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, UA

[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 20 articles were found that matched the inclusion criteria. Papers were reviewed for quality of methodologic reporting of sample and DNA quality, PCR assay characteristics, and analytic approaches to determine final TL was low as assessed by two different recommended reporting guidelines. Additionally, few papers reported whether samples were blinded for analysis. Reported correlation between methods (as assessed by Pearson's r) varied, as did reported coefficients of variation or intra-class correlation coefficient. The sample size for nearly all studies was less than 100, raising significant 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. To assist future investigators, we present recommendations for reporting guidelines of PCR-based TL measurement methods and present data to demonstrate the impact of assay variability on statistical power. Future cross-laboratory studies with rigorous methodologic and statistical reporting, adequate sample size, and blinding are essential to determine the true relation and variability of various TL measurement methods and are a key directive to the newly initiated Telomere Research Network funded jointly by then NIA and NIEHS.

21 # Introduction:

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

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

41 Currently, over two dozen assays have been developed to measure TL in human studies (Fig 2)
42 [4-33]. These assays are classifiable into four broad categories: hybridization-based, PCR-based,
43 sequence-based, and mixed methods (e.g. hybridization/PCR combination). These assays vary in which

component of TL is measured. While most focus on the measurement of the average TL within the sample, assays also measure chromosome-specific TL, the complete distribution of TL in a cell population[33], or the shortest TL. 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 [34, 35]. Several recent reviews have discussed the overall advantages and disadvantages of each method [26, 36] 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 coefficient of variation (CV) compared to intraclass correlation coefficients (ICC) [26, 37, 38]. While studies of basic telomere biology continue to explore the complex role that telomeres play in cellular and organismal function, studies testing TL as markers of disease risk or environmental exposure must balance biological relevance, methodologic precision, and experimental practicality, similar to other epigenetic markers such as DNA methylation[39].

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

Over the last decade, debates have arisen over the utility and measurement of TL. This debate is partially fueled by concerns related to the reproducibility and replicability of TL measurements across studies, methods, and laboratories, particularly as new methods were developed and more laboratories began performing these assays. 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) [40, 41]. Others tested the relative correlation of the TL measured by different assays with an expected phenotype (e.g. aging, parent-offspring correlation) [42], or examined the relative ability of different assays of TL to predict a specific disease or health outcome[43, 44]. Each of these approaches requires a

68 different analytic strategy and study design. To date, the existing evidence remains insufficient to
69 answer key methodologic questions related to differences in reproducibility and replicability across
70 measurement assays and how/whether these differences affect the ability of TL to serve as a biological
71 indicator of exposure or a predictor of disease or health risk [45]. Beyond these concerns, there remains
72 a lack of consensus as to which, if any, methodology is the "gold standard," as even the classic Southern
73 blot method is challenged by its inability to capture potentially critical metrics (e.g. full distribution,
74 shortest telomere length, inclusion of the subtelomeric region). Before applying TL as a biomarker of
75 environmental exposure and/or a predictor of a disease, measured by any method that is applicable to
76 human population studies or ecological studies, it is critical to systematically test fundamental issues
77 related to assay reproducibility and replicability [46].

78 As part of a joint National Institute of Aging and National Institute of Environmental Health
79 Science initiative that funded a U24 cooperative award and four separate U01 awards, a Telomere
80 Research Network (TRN) was established in 2019 (trn.tulane.edu). This network is facilitating
81 interdisciplinary collaborations and dialogues across all scientific fields incorporating telomere research
82 in relation to environmental exposures, psychosocial stress, and disease susceptibility in a variety of
83 health-related research studies. The TRN is coordinating cross-method comparison studies with the
84 long-term goal of developing methodological guidelines and recommendations for telomere research
85 applicable to population-based studies. As a first step towards the goals of this network, we undertook
86 a systematic literature review of published studies that directly compared TL measured using at least
87 one PCR-based method and another approach to determine how these studies might inform the field,
88 with particular attention to assay precision and accuracy of different measurement assays and what
89 research gaps remain.

90 As defined by the Committee on Reproducibility and Replicability in Science, precision is the
91 closeness of agreement between measured quantities obtained by replicate measurements, while

accuracy is the closeness of agreement between a measured quantity and a true value [46]. Reproducibility is defined as precision in measurement under conditions that involve different locations or different measurement procedures, while repeatability is defined as precision in measurements that include the same procedures/locations. Beginning from this perspective, this systematic review evaluated the existing literature related to cross-method comparisons. This review focuses on PCR-based methodologies due to their increasing use in population-based studies, their central role in the debate related to assay precision, and the existence of two reporting guidelines - one created through the TRN, and a second one created by a separate group in a recently published manuscript [47]. The majority of PCR-based methods are derived from two seminal methodologic papers by Richard Cawthon, the first describing a monoplex based assay (qPCR) and the second describing a multiplex assay (MMqPCR) [13, 18]. Our review focuses specifically on the comprehensiveness of methodologic reporting, correlation between TL measured by different assays, assay repeatability and reproducibility, and overall scientific design of methodological comparisons. Finally, we suggest areas of needed scientific examination and provide some guidance related to study design, necessary sample size, and analytic approach, to address key remaining questions: (1) What is known about the relationship between TL measured using PCR-based methods and other assays? (2) What is known about the reproducibility and repeatability of PCR-based methods and how does this relate to other TL measurement techniques? (3) Given variable methodologic precision, what are the implications for 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

115 ## Manuscript search

116 To identify relevant papers that reported on cross method comparisons of any qPCR-based method
117 (qPCR, absolute TL (aTL), and MMqPCR) and another method of TL assessment (PCR-based or otherwise)
118 or the same PCR-based method conducted in separate laboratories, we conducted a critical review
119 beginning with a literature search (Fig 3). The following key terms "telomere," "telomere length," and
120 "human" were searched in PubMed. From these initial results, a second search included the keyword
121 "PCR" to identify the initial titles for screening. Search criteria included papers published since January
122 1, 2002 (the year the first method to measure TL by qPCR was published) through May 10, 2020. The
123 references of selected papers were also reviewed to identify any additional papers. A list of identified
124 papers was presented to the TRN Steering Committee, who also suggested additional papers. Initial
125 review of papers for inclusion was accomplished through evaluation of both the abstract and methods
126 section, as some manuscripts were not directly focused on methodological comparison and instead only
127 reported the cross-method comparison on a subset of samples. Final inclusion in this review met the
128 following baseline requirements:

129 1. Article published in a peer-reviewed journal (abstracts and pre-prints not included).
130 2. Article was not presenting the initial development of a new method or a substantial refinement
131 of an existing methodology. This type of study was excluded due to the intrinsic bias in reporting
132 novel method development.
133 3. Included a direct comparison of TL using the same biological measured with two distinct TL
134 assays or in two separate laboratories.
135 4. At least one of the methods used to measure TL was PCR-based.

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

139 Quality review:

140 Included papers were evaluated for quality of methodologic reporting using two different
141 indices of reporting guidelines for PCR-based telomere studies. The first was created through consensus
142 of the initial participants in the TRN (Supplemental Table 1). The second was derived from
143 recommendations published by Morinha et al. 2020 (Supplemental Table 2) [47]. We included both
144 guidelines for two reasons. First, there is not empirical data to distinguish between the two guidelines in
145 terms of ensuring rigor and reproducibility for the field. Second, as several of the papers reviewed were
146 authored by participants involved in the creation of the TRN guidelines, the inclusion of both guidelines
147 provided some degree of impartiality. Both guidelines contain overlap with the MIQE guidelines and
148 include characterization of the importance of each recommendation [48]. In terms of specific
149 differences, the Morinha recommendations included several pre-analytic considerations not included in
150 the TRN guidelines (e.g. volume of sample processed, robotic instrumentation vs manual), while storage
151 buffer and the percentage of samples tested for DNA integrity were included in the TRN guidelines but
152 not the Morinha. The Morinha guidelines also required greater detail for qPCR validation such as the
153 standard curve and calibration samples, as well as a requirement for the melt curve and Ct of the
154 negative control, which were not included in the TRN guidelines. Both guidelines assess the
155 comprehensiveness of the information describing the PCR assay itself as well as analytic considerations
156 for final TL determination. A grading rubric for each set of guidelines was developed to reduce
157 subjective reviewer interpretation (Supplemental Table 3). A composite assessment for each index was
158 divided into three subcategories for the TRN guidelines and five subcategories for the Morinha

159 guidelines. These broadly encompassed sample collection and processing, DNA quality metrics and
160 storage; PCR assay components and quality control; and data analysis. Two of three reviewers
161 independently assessed each article for fulfillment of reporting guidelines (ARL, LWYM, and SSD). The
162 scores for each individual item were compared and discrepancies resolved by the third reviewer.

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

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

178 ICC calculation:

179 Given established analytic shortcomings related to the use of the CV as a metric of testing the
180 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 [38, 49, 50]. ICCs for one study were also calculated using a two-way, single measurement, absolute agreement, random effects model, known as ICC(A,1) and for average measurements ICC(A,k) in McGraw & Wong's (1996) terminology [51].

To provide guidance for future study design, we present several different power analyses outlining the relation between sample size, ICC, and ability to detect group differences. For cross sectional studies, calculations assumed a realistic (true) standard deviation of 650 base pairs (bp). Power analysis was done using G*Power[52]. N is the combined n of the two groups to be compared and was assumed to be equally distributed among the two groups. To examine the impact of variation in ICC on longitudinal TL studies, the statistical power to detect a significant change (paired-t-test) in telomere length of 25 bp/year for sample sizes of 25, 50 and 100 individuals, and an interval of 8 years between baseline and follow-up (i.e. on average 200 bp in total), as a function of measurement repeatability (e.g. reliability) expressed as the ICC (A,1). Population SD of telomere length was assumed to be 650 bp at both time points and telomere shortening was simulated assuming a Poisson distribution with mean/variance of 25bp/year, which is within the range of observed shortening in adults, although the true clinical or phenotypic implications of specific differences in bps has yet to be empirically determined.

