WGS. To our knowledge no studies have compared WGS data with aTL, although studies have compared TRF and WGS [23].

326 In five of the papers in this review, linear regression was used to extrapolate TL into kilobases
327 (kb) from the T/S ratio using TRF values. One paper converted T/S ratio to kb before analysis of the
328 correlation between methods [46]. In two cases, T/S ratio were converted to kb prior to TL comparison
329 utilizing Bland-Altman analysis [68, 73]. In two of the five papers, the conversion of the T/S ratio to bp
330 was based on analyses extrapolated from different data or measured on a different sample type, raising
331 substantial concerns on the true measurement with uncertain implications for the r value [46, 73].
332 Beyond concerns related to the source of the data utilized for conversion from T/S to bp before
333 comparison across methods, this analytic approach likely to leads to inaccurate reporting [46]. Only two
334 studies utilized the TRN-recommended procedure of transformation to z-score before comparison [43,
335 66]. When comparing relative TL estimates such as the T/S ratio generated from qPCR, transformation of
336 these values to z-scores will yield more informative results and improve ability to compare results
337 between laboratories or assays [76].

338 Repeatability

339 Repeatability, the precision in measurements that include the same procedure/locations,
340 revealed the greatest variation in both lab and assay specific precision and between methods. In these
341 studies, the number of replicates for a specific DNA sample ranged from 1 replicate (e.g. sample
342 analyzed twice) to five replicates (each sample analyzed six different times). Additionally, only four
343 reviewed studies reported the number of samples that were repeated due to within replicate variance,
344 despite clear acknowledgement in the field that a proportion of all studies will ultimately require
345 repeated assays of telomere length as a result of between-replicate variance. While of limited utility in
346 confirming precision, the intra-assay reported CVs for PCR-based methods (qPCR, MMqPCR) in this
347 review ranged from 2.5 to 12%, and the inter-assay CVs ranged from 3.97 to 15.9%. Inter-assay CVs for
348 TRF ranged from 1.25 to 6.3%, with intra-assay CV reported in only one paper as 1.20%. Inter-assay CVs

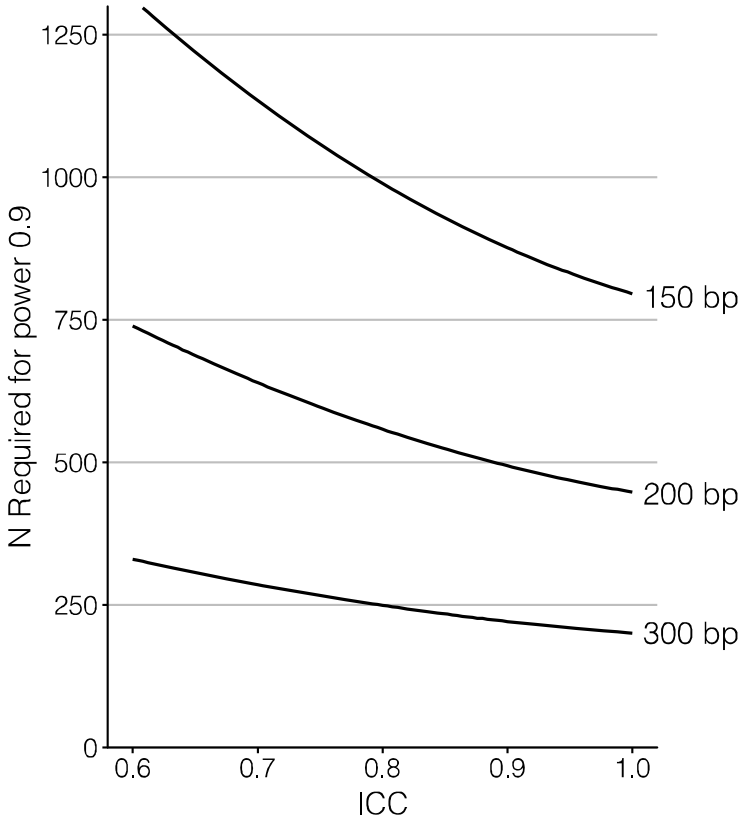
349 for flow-FISH were reported as 9.3% and 10.8% in two papers, with only one reporting an intra-assay CV
350 of 9.6%. One paper examining the aTL assay reported its inter-assay CV as 6.7% and intra-assay CV as
351 2.5%. We emphasize, however, that there are analytic concerns related to the use of CVs for cross
352 laboratory comparisons [40, 52], and directly converting CVs to ICC values is not possible.

