Review and extension of a meta-analysis

Matthew Finster
Master of Data Science student
LaTrobe University

Word count: 2233 excluding tables, figures and references

September 2024

Contents

1 Introduction

A meta-analysis[1] by Sánchez-González et al, published in the JMIR Public Health and Surveillance journal, was reviewed and extended. It included 8 randomised controlled trials (RCTs) and 1 clinical trial (CT) on 1058 diverse participants performed under a range of study conditions lasting from 8 weeks to 1 year.

All studies employed polymerase chain reactions (PCR) based techniques to measure telomere lengths (TLs). PCR-based techniques are the most commonly used and efficient method for TL measurements, despite not being the most precise[2]. PCR-based techniques provide a relative telomere length (T) compared to a single (S) copy gene and results are expressed as a T/S ratio[2]. Smaller T/S ratios have been associated with accelerated aging, including an increase in age-related diseases such as osteoporosis, cancer, and dementia[1]. Physical exercise, on the other hand, is a well-known strategy used to prevent the onset of age-related diseases. The pre-specified objective of this meta-analysis was to therefore understand the impact of different physical exercise modalities on telomere length in healthy people.[1]

The claimed key finding from the study is that interventional exercise generally did not significantly increase TL compared with the control interventions, however high-intensity interval training (HIIT) did significantly increase TL compared to the control interventions.

2 Methods

The authors provide a completed checklist for their adherence to the PRISMA[3] statement. Details of the search strategy and study selection are provided by the authors but are omitted here. 5/8 RCTs showed a high risk of bias and 3/8 showed concerns. Prior to conducting the meta-analysis, the authors removed participants and subdivided studies. An inspection of each study confirmed several changes made by authors, outlined in Table 2 in the Appendix.

The outcome of interest was the difference between T/S ratios before and after exercise interventions, which were reported as Mean Differences (MDs) between the treatment and control groups at post-intervention. The Restricted Maximum Likelihood Method (REML) estimator $\hat{\tau}^2_{\text{REML}}$ was used for the between-study variance and the Hartung and Knapp (HK) adjustment was also applied. The HK adjustment is typically employed when the number of studies is small or if there is significant heterogeneity between studies[4], producing more conservative CI and prediction interval (PI) estimates. The authors used the Cochran Q statistic to assess the presence of between-study heterogeneity.

Although this practice is misguided[5], the authors stated that Random Effects Models (REMs) would be used when a Q test was significant at $P < 0.1$ and I^2 was greater than 25%. As the I^2 statistic exceeded 25%, the authors subsequently sought to reduce this statistic by omitting studies. The authors conducted an influential case analysis graphically, using recommendations by Baujat, et al.[6]. Finally, with the remaining 10 studies, the authors conducted a subgroup analysis based on the type of exercise performed in each study (resistance training [RT], aerobic exercise [AE], or HIIT). The authors produced all forest plots using the metafor package in R.

The authors assessed publication bias via a contour-enhanced funnel plot. Lastly, the authors performed a meta-regression analysis to determine whether exercise intensity and duration, year of publication, and methodological quality influenced the effect sizes.

3 Main findings

The authors reported that *"overall, exercise did not produce a significant increase in TL compared with that of the control groups" (MD 0.02; 95% CI* $[-0.10, 0.13]$*; P = .77; N = 1058).* Between-study variability was estimated with an I^2 value of 70%, which indicated significant heterogeneity among the studies included in the analysis *($P < 0.01$).*

The authors found that two studies should be omitted due the results of their influential case analysis. After the authors these removed two studies, there was still no evidence for a significant difference in TL between the exercise and control groups *(MD 0.02; 95% CI* $[-0.03, 0.07]$*; P = 0.50; N = 772; I^2 = 23%)*

The authors reported the following outcomes from the subgroup analysis: (i) the RT groups showed no significant difference to the control groups *(MD* -0.02*, 95% CI [-0.10,0.05]; P = .54; I^2=16%)*; (ii) the AE groups showed no significant difference to the control groups *(MD* -0.01*, 95% CI* $[-0.04, 0.06]$*; P = .64; I^2 =0%)*; and (iii) the HIIT groups showed significant differences compared with the control groups, with greater TLs observed in the HIIT groups *(MD 0.15, 95% CI* $[0.03, 0.26]$*; P = .01; I^2=0%).*

The funnel plot produced by the authors showed obvious signs of asymmetry. Although the authors reported that *"the studies included in this analysis were not significant; therefore, publication bias was ruled out"*, publication bias is only one of many plausible explanations for the asymmetry[7]. The authors stated that meta-regression analyses showed that exercise intensity and duration, year of publication, and methodological quality did not influence the observed effect sizes.