Results

The initial search revealed 5427 articles and the inclusion of "PCR" as a search term limited the results to 767 articles, whose abstracts and methods were read (Figure 3). An additional 25 articles were identified through assessment of the references of method validation papers and other included cross-method validation studies. A review of these 792 abstracts identified 65 articles for assessment of the full text and supplemental information to determine inclusion in this review. 26 articles were

204 determined to be novel method validation and excluded. 18 articles were excluded as they were either
205 reviews or did not include direct method comparisons. One article was excluded due to the
206 determination that the DNA samples used for cross-method comparison were obtained at different time
207 points. We also included 6 papers that, while not specifically designed as a cross method comparison,
208 included sufficient details comparing TL measurement using different assays. This resulted in a total of
209 20 articles included in this review (Table 1).

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

Study	Year	Methods	Multi-lab?	Sample Type	Sample Size	Age Range (years)	TL Mean ± SD
Ehrlenbach*†[53]	2009	qPCR x TRF		blood	56	51-81	T/S = 1.31 (0.95-1.90) (n=669)
Pavesi*[54]	2009	qPCR x TRF		blood, cord blood	28	0-97	NR
Aviv[40]	2011	qPCR x TRF	Yes	blood	50	41-70	NR
Imam*[55]	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[56]	2013	qPCR x MMqPCR		blood	32	2-59	NR
Gutierrez-Rodrigues[44]	2014	qPCR x FlowFISH x TRF		blood, cord blood	70, 51	0-88	NR
Martin-Ruiz[41]	2015	qPCR x MMqPCR x TRF x STELA	Yes	cell lines, placenta	12	N/A	NR
Eisenberg[42]	2015	MMqPCR x TRF	Yes	blood	190	NR	NR
Jodcyzyk*[57]	2015	qPCR x TRF		blood	20	0-35/50	T/S = 1.184 ± 0.371 (n=677); 1.104 ± 0.153 (n=351)
Panero*[58]	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[59]	2016	qpCR x MMqPCR x TRF x flowFISH		blood, cord blood, DBS, placenta, cell lines	33 - 84	NR	NR
Wand[60]	2016	qPCR x TRF x FlowFISH		blood	11	24-69	NR
Behrens[61]	2017	qPCR x FlowFISH x T/C-FISH x TRF		blood, cord blood	83 - 99	0-81	NR
Khincha†[62]	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[63]	2017	qPCR x TRF x WGS		cell lines	20	N/A	NR
Tarik[64]	2018	MMqPCR x TRF		blood	94	NR	T/S = 1.02 ± 0.32
Wang[65]	2018	qPCR x FlowFISH		blood	181	19-53	NR
Gadalla*[43]	2018	qPCR x qPCR	Yes	blood	100	NR	NR
Ventura Ferreira[66]	2019	MMqPCR x FlowFISH		blood	105	24-66	NR
Ropio[67]	2020	qPCR x TRF x aTL		cell lines	20	N/A	4.32 ± 0.143 kb

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

219 Paper Characteristics

220 The most common methods comparison among the 20 papers evaluated in this review was
221 monoplex quantitative PCR (qPCR) and the telomere restriction fragment (TRF) method by Southern blot
222 (n=12). Four studies compared multiplex qPCR (MMqPCR) with TRF. Seven studies compared qPCR with
223 the flow-FISH method, and two studies compared MMqPCR with flow-FISH. Three studies involved
224 comparisons of MMqPCR and qPCR, and two studies involved direct comparisons of qPCR methods or
225 separate laboratories. Other methods examined included STELA (compared with qPCR, MMqPCR, and
226 TRF in one study) [41], and absolute TL (two studies, one comparing aTL to TRF and one comparing to
227 aTL to qPCR) [58, 68]. T/C-FISH was also examined in one study as it related to qPCR [61]. Additionally,
228 one paper compared the correlation of several whole genome sequencing (WGS) platforms to qPCR-
229 based measurement[63]. Note some studies compared more than two methods [41, 44, 59-63, 65, 68].

230 Whole blood was the most common sample type used (n=12), but cord blood (n=5), peripheral
231 blood mononuclear cells (PBMCs) (n=4), and cell lines (n=4) were also utilized as well as a range of other
232 sample types. Several studies reported on more than one sample type.

233 The reported sample size for the cross-method comparisons ranged from 12 to 181 and only 7
234 papers had a total sample size greater than 100. Five studies reported the means and standard

235 deviations and two reported the median and range of TL measurements for the study. Two studies
236 provided the raw values of the PCR-based TL measurements.

237 Overall quality of reporting of PCR assay methodology

238 Of the 20 studies included in this systematic review, the average completion score across both
239 reporting guidelines was 54%, with an average of 55% for the TRN guidelines and 52% for the Morinha
240 guidelines (Table 2). Overall, papers included between 26-75% and 29-78% of the recommended
241 reporting metrics for the TRN and Morinha guidelines, respectively. Some metrics were consistently
242 reported in nearly all papers, including the sample type, single copy gene name, and type of PCR method
243 utilized. However, only 10% of the included papers reported on sample storage, PCR efficiencies, or the
244 number of samples excluded due to quality concerns with the assay.

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

	Category	**Average Score**
TRN Reporting Guidelines	Sample	41%
	qPCR assay	77%
	Data analysis	42%
	Total	**55%**
Morinha Reporting Guidelines	Sample	47%
	DNA extraction	34%
	qPCR validation	15%
	qPCR protocol	66%
	Data analysis	70%
	Total	**52%**
	Average Total Score	**54%**

246

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

249 DNA

250 For both the TRN guidelines and Morinha guidelines, reporting of sample type, storage, DNA extraction,
251 and DNA quality/integrity was poor, with an average of 41% for the TRN Sample/DNA category, 47% for
252 the Morinha Sample category, and 34% for the DNA category of the Morinha guidelines. Storage
253 conditions for both the biological samples and extracted DNA were poorly reported, with 20% or less of
254 studies providing this information (Table 3). Less than half of the studies reported on metrics related to
255 DNA integrity.

256 **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	20
	Sample storage temperature	4
	Sample storage time before extraction	2
	Sample storage buffer	6
	DNA extraction method	16
	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	20
	PCR machine	16
	Source of master mix & reagents	17
	Final reaction volume	15
	Telomere primer sequences	18
	Telomere primer concentration	16
	Single copy gene name	20
	Single gene primer sequence	18
	Singe gene primer concentration	15
	Full PCR program	14

	PCR efficiency of both primers	2
	Source of control samples	11
	Concentration of DNA standard	9
Data analysis	Mean and standard deviation/median and range of TL	10
	Number of sample replicates	19
	Level of independence of replicates	9
	Analytic method to determine final TL	18
	Method of accounting for variation between replicates	11
	Method of accounting for well position effects	4
	Method of accounting for between-plate effects	12
	% of samples repeated as a result of failed QC	3
	% of samples excluded from analysis	2
	Acceptable range of PCR efficiency for primers	5
	ICCs of sample/study groups	2
	T/S ratio transformed to Z score prior to analysis	1

PCR assay

Reporting on PCR assay conditions and quality control varied. While many metrics of the PCR assay required in the TRN guidelines were well-reported, only 14 of 20 studies reported the full cycling conditions. The lowest reporting metric related to PCR efficiencies with only 10% reporting on actual PCR efficiencies. Additionally, just over half (55%) 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 three 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 9 studies reported on the level of independence of sample replicates (e.g. replicates run on the same plate or on different

271 plates/different times) and only half of the studies reported the means and standard deviations (or
272 median and range) of the T/S ratio.

273 Cross-laboratory studies

274 Only four studies compared analyses across more than one laboratory (Table 1) [40-43]. Of
275 these studies, three described how samples were blinded before analyses. Further, of these cross-
276 laboratory studies, only two studies included the same assay performed in different laboratories [41,
277 43].

278 Reproducibility

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

284 Relative reproducibility

285 The current literature does not provide sufficient data to address the relative reproducibility,
286 the relation of measurements using the same assay in different laboratories as, to date, only two studies
287 have tested this directly. In one study that blinded comparison samples before they were sent to the
288 external laboratories, the median CV across laboratories for qPCR was 18.3%, while the median CV for
289 STELA/TRF based TL measurement was 9.2%. However given the dependence of CVs on the y-intercept,
290 the interpretation of these CVs remains challenging [38]. In the other study, where samples were not
291 blinded before being assayed, the reported within lab CVs for replicate measurements were 2.5% and

292 8.6%. As the laboratories involved utilized different PCR primers and slightly different methods, it was
293 not possible to specifically evaluate relative reproducibility.

294 Reproducibility precision

295 Reproducibility precision, e.g. the closeness of two or more measurements using different
296 techniques, was addressed to some extent in 19 of the 20 studies reviewed. However, only four studies
297 involved assays performed in different laboratories. The correlation of TL measurement with qPCR-
298 based assays to other assays was, for the most part, reported as linear regression and correlations
299 coefficients (Figure 4).

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

304 The widest variability in correlation was found with qPCR and Flow-FISH or TRF. Less variation in
305 the correlation between measurement was found, in general, with MMQPCR and other measures,
306 however there are, to date, fewer laboratories and studies using this methodology. No studies have
307 compared MMQPCR to aTL, or whole genome sequencing (WGS). Only one paper each compared qPCR
308 and qPCR (in separate labs), aTL and TRF, qPCR and aTL, and qPCR with WGS. To our knowledge no
309 studies have compared WGS data with aTL, although studies have compared TRF and WGS [23].