353 Only two studies utilized ICC analyses to examine the repeatability of replicates, reporting ICCs
354 of 0.89 and 0.92 for PCR-based measurement [69, 71]. To expand data on the repeatability of PCR-based
355 and other TL methods, raw data was obtained from authors of a subset of these papers, and ICCs
356 independently calculated. Calculated/reported ICC for TRF methods ranged from 0.92- 0.99 in the
357 studies included in this review and are consistent with the ICCs reported in existing studies utilizing the
358 TRF (0.95 to 0.99). However, it is of note that these ICCs were almost entirely the result of TRF
359 measurement in one laboratory. The ICCs for qPCR-based methods in reviewed papers ranged from
360 0.89-0.92, including the two reported in manuscripts and an additional ICC calculated from raw data (ICC
361 = 0.915, SE = 0.023, 95% confidence interval: [0.860, 0. 946], P<0.001; reported CV 6.5%) [40, 77]. ICCs
362 for MMqPCR (triplicates on the same plate) from one study were run separately based on year of
363 analyses. In one set (n=873) run across different PCR plates in initial and duplicate runs, the ICCs were
364 ICC(A,1) = 0.82 (95% CI 0.79-0.84) and ICC(A,k)=0.90 (95% CI 0.88-0.91). Because these samples were re-
365 run due to initially high intra-assay CVs, this is possibly an under-estimate of the true ICC value. For
366 these same samples, TRF ICCs were calculated from duplicate gels on a subset (n=159) and the inter-gel
367 ICC = 0.96 (95% CI 0.94-0.97). However, we note that these TRF analyses were conducted by a trainee
368 which likely decreased repeatability compared to what is typical of experienced technicians. Given the
369 significant variation in methodologic and raw data reporting, and the wide variability in published CVs, it
370 is likely that the majority of existing TL studies not specifically comparing methods would have
371 significantly lower ICCs.

Determination of effects of ICC variability on sample size and study power

Our systematic review revealed wide variation in TL measurement repeatability. No biologic assay is perfect, and laboratories measuring any biologic substrate vary in their own internal quality control and repeatability. To provide general guidance for investigators, we therefore conducted analyses to evaluate the impact on power and sample size across a range of ICCs.

In Figure 5A, we present the sample size required to test effect sizes of 150, 200, and 300 bp with a t-test with a power of 0.9, as a function of measurement error as expressed in the ICC. To contextualize the differences: 150 bp is the approximate difference found between the sexes, and 300 bp is the approximate difference observed between individuals with and without atherosclerotic cardiovascular disease [78]. As directly converting bp to T/S ratios is not feasible in this analysis and the analyses below, we suggest that investigators using T/S or other relative TL measurements use standard deviation (SD) differences to estimate power. For example. a difference of 150 bp is equal to 150 / 650 = 0.23 SD, which can be converted to a T/S difference when the SD of the T/S measurements is known. Estimates



395 of potential difference can be extracted from existing literature related to their exposure or outcome of
396 interest when considering study design and sample size.

397 **Fig. 5A.** The sample size required to test effect sizes of 150, 200 and 300 bp with a t-test with a power
398 of 0.9, as a function of measurement error as expressed in the ICC. Calculations assumed a realistic
399 (true) standard deviation of 650 bp and power analysis was done using G*Power [54]. N is the combined
400 n of the two groups to be compared and was assumed to be equally distributed among the two groups.

401 Finally, we present the statistical power of different sample sizes to detect a significant
402 difference in telomere shortening rate of 33% using longitudinal data, as a function of measurement
403 reliability expressed as ICC (Fig. 5B, C). This analysis revealed that even with a high ICC (>0.9), large
404 sample sizes are required to yield sufficient statistical power to detect even substantial differences in
405 telomere shortening rate, in particular when the follow-up period is short. This is due to the rate of
406 telomere shortening being low (here 25 bp/year) compared to the TL variation between individuals
407 (here an SD of 650). The rate of base pair loss in infants and children is likely significantly different and,
408 but as of yet is poorly characterized (but see [79]).

