In recent years, qPCR techniques have gained widespread acceptance as the method of choice for telomere length measurements because of their ease of use and simple adaptation to semi-high throughput approaches. However, informative differences in telomere length are typically small, requiring a measurement precision and reproducibility that are not easily obtained. Thus the continued use of more laborious techniques such as Southern blotting (SB) or STELA might still be justified if they create additional information (STELA) or if they result in higher precision (as had been claimed for SB). Performing the first study ever that assessed objectively reproducibility of telomere length measurements in more than a single laboratory per technique, we found no differences in best reproducibility achieved in laboratories using SB as compared with those using qPCR, and no differences in median reproducibility between SB and qPCR laboratories. In a commentary to our paper, Verhulst et al. now express their belief that this conclusion is not sufficiently justified by our data. Their main argument for this critical view is that we should have used only human leukocytes as DNA source because of their relevance for epidemiology and the smaller age group-specific range of telomere lengths. We cannot follow this argument.
We agree that leukocytes are the most frequently used source of DNA in human epidemiology. However, there are other telomere studies as well, in both human and non-human species. Moreover, whereas most individual epidemiological studies in human peripheral blood cells will cover a restricted range of 3 to 4 kb in telomere length as pointed out by Verhulst et al., a range between 2 and 12 kb is easily covered if multiple studies over the whole human age range are taken into account. Therefore, the aim of our study was to compare methodology for telomere length measurements over a range of telomere lengths relevant for a range of applications and not specifically for epidemiological studies using leukocytes.

We agree that a wider range of telomere length in our test samples might generate ‘inflated’ rank correlations (although we would discuss whether these are ‘inflated . . . beyond what is relevant for LTL in epidemiologic studies’5 for the reason given above). However, this could not have led us ‘to the erroneous conclusion that the SB and qPCR methods yielded similar results’,5 because rank correlations between laboratories or even indicators of variation between laboratories (inter-lab CVs) are not direct measures of technical reproducibility (as our results in Table 2 and Supplementary Table S4 clearly illustrate4).

We are aware that on a dataset like ours, rank correlations are sensitive to relatively minor variation in the underlying data. Relatively small changes in the way telomere lengths are calculated, for instance a shift from Gaussian to arithmetic fit for the quantitative evaluation of distances in the gel, can greatly change the range of correlation coefficients as illustrated in our erratum6 and response7 to a separate comment8 to our paper. Therefore, our conclusion of similar accuracy of gel-based and qPCR methods was primarily based on the intra-lab reproducibility of blinded repeat measurements and not on rank correlations.

We are very surprised by the unsupported statement by Verhulst et al. that ‘the statistical tests used by MR . . . are underpowered and consequently of limited value.’5 We have in fact indicated the power of our experimental design in the paper. Eight internal repeat samples per laboratory are fully sufficient to detect a difference between SB and qPCR results of the magnitude found in an earlier paper,3 although they are not enough to detect differences between individual laboratories with certainty in a multi-lab comparison (which was not our goal with this study). All data on which these power calculations are based are published, and we invite Verhulst et al. to reproduce them to resolve their puzzlement.

In essence, however, the truth in this matter does not require any elaborate statistical reasoning. Verhulst et al. base their trust in an a priori better reliability of the SB technique on a comparison between a single SB laboratory with a single qPCR lab,3 further supported only by some very indirect argument.5 This generalization is evidently illogical, and our data (see Figure 34) prove it to be wrong, irrespective of our choice of DNA samples or of statistical considerations. We conclude that both SB and qPCR are capable of similarly precise (or rather imprecise) measurements of relative telomere length. We fully agree with Verhulst et al. that large-scale epidemiological LTL studies comparing techniques as performed in multiple experienced laboratories are necessary to improve sensitivity and specificity of this marker. However, our data suggest that the most important determinant of result precision is not the choice between SB and qPCR but rather the optimization of a number of factors within each technique and laboratory. As discussed,4 the specific nature of these factors and their improvement await further examination.

References