310 In five of the papers in this review, linear regression was used to extrapolate TL into kilobases
311 (kb) from the T/S ratio using TRF values. One paper converted T/S ratio to kb before analysis of the
312 correlation between methods [44]. In two cases, T/S ratio were converted to kb prior to TL comparison
313 utilizing Bland-Altman analysis [61, 66]. In two of the five papers, the conversion of the T/S ratio to bp

314 was based on analyses extrapolated from different data or measured on a different sample type, raising
315 substantial concerns on the true measurement with uncertain implications for the r value [44, 66].
316 Beyond concerns related to the source of the data utilized for conversion from T/S to bp before
317 comparison across methods, this analytic approach likely to leads to inaccurate reporting [44]. Only one
318 study utilized the TRN-recommended procedure of transformation to z-score before comparison [41].

319 Repeatability

320 Repeatability, the precision in measurements that include the same procedure/locations,
321 revealed the greatest variation in both lab and assay specific precision and between methods. In these
322 studies, the number of replicates for a specific DNA sample ranged from 1 replicate (e.g. sample
323 analyzed twice) to five replicates (each sample analyzed six different times). Additionally, only three
324 reviewed studies reported the number of samples that were repeated due to within replicate variance,
325 despite clear acknowledgement in the field that a proportion of all studies will ultimately require
326 repeated assays of telomere length as a result of between-replicate variance. While of limited utility in
327 confirming precision, the intra-assay reported CVs for PCR-based methods (qPCR, MMqPCR) in this
328 review ranged from 2.5 to 12%, and the inter-assay CVs ranged from 3.97 to 15.9%. Inter-assay CVs for
329 TRF ranged from 1.68 to 6.3%, with intra-assay CV reported in only one paper as 1.20%. Inter-assay CVs
330 for flow-FISH were reported as 9.3% and 10.8% in two papers, with only one reporting an intra-assay CV
331 of 9.6%. One paper examining the aTL assay reported its inter-assay CV as 6.7% and intra-assay CV as
332 2.5%. We emphasize, however, that CVs have several flaws which make comparisons of them across
333 laboratories and methods particularly problematic [38, 50].

334 Only two studies utilized ICC analyses to examine the repeatability of replicates, reporting ICCs
335 of 0.89 and 0.92 for PCR-based measurement [62, 64]. To expand data on the repeatability of PCR-based
336 and other TL methods, raw data was obtained from authors of a subset of these papers, and ICCs

337 independently calculated. Calculated/reported ICC for TRF methods ranged from 0.92- 0.99 in the
338 studies included in this review and are consistent with the ICCs reported in existing studies utilizing the
339 TRF (0.95 to 0.99). However, it is of note that these ICCs were almost entirely the result of TRF
340 measurement in one laboratory. The ICCs for qPCR-based methods in reviewed papers ranged from
341 0.89-0.92, including the two reported in manuscripts and an additional ICC calculated from raw data (ICC
342 = 0.915, SE = 0.023, 95% confidence interval: [0.860, 0. 946], P<0.001; reported CV 6.5%) [38, 69]. ICCs
343 for MMqPCR (triplicates on the same plate) from one study were run separately based on year of
344 analyses. In one set (n=873) run across different PCR plates in initial and duplicate runs, the ICCs were
345 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-
346 run due to initially high intra-assay CVs, this is likely an under-estimate of the true ICC value. For these
347 same samples, TRF ICCs were calculated from duplicate gels on a subset (n=159) and the inter-gel
348 ICC(A,1) = 0.92 (95% CI 0.89-0.95) and ICC(A,k)=0.96 (95% CI 0.94-0.97). However, we note that these
349 TRF analyses were conducted by a trainee which likely decreased repeatability compared to what is
350 typical of experienced technicians. Given the significant variation in methodologic and raw data
351 reporting, and the wide variability in published CVs, it is likely that the majority of existing TL studies not
352 specifically comparing methods would have significantly lower ICCs.

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

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

359 In Figure 5A, we present the sample size required to test effect sizes of 150, 200, and 300 bp
360 with a t-test with a power of 0.9, as a function of measurement error as expressed in the ICC. To
361 contextualize the differences: 150 bp is the approximate difference found between the sexes, and 300
362 bp is the approximate difference observed between individuals with and without atherosclerotic
363 cardiovascular disease [70]. As directly converting bp to T/S ratios is not feasible in this analysis, we
364 suggest that investigators using PCR-based methods conservatively estimate their potential bp
365 difference from the existing literature related to their exposure or outcome of interest when considering
366 study design and sample size. Calculations assumed a realistic (true) standard deviation of 650 bp and
367 power analysis was done using G*Power [52]. N is the combined n of the two groups to be compared
368 and was assumed to be equally distributed among the two groups.

369 **Fig 5A. ICC effect on statistical power (cross-sectional studies).** The sample size required to test effect
370 sizes of 150, 200 and 300 bp with a t-test with a power of 0.9, as a function of measurement error as
371 expressed in the ICC (intraclass correlation coefficient). Calculations assumed a realistic (true) standard
372 deviation of 650 bp and power analysis was done using G*Power. N is the combined n of the two groups
373 to be compared and was assumed to be equally distributed among the two groups.

374 Fig. 5B and 5C present the power of different sample sizes to detect a 200bp difference at
375 various ICC levels with a presumed mean TL of 650bp in a longitudinal study. For small sample sizes, e.g.
376 25 subjects, the impact of even moderate variability in ICCs on power is substantial. This calculation was
377 based on the average loss of 25 base pairs per year in adults. The rate of base pair loss in infants and
378 children is likely significantly different and, but as of yet is poorly characterized.

379 **Fig 5B, C. ICC effect on statistical power (longitudinal studies).** Statistical power to detect a significant
380 change (paired-t-test) in telomere length of 25 bp/yearfor sample sizes of 25, 50 and 100 individuals,
381 and with an interval of 4 (B) or 8 years (C) between baseline and follow-up(i.e. respectively on average

382 100 and 200 bp in total), as a function of measurement reliability expressed as the Intraclass Correlation
383 Coefficient. Population SD of true telomere length (i.e. without measurement error) was transformed to
384 be 0.65 kb at both time points and telomere shortening was simulated assuming a Poisson distribution
385 with mean / variance of 25.

386 # Discussion:

387 This systematic review found a total of 20 papers documenting comparison between TL
388 measured using a PCR based methodology and another TL assay. Until recently, there were no
389 publication reporting guidelines for PCR TL measurement. Our review specifically focused on method
390 comparison studies with the expectation that reporting of critical assay parameters would be expected
391 to be higher than in typical studies. Our analyses, using two separately developed reporting guidelines,
392 found that on average only half of the recommended factors were documented, indicating the need for
393 increased methodologic reporting. The lowest reporting was related to information about the validation
394 of PCR-based assays outlined in the Morinha guidelines, with only five papers including any of the
395 recommended factors. PCR efficiencies, a key reporting requirement in both guidelines and also
396 considered essential in the MIQE guidelines, was absent from the majority of papers with only five
397 mentioning the PCR efficiency parameters and only two documenting the actual PCR efficiencies. Given
398 that all PCR-based methods either assume or specifically calculate the PCR efficiency when determining
399 the T/S ratio, the absence of this key metric is concerning, particularly if there are significant differences
400 in PCR efficiencies between the single copy and the telomere primers. A surprising number of studies
401 failed to comment on pre-analytic factors, specifically sample storage time and conditions, freeze-thaw
402 cycles, and evaluation of DNA quality and integrity, all potential sources of assay variability for both PCR
403 and non-PCR based TL assays that may contribute to current debates in the field about the utility of TL
404 [67, 71, 72]. Lastly, the reporting of the number of samples failing initial quality control, repeated, or

405 unable to be assayed was low. In laboratories routinely performing TL measurement using any assay, a
406 certain percentage of samples for each study will require repeating and regularly a small subset may be
407 unanalyzable for various reasons. While it is possible that these factors were considered and monitored,
408 the lack of reporting for this metric heightens the need for increased attention to the proposed
409 reporting recommendations. Moving forward, the widespread dissemination of these qPCR reporting
410 guidelines to study sections, peer reviewers, and scientists represents an important next step in
411 enhancing the scientific rigor of the field.

412 At this time, evaluation of the existing literature fails to provide sufficient evidence of the
413 relative or precision reproducibility of different TL assays. Our review identified only four studies that
414 included cross laboratory comparisons and, of these, only two evaluated PCR-based assays performed in
415 more than one lab. As the number of laboratories performing TL studies using PCR and other methods
416 continue to increase, the lack of clear data about cross-laboratory reproducibility and the absence of
417 existing DNA standards or other methods to account for cross-laboratory variation substantially limits
418 the ability to characterize relative reproducibility. In terms of precision reproducibility across different
419 methods (e.g. PCR and TRF, or PCR and FISH), the current variability in findings, particularly when
420 coupled with limited methodologic reporting, highlights the need for additional rigorous and blinded
421 cross laboratory studies that are adequately powered to accurately determine how TL in a population
422 measured using different assays truly relates. Although eleven studies evaluated the relationship
423 between qPCR and TRF, due to the wide range in reported correlations between TL measurements, the
424 relatively small samples sizes, and the insufficient analytic and assay blinding, there is currently
425 insufficient data to draw firm conclusions on the general correlation between TL measured with
426 different assays. The analytic consequences of using CVs to test the relationship between TL measured
427 using different assays has been discussed previously, as has the issues caused by the use of analytic
428 strategies such as conversion to base pairs instead of z scores, especially when extrapolating from data

429 not produced in the same laboratory or with the same samples [38, 49]. In this review, we utilized
430 existing raw data from 3 included studies to provide preliminary data about precision reproducibility for
431 PCR and TRF studies. The wide range of ICCs calculated from these few studies, particularly for PCR-
432 based methods, and the low reporting of ICCs in the papers included in this review highlights the need
433 to increase attention to this specific reporting requirement. For many of the existing studies, the small
434 sample size and the lack of reporting of the means and standard deviations of TL prevents objective
435 determination of whether any of the current studies were adequately powered. Beyond these concerns,
436 as many studies utilized data, particularly for TRF, from a small number of specialized laboratories, the
437 applicability of much of the existing data to the wider telomere field is uncertain. For aTL and MMqPCR
438 based TL measurements, the current paucity of published cross method comparisons limits the ability to
439 form an opinion of how TL measured with these assays relates to other methods.