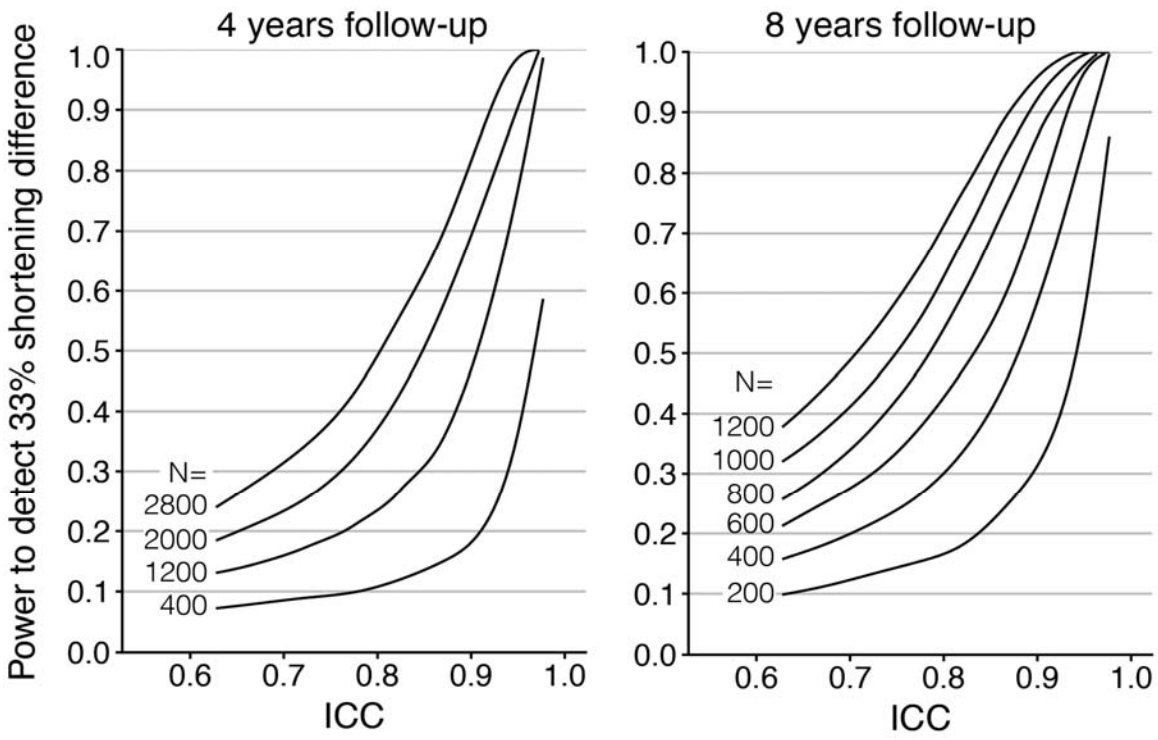


Fig. 5B, C. Power to detect a 33% change of telomere shortening rate, up or down, with $p<0.05$ relative to a baseline shortening rate of 25 bp/year. D. Four-year follow-up period. E. Eight-year follow-up period. Power was calculated for sample sizes as shown (200 – 2800), equally divided over the two levels of telomere shortening rate. Baseline telomere shortening was simulated assuming a Poisson distribution with mean/variance of 25, and population SD of telomere length was maintained at 0.65 kb at both time points.

Discussion:

This systematic review found a total of 25 papers documenting comparison between TL measured using a PCR-based methodology and another TL assay. Until recently, no publication reporting guidelines existed for qPCR-based TL measurement. Our review focused on method comparison studies with the expectation that critical assay parameters and methodologic description would be more detailed and specific. Our review, using two separately developed reporting guidelines, found that, on

422 average, only half of the recommended factors were documented, indicating the need for increased
423 methodologic reporting and wider awareness of reporting recommendations. The lowest reporting was
424 related to information about the validation of PCR-based assays outlined in the Morinha guidelines, with
425 only seven papers including any of the recommended factors. PCR efficiencies, a key reporting
426 requirement in both guidelines and the MIQE guidelines, was absent from the majority of papers with
427 only six mentioning the PCR efficiency parameters and only three documenting the actual PCR
428 efficiencies. Given that all PCR-based methods either assume or specifically calculate the PCR efficiency
429 when determining the T/S ratio, and that, in general, the determination of the T/S ratio assumes similar
430 efficiencies for the single gene and the telomeric primers, the absence of this key metric is concerning.
431 Fewer than half of studies failed to comment on key pre-analytic factors, specifically sample storage
432 time and conditions, freeze-thaw cycles, and evaluation of DNA quality and integrity, all potential
433 sources of assay variability for both PCR and non-PCR-based TL assays that may contribute to current
434 debates in the field about the utility of TL [74, 80, 81]. Lastly, the reporting of the number of samples
435 failing initial quality control, repeated, or unable to be assayed was low. In laboratories routinely
436 performing TL measurement using any assay, a certain percentage of samples for each study will require
437 repeating and regularly a small subset may be unanalyzable for various reasons. While it is possible that
438 these factors were considered and monitored, the lack of reporting for this metric heightens the need
439 for increased attention to the proposed reporting recommendations. Moving forward, the widespread
440 dissemination of these qPCR reporting guidelines to study sections, peer reviewers, and scientists
441 represents an important next step in enhancing the scientific rigor of the field.

442 At this time, evaluation of the existing literature fails to provide sufficient evidence of the
443 relative or precision reproducibility of different TL assays. Our review identified only six studies that
444 included cross laboratory comparisons and, of these, only three evaluated PCR-based assays performed
445 in more than one lab. As the number of laboratories performing TL studies using PCR and other

446 methods continue to increase, the lack of clear data about cross-laboratory reproducibility and the
447 absence of existing DNA standards or other methods to account for cross-laboratory variation
448 substantially limits the ability to characterize relative reproducibility. In terms of reproducibility across
449 different methods (e.g. PCR and TRF, or PCR and FISH), the current variability in findings, particularly
450 when coupled with limited methodologic reporting, highlights the need for additional rigorous and
451 blinded cross laboratory studies that are adequately powered to accurately determine how TL in a
452 population measured using different assays truly relates. Although 17 studies evaluated the relationship
453 between qPCR and TRF, due to the wide range in reported correlations between TL measurements, the
454 relatively small samples sizes, and the insufficient analytic and assay blinding, there is currently
455 insufficient data to draw firm conclusions on the general correlation between TL measured with
456 different assays. The analytic consequences of using CVs to test the relationship between TL measured
457 using different assays has been discussed previously, as has the issues caused by the use of analytic
458 strategies such as conversion to base pairs instead of z scores, especially when extrapolating from data
459 produced in different laboratories or using different samples [40, 51]. In this review, we utilized existing
460 raw data from 3 included studies to provide preliminary data about precision reproducibility for PCR and
461 TRF studies. The wide range of ICCs calculated from these few studies, particularly for PCR-based
462 methods, and the low reporting of ICCs in the papers included in this review highlights the need to
463 increase attention to this specific reporting requirement. For many of the existing studies, the small
464 sample size and the lack of reporting of the means and standard deviations of TL prevents objective
465 determination of whether any of the current studies were adequately powered. Beyond these concerns,
466 the over-representation of data from specialized laboratories, particularly for TRF, the applicability of
467 much of the existing data to the wider telomere field is uncertain. For aTL and MMqPCR-based TL
468 measurements, the current paucity of published cross method comparisons limits the ability to form an
469 opinion of how TL measured with these assays relates to other methods.

470 Measurement precision is critical, in particular for longitudinal studies. Methodologies that are
471 low cost, practical, and simple to implement with standard laboratory equipment, especially when they
472 are innovative or high impact, are often rapidly implemented across laboratories with various levels of
473 expertise in the new methodology. Invariably this results in diverse protocols, analyses, and
474 methodologic reporting – consequences that are even more problematic when there is an absence of
475 consensus on best practices [50]. As with many other biologic assays, the development of reporting
476 guidelines for TL measurement has lagged behind the broad implementation of the methods themselves
477 [82-86]. The lack of consolidated guidance about factors, both pre-analytic and within the assay itself,
478 that contribute to measurement error when combined with the wide popularity of PCR-based TL
479 measurement undoubtedly contributed to discrepancies in the existing literature and failed study
480 replications. Similar to the MIQE guidelines, the reporting guidelines presented and tested in this
481 systematic review for PCR-based TL assays are meant as minimal reporting recommendations focused
482 on enhancing the reliability of results, consistency between different laboratories performing the same
483 assay, and increased experimental transparency and accuracy [50]. To assist investigators and reviewers
484 we highlight the overlapping recommendations with the MIQE guidelines, indicate whether a particular
485 requirement is desirable or essential, and provide references that support the selection of the particular
486 reporting requirement. Over the course of the next four years, the TRN expects to develop similar
487 reporting recommendations for other types of TL assays while conducting adequately powered and
488 scientifically rigorous studies to support these reporting guidelines, recognizing that individual
489 recommendations have varying levels of initial empirical support [87].

490 Despite the strengths of this review, there are several limitations. First, this review only focused
491 on assays applicable to population-based studies in humans. It does not address issues in other species
492 or assays that may have clinical utility but for which the requirements for sample types (e.g. fresh tissue
493 and/or live cells) or the cost/labor/expertise requirements (e.g. TeSLA, STELA) limit utilization in

494 population based studies. A second limitation is that we utilized reporting guidelines for qPCR-based
495 assays only. To date, specific protocol recommendations and reporting guidelines have not been
496 published for other TL assays (e.g. TRF, FISH) although detailed methodologic protocols do exist [88].
497 Additionally, it is possible that additional articles comparing TL assays may be available in other
498 databases or pre-print servers. However, many of the articles included in this review were not
499 specifically designed solely to compare TL measurement methods and would not be found through
500 standardized database searches. Further, it is unlikely that additional articles would change the general
501 picture emerging from this review. Finally, we note that while this article focused on precision and
502 reproducibility, accuracy of measurement is as important. Precisely inaccurate measures will be of
503 limited use to the scientific field, a factor that becomes more problematic when using relative estimates
504 and not true values as is the case in many TL assays. In the absence of a clear gold standard
505 measurement technique, accuracy is difficult to discern.

506 **Conclusions**

507 After careful examination of the existing literature, it is apparent that rigorous cross laboratory
508 and methodological studies must be an immediate priority for the field. To assist the field moving
509 forward, we include reporting guidelines for PCR-based TL assays and indicate specific scientific papers
510 that support these recommendations originally developed through consensus of the initial TRN
511 members and consultants. These guidelines do not outline a specific PCR methodology and, at this time,
512 we do not believe there is sufficient data to provide guidance on specific assay approaches or
513 components. Rather, these guidelines are provided to ensure reviewers and readers can adequately
514 assess the methodology and consider the implications of these factors for each study's findings. The
515 consistency in results across reporting guidelines (TRN, Morinha, MIQE) related to the integrity and
516 quality of both the initial biological sample and the DNA itself support the critical nature of this