440 Measurement precision is critical. Methodologies that are low cost, practical, and simple to
441 implement with standard laboratory equipment, especially when they are innovative or high impact, are
442 often rapidly implemented across laboratories with various levels of expertise in the new methodology.
443 Invariably this results in diverse protocols, analyses, and methodologic reporting, consequences that are
444 even more problematic when there is an absence of consensus on best practices[48]. As with many
445 other biologic assays, the development of reporting guidelines for TL measurement has lagged behind
446 the broad implementation of the methods themselves [73-77]. The lack of consolidated guidance about
447 factors, both pre-analytic and within the assay itself, that contribute to measurement error when
448 combined with the wide popularity of PCR-based TL measurement undoubtedly contributed to
449 discrepancies in the existing literature and failed study replications. Similar to the MIQE guidelines, the
450 reporting guidelines presented and tested in this systematic review for PCR-based TL assays are meant
451 as minimal reporting recommendations focused on enhancing the reliability of results, consistency
452 between different laboratories performing the same assay, and increased experimental transparency

453 and accuracy [48]. To assist investigators and reviewers we highlight the overlapping recommendations
454 with the MIQE guidelines, indicate whether a particular requirement is desirable or essential, and
455 provide references that support the selection of the particular reporting requirement. Over the course
456 of the next four years, the TRN expects to develop similar reporting recommendations for other types of
457 TL assays while conducting adequately powered and scientifically rigorous studies to support these
458 reporting guidelines, recognizing that individual recommendations have varying levels of initial empirical
459 support [78].

460 Despite the strengths of this review, there are several limitations. First, this review only focused
461 on assays applicable to population-based studies in humans. It does not address issues in other species
462 or assays that may have clinical utility but for which the requirements for sample types (e.g. fresh tissue
463 and/or live cells) or the cost/labor/expertise requirements (e.g. TESLA) limit utilization in population
464 based studies. As part of its scientific charge to establish telomeres as indicators of psychosocial and
465 environmental exposures and sentinels of disease, the TRN also expects to support both the adaptation
466 of existing TL assays not currently scalable to population studies as well as the discovery of new
467 methodologies. A second limitation is that we utilized reporting guidelines for qPCR-based assays only.
468 To date, specific protocol recommendations and reporting guidelines have not been published for other
469 TL assays (e.g. TRF, FISH) although detailed methodologic protocols do exist [79]. Finally, we note that
470 while this article primarily focused on precision and reproducibility, accuracy of measurement is also
471 critically important. Precisely inaccurate measures will be of no use to the scientific field. In the absence
472 of a clear gold standard measurement technique, accuracy is difficult to discern. In addition to validating
473 methods across laboratories, we suggest using measures known to be correlated with TL (e.g. age,
474 familial relatedness, baseline to follow up TL correlations) to better establish external validity.

475 **Conclusions**

476 After careful examination of the existing literature, it is apparent that rigorous cross laboratory
477 and methodological studies must be an immediate priority for the TRN and the field. To assist the field
478 moving forward, we include reporting guidelines for PCR-based TL assays and indicate specific scientific
479 papers that support these recommendations originally developed through consensus of the initial TRN
480 members and consultants. These guidelines do not outline a specific PCR methodology and, at this time,
481 we do not believe there is sufficient data to provide guidance on specific assay approaches or
482 components. Rather, these guidelines are provided to ensure reviewers and readers can adequately
483 assess the methodology and consider the implications of these factors for each study's findings. The
484 consistency in results across reporting guidelines (TRN, Morinha, MIQE) related to the integrity and
485 quality of both the initial biological sample and the DNA itself support the critical nature of this
486 reporting metric. In terms of assay reporting, increased attention of investigators and reviewers to
487 ensuring complete reporting of assay reagents and PCR efficiencies is also expected to enhance the rigor
488 of the field. TRN investigators are currently testing the impact of different pre-analytic factors, DNA
489 integrity, and PCR conditions to provide extant evidence of the importance of these parameters in
490 relation to precision and reproducibility. We recommend that studies be required to report ICCs in lieu
491 of CVs, as well as either the median or mean and standard deviation of TL. We also provide specific
492 guidance related to sample size and power that is contingent upon the ICC given the substantial impact
493 of differences in assay precision on the ability to determine true relationships and with the expectation
494 that this will be of use for investigators as they develop new research questions. It is important to
495 balance assay cost, in terms of both time and reagents, with the needed sample size and statistical
496 power. Moving forward, investigators should carefully consider study design from this perspective,
497 recognizing that there is currently no "ideal" approach. Telomere research offers significant potential
498 across a diverse range of scientific fields with potential mechanistic insight into overlapping biological
499 pathways contributing to many of the leading causes of morbidity and mortality. Ensuring the highest

500 scientific rigor and precision, through accurate methodological reporting and rigorous testing of the
501 factors that contribute to assay variability, are requisite steps to ensuring that potential is achieved.

502 Acknowledgements

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

507 # References

508 1. Olonikov, A., *A theory of marginotomy.* J Theor Biol, 1971. **41**: p. 181-190.
509 2. Blackburn, E. and J. Gall, *A tandemly repeated sequence at the termini of the extrachromosomal*
510 *ribosomal RNA genes in Tetrahymena.* Journal of molecular biology, 1978. **120**(1): p. 33-53.
511 3. Moyzis, R., et al., *A highly conserved repetitive DNA sequence,(TTAGGG) n, present at the*
512 *telomeres of human chromosomes.* Proceedings of the National Academy of Sciences, 1988.
513 **85**(18): p. 6622-6626.
514 4. Therkelsen, A., et al., *Staining of human telomeres with primed in situ labeling (PRINS).*
515 Cytogenet Cell Genet, 1995. **68**(1-2): p. 115-8.
516 5. Lansdorp, P., et al., *Heterogeneity in telomere length of human chromosomes.* Hum Mol Genet,
517 1996. **5**(5): p. 685-91.
518 6. Bryant, J., et al., *Measurement of telomeric DNA content in human tissues.* Biotechniques, 1997.
519 **23**(3): p. 476-8, 480, 482, passim.
520 7. de Pauw, E., et al., *Assessment of telomere length in hematopoietic interphase cells using in situ*
521 *hybridization and digital fluorescence microscopy.* Cytometry, 1998. **32**(3): p. 163-9.
522 8. Hultdin, M., et al., *Telomere analysis by fluorescence in situ hybridization and flow cytometry.*
523 Nucleic Acids Res, 1998. **26**(16): p. 3651-6.
524 9. Narath, R., et al., *Automatic telomere length measurements in interphase nuclei by IQ-FISH.*
525 Cytometry Part A: the Journal of the International Society for Analytical Cytology, 2005. **68**(2): p.
526 113-120.
527 10. Nakamura, Y., et al., *Simple, rapid, quantitative, and sensitive detection of telomere repeats in*
528 *cell lysate by a hybridization protection assay.* Clin Chem, 1999. **45**(10): p. 1718-24.
529 11. Law, H. and Y. Lau, *Validation and development of quantitative flow cytometry-based*
530 *fluorescence in situ hybridization for intercenter comparison of telomere length measurement.*
531 Cytometry: The Journal of the International Society for Analytical Cytology, 2001. **43**(2): p. 150-
532 153.
533 12. Rufer, N., et al., *Telomere length dynamics in human lymphocyte subpopulations measured by*
534 *flow cytometry.* Nat Biotechnol, 1998. **16**(8): p. 743-7.
535 13. Cawthon, R., *Telomere measurement by quatitative PCR.* Nucleic Acids Research, 2002. **30**(10):
536 p. e47- 53.
537 14. Meeker, A.K., et al., *Telomere length assessment in human archival tissues: combined telomere*
538 *fluorescence in situ hybridization and immunostaining.* The American Journal of Pathology, 2002.
539 **160**(4): p. 1259-1268.
540 15. Perner, S., et al., *Quantifying Telomere Lengths of Human Individual Chromosome Arms by*
541 *Centromere-Calibrated Fluorescence in Situ Hybridization and Digital Imaging.* The American
542 Journal of Pathology, 2003. **163**(5): p. 1751-1756.
543 16. Baird, D., et al., *Extensive allelic variation and ultrashort telomeres in senescent human cells.* Nat
544 Genet, 2003. **33**(2): p. 203-7.
545 17. Canela, A., et al., *High-throughput telomere length quantification by FISH and its application to*
546 *human population studies.* Proc Natl Acad Sci U S A, 2007. **104**(13): p. 5300-5.
547 18. Cawthon, R., *Telomere length measurement by a novel monochrome multiplex quantitative PCR*
548 *method.* Nucleic Acids Research, 2009. **37**: p. e21.
549 19. Bendix, L., et al., *The load of short telomeres, estimated by a new method, Universal STELA,*
550 *correlates with number of senescent cells.* Aging Cell, 2010. **9**(3): p. 383-97.