517 reporting metric. In terms of assay reporting, increased attention of investigators and reviewers to
518 ensuring complete reporting of assay reagents and PCR efficiencies is also expected to enhance the rigor
519 of the field. TRN investigators are currently testing the impact of different pre-analytic factors, DNA
520 integrity, and PCR conditions to provide extant evidence of the importance of these parameters in
521 relation to precision and reproducibility. We recommend that studies be required to report ICCs in lieu
522 of CVs, as well as either the median or mean and standard deviation of TL. We also provide specific
523 guidance related to sample size and power that is contingent upon the ICC given the substantial impact
524 of differences in assay precision on the ability to determine true relationships and with the expectation
525 that this will be of use for investigators as they embark on new research studies. It is important to
526 balance assay cost, in terms of both time and reagents, with the needed sample size and statistical
527 power. Moving forward, investigators should carefully consider study design from this perspective,
528 recognizing that there is currently no "ideal" approach. Telomere research offers significant potential
529 across a diverse range of scientific fields with potential mechanistic insight into overlapping biological
530 pathways contributing to many of the leading causes of morbidity and mortality. Ensuring the highest
531 scientific rigor and precision, through accurate methodological reporting and rigorous testing of the
532 factors that contribute to assay variability, are requisite steps to ensuring that potential is achieved.

533 # Acknowledgements

534 The authors would like to acknowledge the TRN Advisory Committee (Daniel Belsky, Eileen Crimmins,
535 Sonja Entringer, Daniel Nettle, Jerry Shay, Daniel Nettle), U01 grantee P.I.s (Abraham Aviv, Jue Lin, Idan
536 Shalev, Yun-Ling Zheng), and NIA/NIEHS partners (Max Guo, Michelle Heacock, Lisbeth Nielsen) for their
537 input in creating the TRN qPCR Reporting Guidelines.

538 # References

539 1. Olonikov, A., *A theory of marginotomy.* J Theor Biol, 1971. **41**: p. 181-190.
540 2. Blackburn, E. and J. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal*
541 *ribosomal RNA genes in Tetrahymena.* Journal of molecular biology, 1978. **120**(1): p. 33-53.
542 3. Moyzis, R., et al., *A highly conserved repetitive DNA sequence,(TTAGGG) n, present at the*
543 *telomeres of human chromosomes.* Proceedings of the National Academy of Sciences, 1988.
544 **85**(18): p. 6622-6626.
545 4. Therkelsen, A., et al., *Staining of human telomeres with primed in situ labeling (PRINS).*
546 Cytogenet Cell Genet, 1995. **68**(1-2): p. 115-8.
547 5. Lansdorp, P., et al., *Heterogeneity in telomere length of human chromosomes.* Hum Mol Genet,
548 1996. **5**(5): p. 685-91.
549 6. Bryant, J., et al., *Measurement of telomeric DNA content in human tissues.* Biotechniques, 1997.
550 **23**(3): p. 476-8, 480, 482, passim.
551 7. de Pauw, E., et al., *Assessment of telomere length in hematopoietic interphase cells using in situ*
552 *hybridization and digital fluorescence microscopy.* Cytometry, 1998. **32**(3): p. 163-9.
553 8. Hultdin, M., et al., *Telomere analysis by fluorescence in situ hybridization and flow cytometry.*
554 Nucleic Acids Res, 1998. **26**(16): p. 3651-6.
555 9. Narath, R., et al., *Automatic telomere length measurements in interphase nuclei by IQ-FISH.*
556 Cytometry Part A: the Journal of the International Society for Analytical Cytology, 2005. **68**(2): p.
557 113-120.
558 10. Nakamura, Y., et al., *Simple, rapid, quantitative, and sensitive detection of telomere repeats in*
559 *cell lysate by a hybridization protection assay.* Clin Chem, 1999. **45**(10): p. 1718-24.
560 11. Law, H. and Y. Lau, *Validation and development of quantitative flow cytometry-based*
561 *fluorescence in situ hybridization for intercenter comparison of telomere length measurement.*
562 Cytometry: The Journal of the International Society for Analytical Cytology, 2001. **43**(2): p. 150-
563 153.
564 12. Rufer, N., et al., *Telomere length dynamics in human lymphocyte subpopulations measured by*
565 *flow cytometry.* Nat Biotechnol, 1998. **16**(8): p. 743-7.
566 13. Cawthon, R., *Telomere measurement by quatitative PCR.* Nucleic Acids Research, 2002. **30**(10):
567 p. e47- 53.
568 14. Meeker, A.K., et al., *Telomere length assessment in human archival tissues: combined telomere*
569 *fluorescence in situ hybridization and immunostaining.* The American Journal of Pathology, 2002.
570 **160**(4): p. 1259-1268.
571 15. Perner, S., et al., *Quantifying Telomere Lengths of Human Individual Chromosome Arms by*
572 *Centromere-Calibrated Fluorescence in Situ Hybridization and Digital Imaging.* The American
573 Journal of Pathology, 2003. **163**(5): p. 1751-1756.
574 16. Baird, D., et al., *Extensive allelic variation and ultrashort telomeres in senescent human cells.* Nat
575 Genet, 2003. **33**(2): p. 203-7.
576 17. Canela, A., et al., *High-throughput telomere length quantification by FISH and its application to*
577 *human population studies.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5300-5.
578 18. Cawthon, R., *Telomere length measurement by a novel monochrome multiplex quantitative PCR*
579 *method.* Nucleic Acids Research, 2009. **37**: p. e21.
580 19. Bendix, L., et al., *The load of short telomeres, estimated by a new method, Universal STELA,*
581 *correlates with number of senescent cells.* Aging Cell, 2010. **9**(3): p. 383-97.