551 20. O'Callaghan, N. and M. Fenech, *A quantitative PCR method for measuring absolute telomere*
552 *length.* Biological procedures online, 2011. **13**: p. 3-3.
553 21. Kimura, M. and A. Aviv, *Measurement of telomere DNA content by dot blot analysis.* Nucleic
554 Acids Research, 2011. **39**(12): p. e84-e84.
555 22. Kibriya, M., et al., *Measurement of telomere length: a new assay using QuantiGene chemistry on*
556 *a Luminex platform*. 2014, AACR.
557 23. Ding, Z., et al., *Estimating telomere length from whole genome sequence data.* Nucleic Acids Res,
558 2014. **42**(9): p. e75.
559 24. Mender, I. and J. Shay, *Telomere Dysfunction Induced Foci (TIF) Analysis.* Bio-protocol, 2015.
560 **5**(22): p. e1656.
561 25. Nersisyan, L. and A. Arakelyan, *Computel: Computation of Mean Telomere Length from Whole-*
562 *Genome Next-Generation Sequencing Data.* PLOS ONE, 2015. **10**(4): p. e0125201.
563 26. Lai, T., et al., *A method for measuring the distribution of the shortest telomeres in cells and*
564 *tissues.* Nature communications, 2017. **8**(1): p. 1-14.
565 27. McCaffrey, J., et al., *High-throughput single-molecule telomere characterization.* Genome
566 research, 2017. **27**(11): p. 1904-1915.
567 28. Jasmine, F., et al., *A novel pooled-sample multiplex luminex assay for high-throughput*
568 *measurement of relative telomere length.* American Journal of Human Biology, 2018. **30**(4): p.
569 e23118.
570 29. Dahlgren, P., et al., *Development of a New Monochrome Multiplex qPCR Method for Relative*
571 *Telomere Length Measurement in Cancer.* Neoplasia, 2018. **20**(5): p. 425-431.
572 30. Farmery, J.H.R., et al., *Telomerecat: A ploidy-agnostic method for estimating telomere length*
573 *from whole genome sequencing data.* Scientific Reports, 2018. **8**(1): p. 1300.
574 31. Lu, A., et al., *DNA methylation-based estimator of telomere length.* Aging, 2019. **11**(16): p. 5895-
575 5923.
576 32. Feuerbach, L., et al., *TelomereHunter–in silico estimation of telomere content and composition*
577 *from cancer genomes.* BMC bioinformatics, 2019. **20**(1): p. 1-11.
578 33. de Pedro, N., et al., *Analytical Validation of Telomere Analysis Technology® for the High-*
579 *Throughput Analysis of Multiple Telomere-Associated Variables.* Biological Procedures Online,
580 2020. **22**(1): p. 2.
581 34. Hemann, M., et al., *The shortest telomere, not average telomere length, is critical for cell*
582 *viability and chromosome stability.* Cell, 2001. **107**(1): p. 67-77.
583 35. Xu, Z., et al., *The length of the shortest telomere as the major determinant of the onset of*
584 *replicative senescence.* Genetics, 2013. **194**(4): p. 847-857.
585 36. Montpetit, A., et al., *Telomere length: a review of methods for measurement.* Nursing research,
586 2014. **63**(4): p. 289.
587 37. Aubert, G., M. Hills, and P. Lansdorp, *Telomere length measurement-caveats and a critical*
588 *assessment of the available technologies and tools.* Mutat Res, 2012. **730**(1-2): p. 59-67.
589 38. Eisenberg, D., *Telomere length measurement validity: the coefficient of variation is invalid and*
590 *cannot be used to compare quantitative polymerase chain reaction and Southern blot telomere*
591 *length measurement techniques.* International Journal of Epidemiology, 2016. **45**(4): p. 1295-
592 1298.
593 39. Ehrlich, M., *DNA hypermethylation in disease: mechanisms and clinical relevance.* Epigenetics,
594 2019. **14**(12): p. 1141-1163.
595 40. Aviv, A., et al., *Impartial comparative analysis of measurement of leukocyte telomere*
596 *length/DNA content by Southern blots and qPCR.* Nucleic Acids Research, 2011. **39**(20): p. e134.
597 41. Martin-Ruiz, C., et al., *Reproducibility of telomere length assessment: an international*
598 *collaborative study.* International journal of epidemiology, 2015. **44**(5): p. 1673-1683.

599 42. Eisenberg, D., C. Kuzawa, and M.G. Hayes, *Improving qPCR telomere length assays: Controlling*
600 *for well position effects increases statistical power.* American Journal of Human Biology, 2015: p.
601 n/a-n/a.
602 43. Gadalla, S., et al., *No association between donor telomere length and outcomes after allogeneic*
603 *unrelated hematopoietic cell transplant in patients with acute leukemia.* Bone marrow
604 transplantation, 2018. **53**(4): p. 383-391.
605 44. Gutierrez-Rodrigues, F., et al., *Direct Comparison of Flow-FISH and qPCR as Diagnostic Tests for*
606 *Telomere Length Measurement in Humans.* PLoS ONE, 2014. **9**(11): p. e113747.
607 45. Bodelon, C., S. Savage, and S. Gadalla, *Telomeres in molecular epidemiology studies.* Prog Mol
608 Biol Transl Sci, 2014. **125**: p. 113-31.
609 46. Plant, A., R. Hanisch, and National Institute of Standards and Technology, *Reproducibility and*
610 *Replicability in Science, A Metrology Perspective* 2018: A Report to the National Academies of
611 Sciences, Engineering and Medicine Committee on Reproducibility and Replicability in Science.
612 47. Morinha, F., P. Magalhães, and G. Blanco, *Standard guidelines for the publication of telomere*
613 *qPCR results in evolutionary ecology.* Molecular Ecology Resources, 2020. **20**(3).
614 48. Bustin, S., et al., *The MIQE Guidelines: Minimum Information for Publication of Quantitative*
615 *Real-Time PCR Experiments.* Clinical Chemistry, 2009. **55**(4): p. 611-622.
616 49. Verhulst, S., et al., *Response to: Reliability and validity of telomere length measurements.*
617 International journal of epidemiology, 2016. **45**(4): p. 1298-1301.
618 50. Nettle, D., et al., *Consequences of measurement error in qPCR telomere data: A simulation study.*
619 PLoS One, 2019. **14**(5): p. e0216118.
620 51. McGraw, K. and S. Wong, *" Forming inferences about some intraclass correlations coefficients":*
621 *Correction.* 1996.
622 52. Faul, F., et al., *Statistical power analyses using G* Power 3.1: Tests for correlation and regression*
623 *analyses.* Behavior research methods, 2009. **41**(4): p. 1149-1160.
624 53. Ehrlenbach, S., et al., *Influences on the reduction of relative telomere length over 10 years in the*
625 *population-based Bruneck Study: introduction of a well-controlled high-throughput assay.*
626 International Journal of Epidemiology, 2009. **38**(6): p. 1725-1734.
627 54. Pavesi, E., et al., *Analysis of telomeres in peripheral blood cells from patients with bone marrow*
628 *failure.* Pediatric Blood & Cancer, 2009. **53**(3): p. 411-416.
629 55. Imam, T., et al., *Leukocyte Telomere Length in HIV-Infected Pregnant Women Treated With*
630 *Antiretroviral Drugs During Pregnancy and Their Uninfected Infants.* JAIDS Journal of Acquired
631 Immune Deficiency Syndromes, 2012. **60**(5): p. 495-502.
632 56. Zanet, D.L., et al., *Blood and Dried Blood Spot Telomere Length Measurement by qPCR: Assay*
633 *Considerations.* PLOS ONE, 2013. **8**(2): p. e57787.
634 57. Jodczyk, S., et al., *Telomere length measurement on the Roche LightCycler 480 Platform.* Genet
635 Test Mol Biomarkers, 2015. **19**(2): p. 63-8.
636 58. Panero, J., et al., *Absolute qPCR for measuring telomere length in bone marrow samples of*
637 *plasma cell disorders.* Molecular biotechnology, 2015. **57**(2): p. 155-159.
638 59. Hsieh, A., et al., *Optimization of a relative telomere length assay by monochromatic multiplex*
639 *real-time quantitative PCR on the LightCycler 480: sources of variability and quality control*
640 *considerations.* The Journal of Molecular Diagnostics, 2016. **18**(3): p. 425-437.
641 60. Wand, T., et al., *Telomere content measurement in human hematopoietic cells: Comparative*
642 *analysis of qPCR and Flow-FISH techniques.* Cytometry A, 2016. **89**(10): p. 914-921.
643 61. Behrens, Y., et al., *Comparison of different methods for telomere length measurement in whole*
644 *blood and blood cell subsets: recommendations for telomere length measurement in*
645 *hematological diseases.* Genes, Chromosomes and Cancer, 2017. **56**(9): p. 700-708.

646 62. Khincha, P., et al., *Correlation of leukocyte telomere length measurement methods in patients*
647 *with dyskeratosis congenita and in their unaffected relatives.* International journal of molecular
648 sciences, 2017. **18**(8): p. 1765.
649 63. Lee, M., et al., *Comparative analysis of whole genome sequencing-based telomere length*
650 *measurement techniques.* Methods, 2017. **114**: p. 4-15.
651 64. Tarik, M., et al., *Validation of quantitative polymerase chain reaction with Southern blot method*
652 *for telomere length analysis.* Future Science OA, 2018. **4**(4): p. FSO282.
653 65. Wang, Y., et al., *Telomere Length Calibration from qPCR Measurement: Limitations of Current*
654 *Method.* Cells, 2018. **7**(11): p. 183.
655 66. Ferreira, M., et al., *Comparison of flow-FISH and MM-qPCR telomere length assessment*
656 *techniques for the screening of telomeropathies.* Annals of the New York Academy of Sciences,
657 2019: p. 1-11.
658 67. Ropio, J., et al., *Reliable blood cancer cells' telomere length evaluation by qPCR.* Cancer Med,
659 2020. **9**(9): p. 3153-3162.
660 68. Ropio, J., et al., *Reliable blood cancer cells' telomere length evaluation by qPCR.* Cancer
661 Medicine, 2020.
662 69. Verhulst, S., et al., *Commentary: The reliability of telomere length measurements.* International
663 Journal of Epidemiology, 2015. **44**(5): p. 1683-1686.
664 70. Benetos, A., et al., *Response by Benetos et al to letter regarding article,"Short leukocyte*
665 *telomere length precedes clinical expression of atherosclerosis: the blood-and-muscle model".*
666 Circulation research, 2018. **122**(8): p. e73-e74.
667 71. Dlouha, D., et al., *Comparison of the relative telomere length measured in leukocytes and eleven*
668 *different human tissues.* Physiol Res, 2014. **63 Suppl 3**: p. S343-50.
669 72. Dagnall, C., et al., *Effect of pre-analytic variables on the reproducibility of qPCR relative telomere*
670 *length measurement.* PLOS ONE, 2017. **12**(9): p. e0184098.
671 73. Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)—toward*
672 *standards for microarray data.* Nature genetics, 2001. **29**(4): p. 365-371.
673 74. Taylor, C., et al., *The minimum information about a proteomics experiment (MIAPE).* Nature
674 biotechnology, 2007. **25**(8): p. 887-893.
675 75. Field, D., et al., *The minimum information about a genome sequence (MIGS) specification.*
676 Nature biotechnology, 2008. **26**(5): p. 541-547.
677 76. Taylor, C., et al., *Promoting coherent minimum reporting guidelines for biological and biomedical*
678 *investigations: the MIBBI project.* Nature biotechnology, 2008. **26**(8): p. 889-896.
679 77. Fiehn, O., et al., *The metabolomics standards initiative (MSI).* Metabolomics, 2007. **3**(3): p. 175-
680 178.
681 78. Brazma, A., *Minimum information about a microarray experiment (MIAME)–successes, failures,*
682 *challenges.* TheScientificWorldJournal, 2009. **9**.
683 79. Kimura, M., et al., *Measurement of telomere length by the Southern blot analysis of terminal*
684 *restriction fragment lengths.* Nature protocols, 2010. **5**(9): p. 1596-1607.

685 # Supporting information

686 **S1 Table. Telomere Research Network (TRN) Reporting Guidelines.** Reporting items of the TRN

687 Guidelines by category, with assigned importance for each item and comparison with the related

688 assigned importance of the MIQE guidelines. E: essential information, should be submitted with the
689 manuscript. D: desired information, should be submitted if available.

690 **S2 Table. Morinha Reporting Guidelines.** Reporting items of the guidelines developed by Morinha et al.
691 by category, with assigned importance for each item and comparison with the related assigned
692 importance of the MIQE guidelines. A: very important. B: somewhat important. C: somewhat important,
693 should be submitted if available. E: essential information, should be submitted with the manuscript. D:
694 desired information, should be submitted if available.

695 **S3 Table. Grading rubric for TRN and Morinha Reporting Guidelines.**

696 **S4 Table. Results of reporting guidelines for individual papers.** Scores for each category are provided as
697 a percentage of items in that category sufficiently reported. A total score for each set of reporting
698 guidelines are calculated from all items as well as an average overall score between the two reporting
699 guidelines.

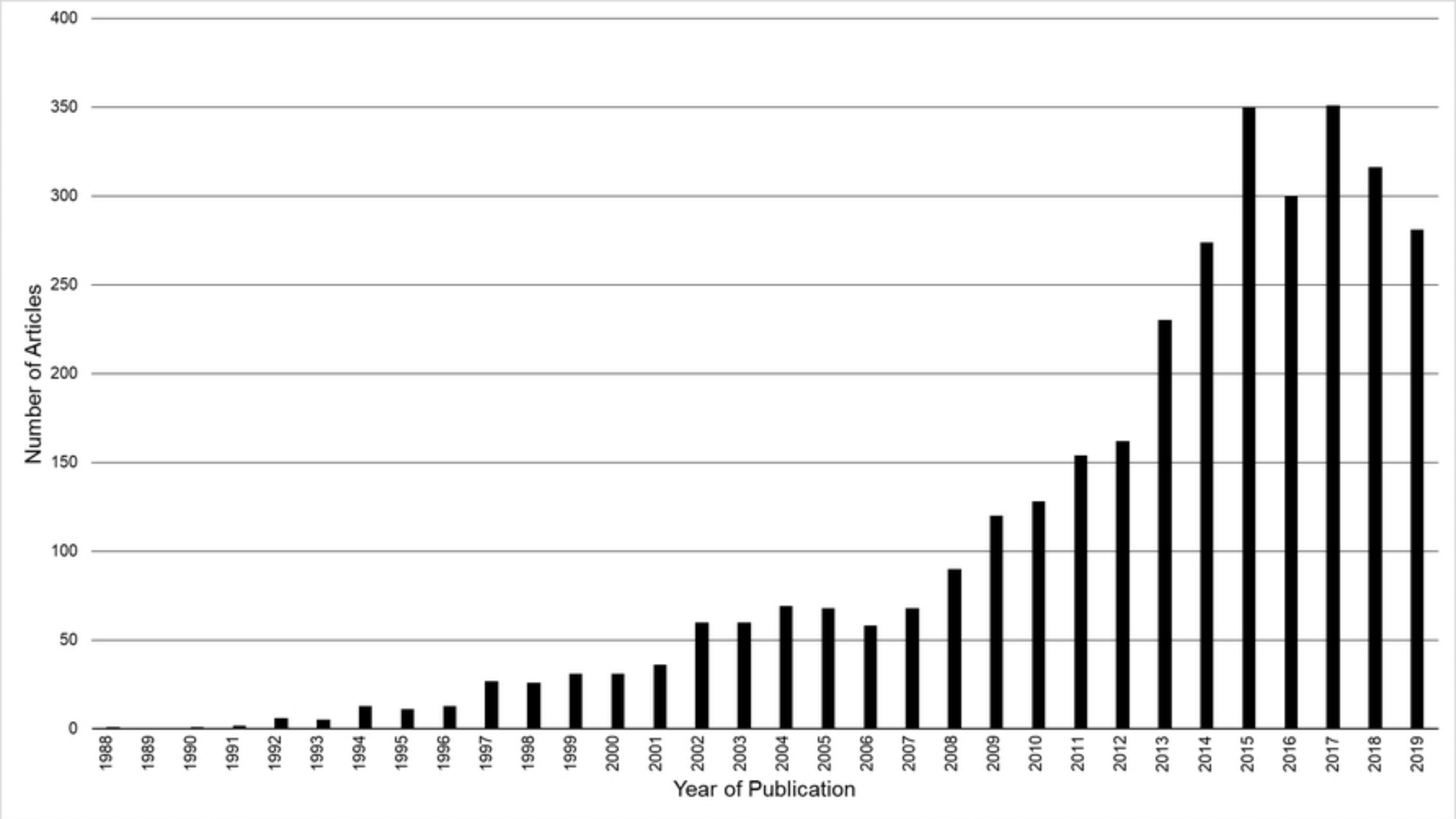

400
350
300
250
200
150
100
50
0
Number of Articles
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year of Publication