582 20. O'Callaghan, N. and M. Fenech, *A quantitative PCR method for measuring absolute telomere*
583 *length.* Biological procedures online, 2011. **13**: p. 3-3.
584 21. Kimura, M. and A. Aviv, *Measurement of telomere DNA content by dot blot analysis.* Nucleic
585 Acids Research, 2011. **39**(12): p. e84-e84.
586 22. Kibriya, M., et al., *Measurement of telomere length: a new assay using QuantiGene chemistry on*
587 *a Luminex platform*. 2014, AACR.
588 23. Ding, Z., et al., *Estimating telomere length from whole genome sequence data.* Nucleic Acids Res,
589 2014. **42**(9): p. e75.
590 24. Nersisyan, L. and A. Arakelyan, *Computel: Computation of Mean Telomere Length from Whole-*
591 *Genome Next-Generation Sequencing Data.* PLOS ONE, 2015. **10**(4): p. e0125201.
592 25. Lai, T., et al., *A method for measuring the distribution of the shortest telomeres in cells and*
593 *tissues.* Nature communications, 2017. **8**(1): p. 1-14.
594 26. McCaffrey, J., et al., *High-throughput single-molecule telomere characterization.* Genome
595 research, 2017. **27**(11): p. 1904-1915.
596 27. Jasmine, F., et al., *A novel pooled-sample multiplex luminex assay for high-throughput*
597 *measurement of relative telomere length.* American Journal of Human Biology, 2018. **30**(4): p.
598 e23118.
599 28. Dahlgren, P., et al., *Development of a New Monochrome Multiplex qPCR Method for Relative*
600 *Telomere Length Measurement in Cancer.* Neoplasia, 2018. **20**(5): p. 425-431.
601 29. Farmery, J.H.R., et al., *Telomerecat: A ploidy-agnostic method for estimating telomere length*
602 *from whole genome sequencing data.* Scientific Reports, 2018. **8**(1): p. 1300.
603 30. Lu, A., et al., *DNA methylation-based estimator of telomere length.* Aging, 2019. **11**(16): p. 5895-
604 5923.
605 31. Feuerbach, L., et al., *TelomereHunter–in silico estimation of telomere content and composition*
606 *from cancer genomes.* BMC bioinformatics, 2019. **20**(1): p. 1-11.
607 32. de Pedro, N., et al., *Analytical Validation of Telomere Analysis Technology® for the High-*
608 *Throughput Analysis of Multiple Telomere-Associated Variables.* Biological Procedures Online,
609 2020. **22**(1): p. 2.
610 33. Sethi, I., et al., *Dual Labeled Fluorescence Probe Based qPCR Assay to Measure the Telomere*
611 *Length.* Gene, 2020: p. 145178.
612 34. Luo, Y., et al., *Massively parallel single-molecule telomere length measurement with digital real-*
613 *time PCR.* Science Advances, 2020. **6**(34): p. eabb7944.
614 35. Hemann, M., et al., *The shortest telomere, not average telomere length, is critical for cell*
615 *viability and chromosome stability.* Cell, 2001. **107**(1): p. 67-77.
616 36. Xu, Z., et al., *The length of the shortest telomere as the major determinant of the onset of*
617 *replicative senescence.* Genetics, 2013. **194**(4): p. 847-857.
618 37. Montpetit, A., et al., *Telomere length: a review of methods for measurement.* Nursing research,
619 2014. **63**(4): p. 289.
620 38. Shay, J.W. and W.E. Wright, *Telomeres and telomerase: three decades of progress.* Nat Rev
621 Genet, 2019. **20**(5): p. 299-309.
622 39. Aubert, G., M. Hills, and P. Lansdorp, *Telomere length measurement-caveats and a critical*
623 *assessment of the available technologies and tools.* Mutat Res, 2012. **730**(1-2): p. 59-67.
624 40. Eisenberg, D., *Telomere length measurement validity: the coefficient of variation is invalid and*
625 *cannot be used to compare quantitative polymerase chain reaction and Southern blot telomere*
626 *length measurement techniques.* International Journal of Epidemiology, 2016. **45**(4): p. 1295-
627 1298.
628 41. Ehrlich, M., *DNA hypermethylation in disease: mechanisms and clinical relevance.* Epigenetics,
629 2019. **14**(12): p. 1141-1163.

630 42. Aviv, A., et al., *Impartial comparative analysis of measurement of leukocyte telomere*
631 *length/DNA content by Southern blots and qPCR.* Nucleic Acids Research, 2011. **39**(20): p. e134.
632 43. Martin-Ruiz, C., et al., *Reproducibility of telomere length assessment: an international*
633 *collaborative study*. International journal of epidemiology, 2015. **44**(5): p. 1673-1683.
634 44. Eisenberg, D., C. Kuzawa, and M.G. Hayes, *Improving qPCR telomere length assays: Controlling*
635 *for well position effects increases statistical power.* American Journal of Human Biology, 2015: p.
636 n/a-n/a.
637 45. Gadalla, S., et al., *No association between donor telomere length and outcomes after allogeneic*
638 *unrelated hematopoietic cell transplant in patients with acute leukemia.* Bone marrow
639 transplantation, 2018. **53**(4): p. 383-391.
640 46. Gutierrez-Rodrigues, F., et al., *Direct Comparison of Flow-FISH and qPCR as Diagnostic Tests for*
641 *Telomere Length Measurement in Humans*. PLoS ONE, 2014. **9**(11): p. e113747.
642 47. Bodelon, C., S. Savage, and S. Gadalla, *Telomeres in molecular epidemiology studies.* Prog Mol
643 Biol Transl Sci, 2014. **125**: p. 113-31.
644 48. Plant, A., R. Hanisch, and National Institute of Standards and Technology, *Reproducibility and*
645 *Replicability in Science, A Metrology Perspective* 2018: A Report to the National Academies of
646 Sciences, Engineering and Medicine Committee on Reproducibility and Replicability in Science.
647 49. Morinha, F., P. Magalhães, and G. Blanco, *Standard guidelines for the publication of telomere*
648 *qPCR results in evolutionary ecology.* Molecular Ecology Resources, 2020. **20**(3).
649 50. Bustin, S., et al., *The MIQE Guidelines: Minimum Information for Publication of Quantitative*
650 *Real-Time PCR Experiments*. Clinical Chemistry, 2009. **55**(4): p. 611-622.
651 51. Verhulst, S., et al., *Response to: Reliability and validity of telomere length measurements.*
652 International journal of epidemiology, 2016. **45**(4): p. 1298-1301.
653 52. Nettle, D., et al., *Consequences of measurement error in qPCR telomere data: A simulation study.*
654 PLoS One, 2019. **14**(5): p. e0216118.
655 53. McGraw, K. and S. Wong, *" Forming inferences about some intraclass correlations coefficients":*
656 *Correction.* 1996.
657 54. Faul, F., et al., *Statistical power analyses using G* Power 3.1: Tests for correlation and regression*
658 *analyses*. Behavior research methods, 2009. **41**(4): p. 1149-1160.
659 55. Martin-Ruiz, C., et al., *Stochastic variation in telomere shortening rate causes heterogeneity of*
660 *human fibroblast replicative life span.* J Biol Chem, 2004. **279**(17): p. 17826-33.
661 56. Gardner, J.P., et al., *Telomere dynamics in macaques and humans.* J Gerontol A Biol Sci Med Sci,
662 2007. **62**(4): p. 367-74.
663 57. Hunt, S.C., et al., *Leukocyte telomeres are longer in African Americans than in whites: the*
664 *National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study.*
665 Aging Cell, 2008. **7**(4): p. 451-8.
666 58. Salpea, K.D., et al., *The association of telomere length with paternal history of premature*
667 *myocardial infarction in the European Atherosclerosis Research Study II*. J Mol Med (Berl), 2008.
668 **86**(7): p. 815-24.
669 59. Ehrlenbach, S., et al., *Influences on the reduction of relative telomere length over 10 years in the*
670 *population-based Bruneck Study: introduction of a well-controlled high-throughput assay.*
671 International Journal of Epidemiology, 2009. **38**(6): p. 1725-1734.
672 60. Pavesi, E., et al., *Analysis of telomeres in peripheral blood cells from patients with bone marrow*
673 *failure.* Pediatric Blood & Cancer, 2009. **53**(3): p. 411-416.
674 61. Imam, T., et al., *Leukocyte Telomere Length in HIV-Infected Pregnant Women Treated With*
675 *Antiretroviral Drugs During Pregnancy and Their Uninfected Infants*. JAIDS Journal of Acquired
676 Immune Deficiency Syndromes, 2012. **60**(5): p. 495-502.