Figure 1

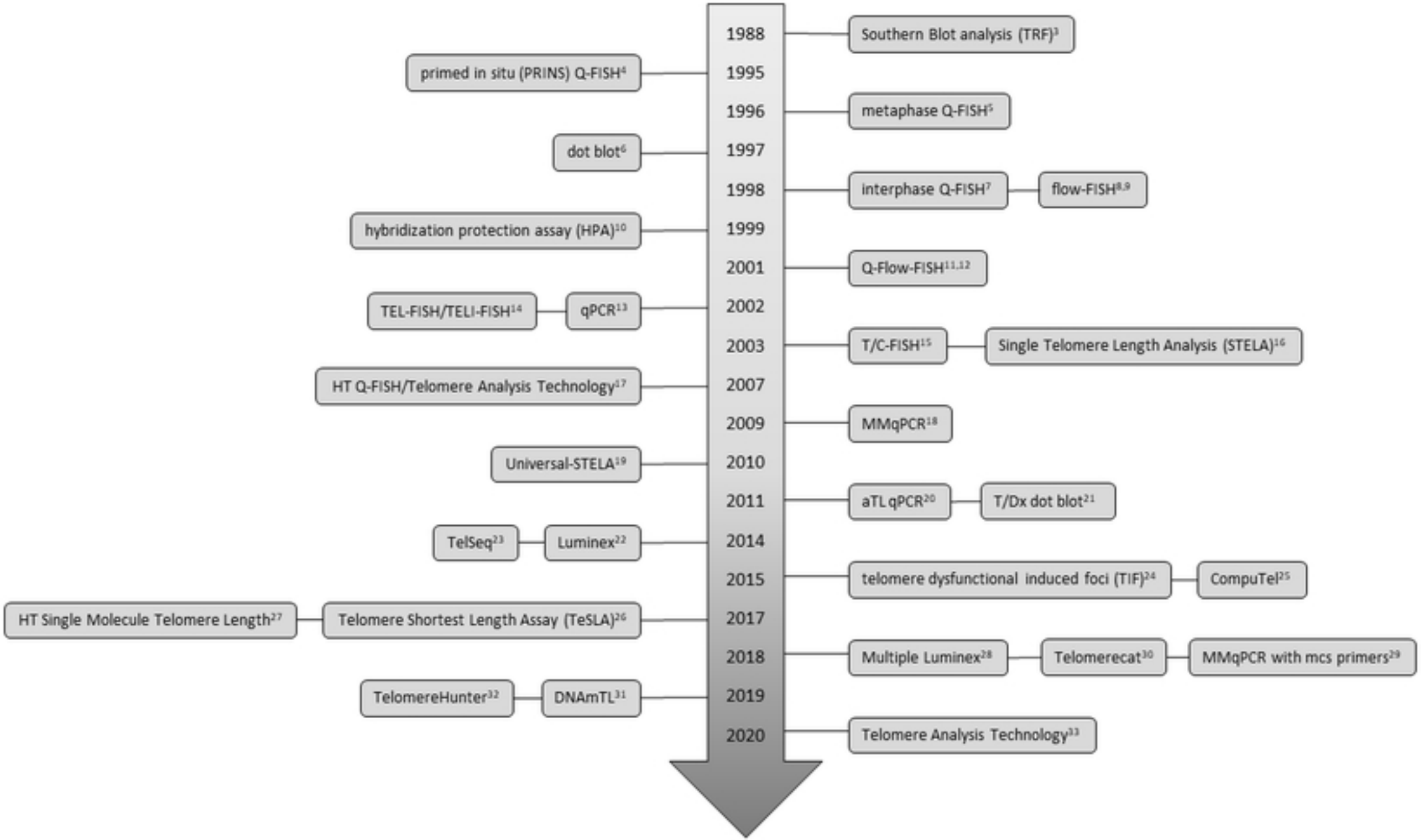
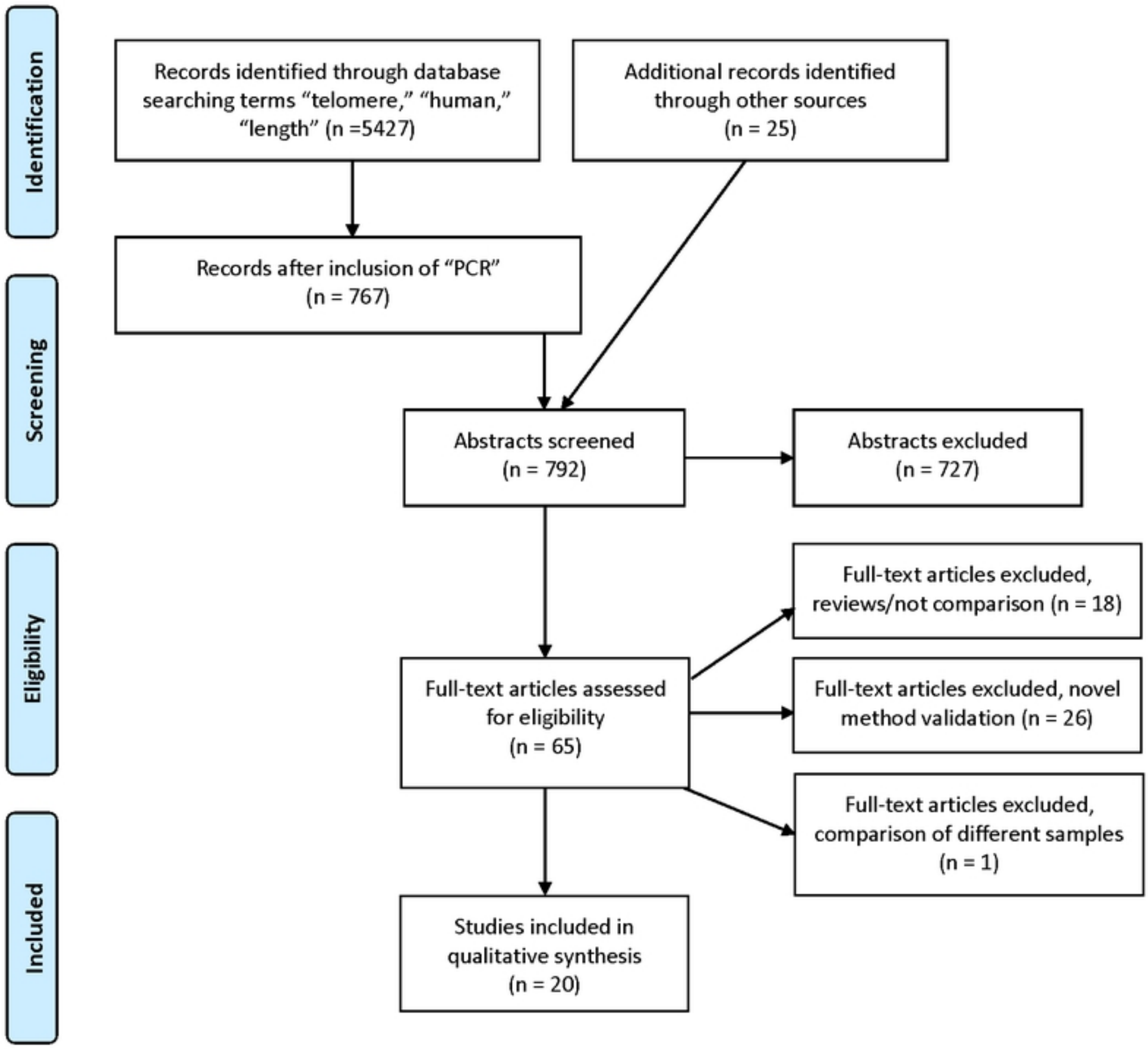

1988
Southern Blot analysis (TRF)[3]
1995
primed in situ (PRINS) Q-FISH[4]
1996
metaphase Q-FISH[5]
1997
dot blot[6]
1998
interphase Q-FISH[7]
flow-FISH[8,9]
1999
hybridization protection assay (HPA)[10]
2001
Q-Flow-FISH[11,12]
2002
TEL-FISH/TELI-FISH[14]
qPCR[13]
2003
T/C-FISH[15]
Single Telomere Length Analysis (STELA)[16]
2007
HT Q-FISH/Telomere Analysis Technology[17]
2009
MMqPCR[18]
2010
Universal-STELA[19]
2011
aTL qPCR[20]
T/Dx dot blot[21]
2014
TelSeq[23]
Luminex[22]
2015
telomere dysfunctional induced foci (TIF)[24]
CompuTel[25]
2017
HT Single Molecule Telomere Length[27]
Telomere Shortest Length Assay (TeSLA)[26]
2018
Multiple Luminex[28]
Telomerecat[30]
MMqPCR with mcs primers[29]
2019
TelomereHunter[32]
DNAmTL[31]
2020
Telomere Analysis Technology[33]

Figure 2

PRISMA 2009 Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Figure 3

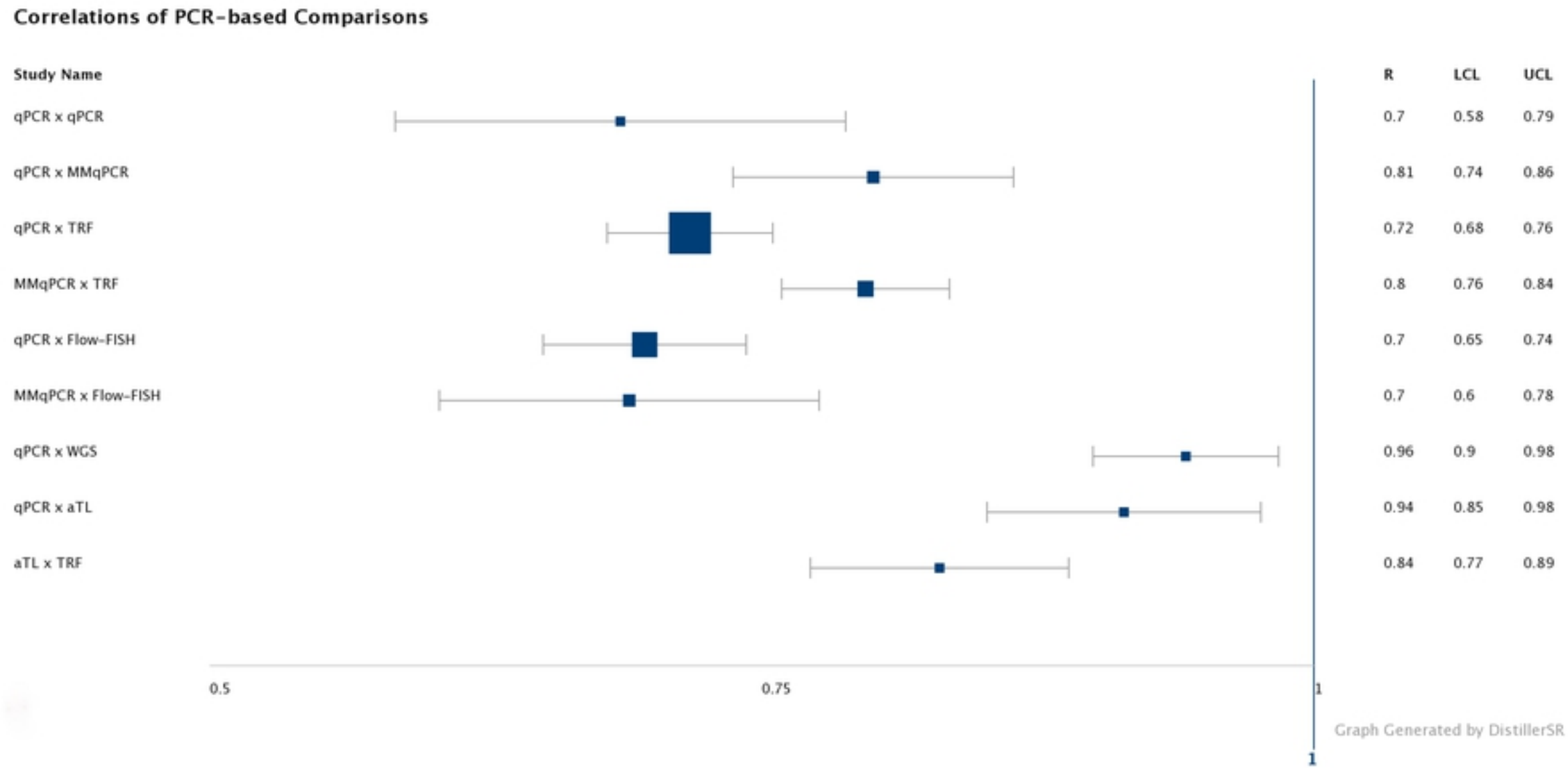

Correlations of PCR-based Comparisons
Study Name
R
LCL
UCL
qPCR x qPCR
0.7
0.58
0.79
qPCR x MMqPCR
0.81
0.74
0.86
qPCR x TRF
0.72
0.68
0.76
MMqPCR x TRF
0.8
0.76
0.84
qPCR x Flow-FISH
0.7
0.65
0.74
MMqPCR x Flow-FISH
0.7
0.6
0.78
qPCR x WGS
0.96
0.9
0.98
qPCR x aTL
0.94
0.85
0.98
aTL x TRF
0.84
0.77
0.89
0.5
0.75
1
Graph Generated by DistillerSR

Figure 4

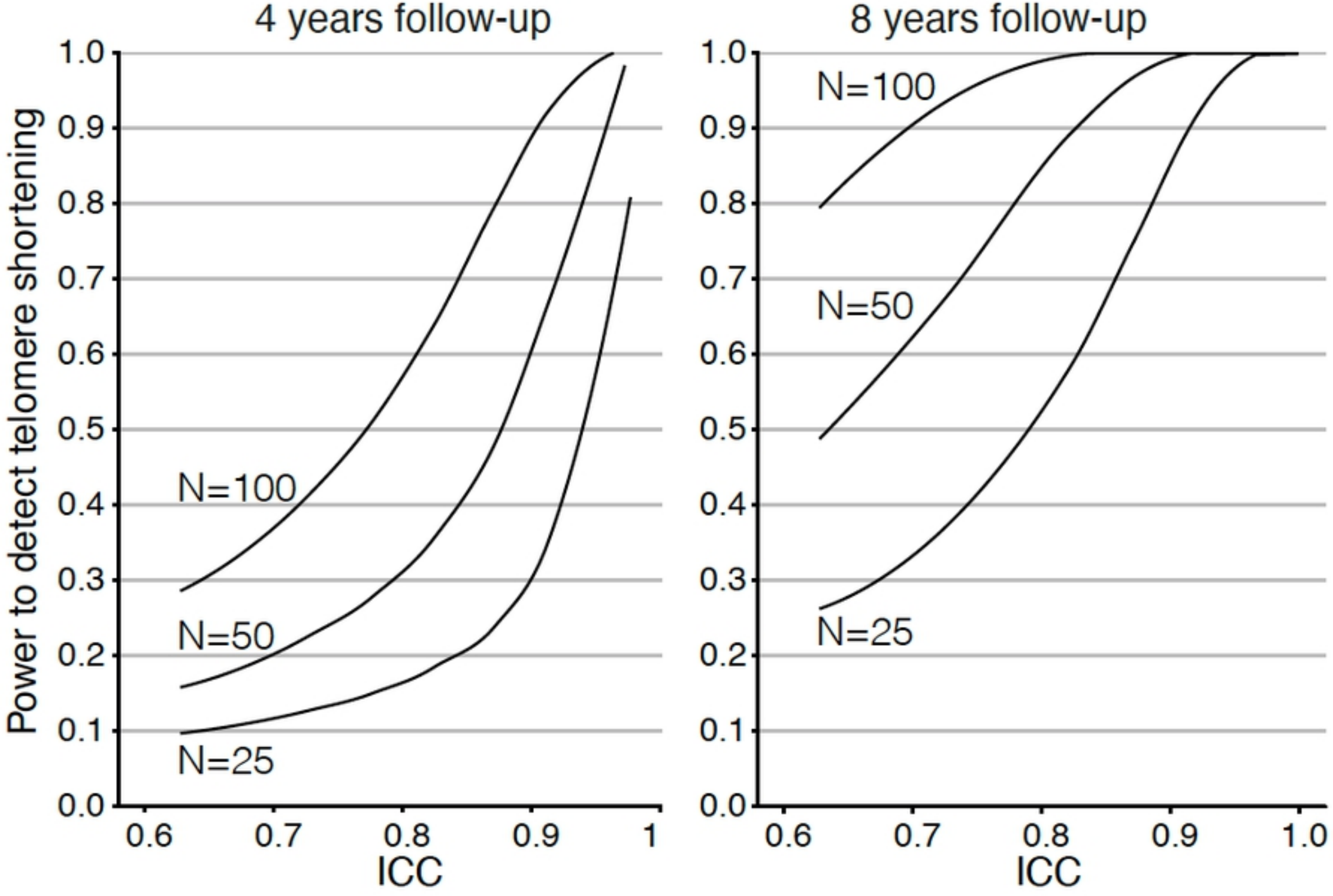

4 years follow-up
8 years follow-up
Power to detect telomere shortening
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
N=100
N=50
N=25
0.6
0.7
0.8
0.9
1
1.0
ICC

Figure 5B 5C

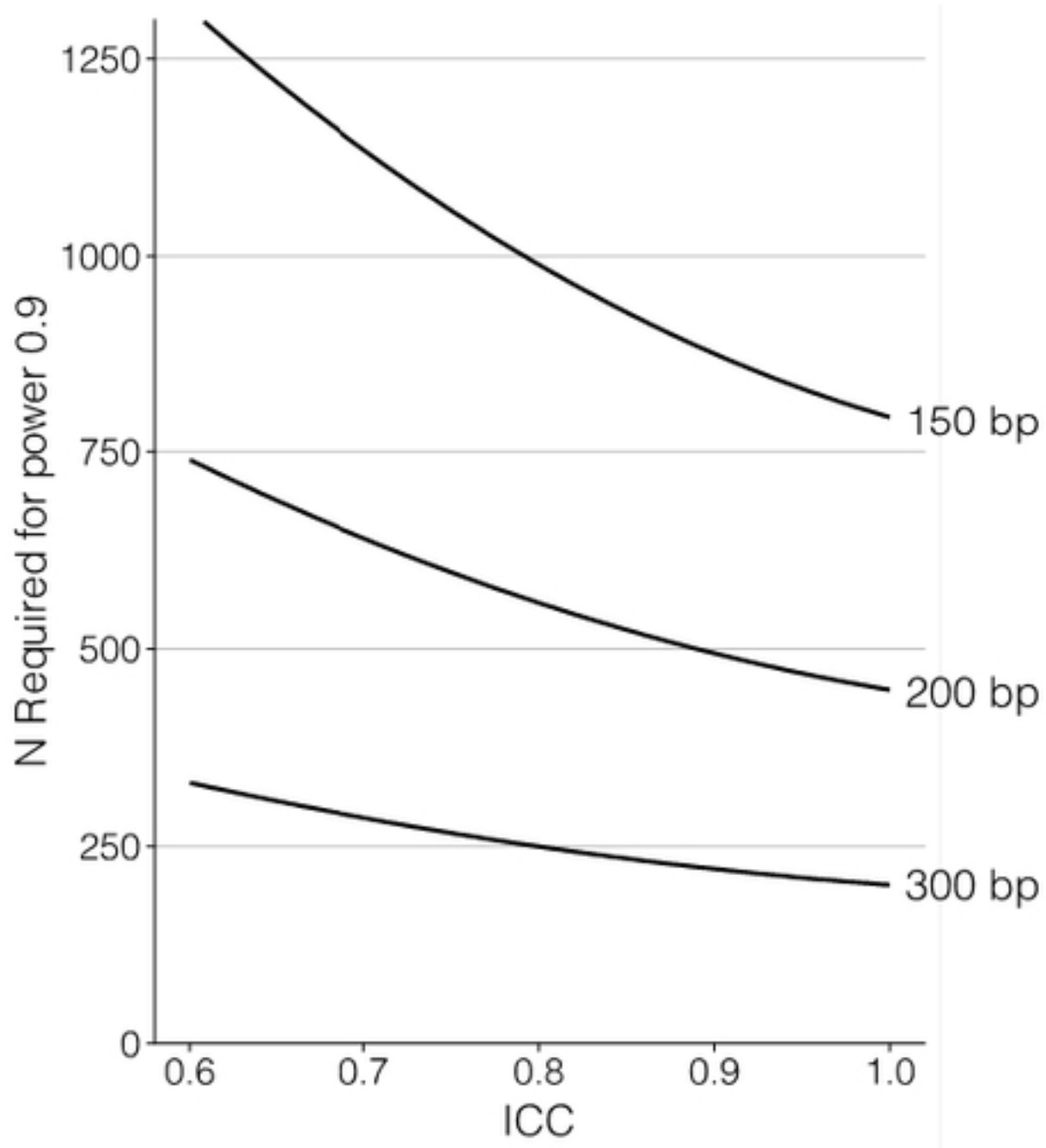

1250
1000
750
500
250
0
N Required for power 0.9
0.6
0.7
0.8
0.9
1.0
ICC
150 bp
200 bp
300 bp

Figure 5A