677 62. Zanet, D.L., et al., *Blood and Dried Blood Spot Telomere Length Measurement by qPCR: Assay*
678 *Considerations.* PLOS ONE, 2013. **8**(2): p. e57787.
679 63. Jodczyk, S., et al., *Telomere length measurement on the Roche LightCycler 480 Platform.* Genet
680 Test Mol Biomarkers, 2015. **19**(2): p. 63-8.
681 64. Panero, J., et al., *Absolute qPCR for measuring telomere length in bone marrow samples of*
682 *plasma cell disorders.* Molecular biotechnology, 2015. **57**(2): p. 155-159.
683 65. Hsieh, A., et al., *Optimization of a relative telomere length assay by monochromatic multiplex*
684 *real-time quantitative PCR on the LightCycler 480: sources of variability and quality control*
685 *considerations.* The Journal of Molecular Diagnostics, 2016. **18**(3): p. 425-437.
686 66. Lynch, S.M., et al., *Race, Ethnicity, Psychosocial Factors, and Telomere Length in a Multicenter*
687 *Setting.* PLOS ONE, 2016. **11**(1): p. e0146723.
688 67. Wand, T., et al., *Telomere content measurement in human hematopoietic cells: Comparative*
689 *analysis of qPCR and Flow-FISH techniques.* Cytometry A, 2016. **89**(10): p. 914-921.
690 68. Behrens, Y., et al., *Comparison of different methods for telomere length measurement in whole*
691 *blood and blood cell subsets: recommendations for telomere length measurement in*
692 *hematological diseases.* Genes, Chromosomes and Cancer, 2017. **56**(9): p. 700-708.
693 69. Khincha, P., et al., *Correlation of leukocyte telomere length measurement methods in patients*
694 *with dyskeratosis congenita and in their unaffected relatives.* International journal of molecular
695 sciences, 2017. **18**(8): p. 1765.
696 70. Lee, M., et al., *Comparative analysis of whole genome sequencing-based telomere length*
697 *measurement techniques.* Methods, 2017. **114**: p. 4-15.
698 71. Tarik, M., et al., *Validation of quantitative polymerase chain reaction with Southern blot method*
699 *for telomere length analysis.* Future Science OA, 2018. **4**(4): p. FSO282.
700 72. Wang, Y., et al., *Telomere Length Calibration from qPCR Measurement: Limitations of Current*
701 *Method.* Cells, 2018. **7**(11): p. 183.
702 73. Ferreira, M., et al., *Comparison of flow-FISH and MM-qPCR telomere length assessment*
703 *techniques for the screening of telomeropathies.* Annals of the New York Academy of Sciences,
704 2019: p. 1-11.
705 74. Ropio, J., et al., *Reliable blood cancer cells' telomere length evaluation by qPCR.* Cancer Med,
706 2020. **9**(9): p. 3153-3162.
707 75. Ropio, J., et al., *Reliable blood cancer cells' telomere length evaluation by qPCR.* Cancer
708 Medicine, 2020.
709 76. Verhulst, S., *Improving comparability between qPCR-based telomere studies.* Molecular Ecology
710 Resources, 2020. **20**(1): p. 11-13.
711 77. Verhulst, S., et al., *Commentary: The reliability of telomere length measurements.* International
712 Journal of Epidemiology, 2015. **44**(5): p. 1683-1686.
713 78. Benetos, A., et al., *Response by Benetos et al to letter regarding article,"Short leukocyte*
714 *telomere length precedes clinical expression of atherosclerosis: the blood-and-muscle model".*
715 Circulation research, 2018. **122**(8): p. e73-e74.
716 79. Benetos, A., et al., *Telomere length tracking in children and their parents: implications for adult*
717 *onset diseases.* Faseb j, 2019. **33**(12): p. 14248-14253.
718 80. Dlouha, D., et al., *Comparison of the relative telomere length measured in leukocytes and eleven*
719 *different human tissues.* Physiol Res, 2014. **63 Suppl 3**: p. S343-50.
720 81. Dagnall, C., et al., *Effect of pre-analytic variables on the reproducibility of qPCR relative telomere*
721 *length measurement.* PLOS ONE, 2017. **12**(9): p. e0184098.
722 82. Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)—toward*
723 *standards for microarray data.* Nature genetics, 2001. **29**(4): p. 365-371.

724 83. Taylor, C., et al., *The minimum information about a proteomics experiment (MIAPE).* Nature
725 biotechnology, 2007. **25**(8): p. 887-893.
726 84. Field, D., et al., *The minimum information about a genome sequence (MIGS) specification.*
727 Nature biotechnology, 2008. **26**(5): p. 541-547.
728 85. Taylor, C., et al., *Promoting coherent minimum reporting guidelines for biological and biomedical*
729 *investigations: the MIBBI project.* Nature biotechnology, 2008. **26**(8): p. 889-896.
730 86. Fiehn, O., et al., *The metabolomics standards initiative (MSI).* Metabolomics, 2007. **3**(3): p. 175-
731 178.
732 87. Brazma, A., *Minimum information about a microarray experiment (MIAME)–successes, failures,*
733 *challenges.* TheScientificWorldJournal, 2009. **9**.
734 88. Kimura, M., et al., *Measurement of telomere length by the Southern blot analysis of terminal*
735 *restriction fragment lengths.* Nature protocols, 2010. **5**(9): p. 1596-1607.

736