4 Replication of main findings

All meta-estimated MDs, associated CIs and I^2 values were able to be replicated precisely for the main meta-analysis. The between-study variance estimator $\hat{\tau}^2_{\text{REML}}$ and HK adjustment, as indicated by the authors, was used. For results, see Figure 1.

To further replicate the findings of the original study, the R metainf package was used to produce a *leave-one-out* analysis. It was able to replicate the finding that I^2 reduced from 70% to 30% when the outlier[8] study was removed. For this analysis, see Figure 2. A Baujat plot was also created using the baujat plot function in the meta package. The Baujat plot did not produce the same results as the authors and, notably, the second study that the authors omitted[9] was not as influential as they proposed in their analysis. For these results, see Figure 3. To continue replicating the authors' results, this influential case was omitted before the sub-group analysis anyway.

When replicating the results of the sub-group analysis, it was noted that the authors dropped the usage of the HK adjustment, as indicated by the publishing of z-scores rather than t-scores. The results of the sub-group analysis were able to be mostly replicated using the meta package in R. Like the authors found, there were no significant differences between exercise and control groups in the RT or the AE subgroups. The difference between the groups in the HIIT subgroup was considered significant *(MD = 0.15; 95% CI [0.03, 0.26], p = 0.01, I^2=0%)*. For these results, see Figure 4.

Lastly, a funnel plot was created using the rma and funnel functions in the metafor package in R. The same asymmetry was observed in our replication of findings. See Figure 5 for further inspection.

Note the meta package was preferred over metafor package here as it tends to be more user-friendly. Having received many updates since it was labelled as inferior in 2010[10]; the meta package is similar in functionality to metafor.

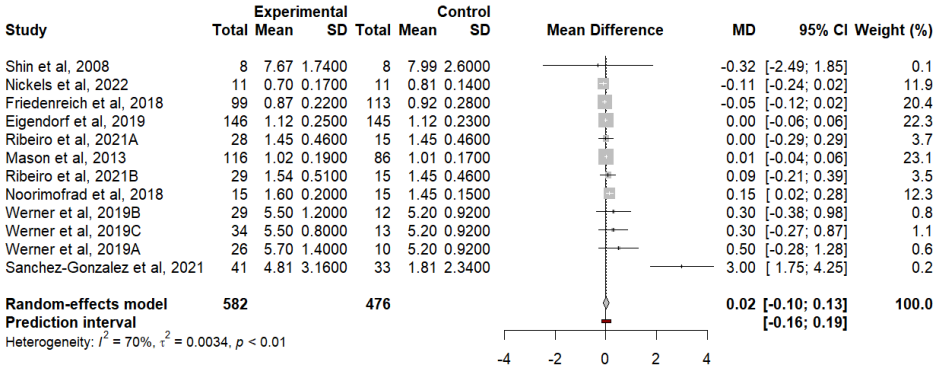

Figure 1: Forest plot for meta-analysis of interventions, produced with forest function using the meta package in R

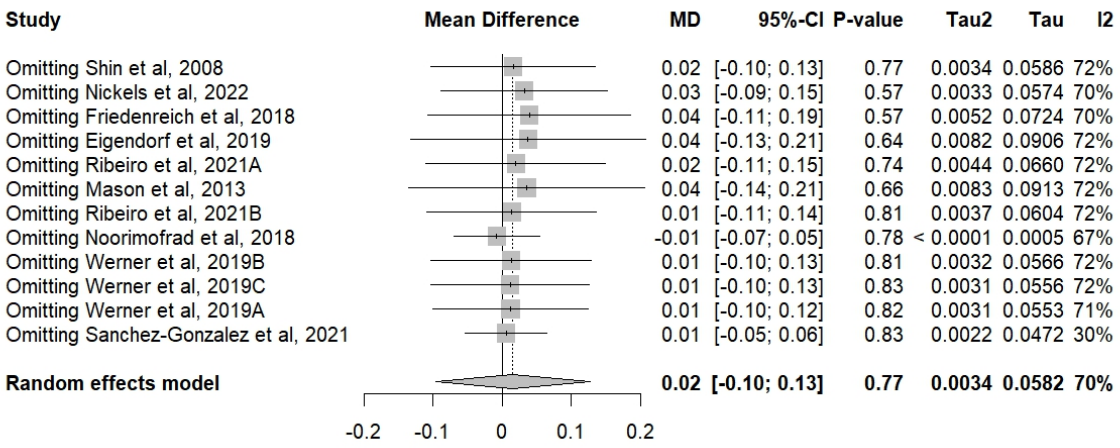

Figure 2: Leave-one-out analysis, produced with forest and metainf functions using the meta package in R

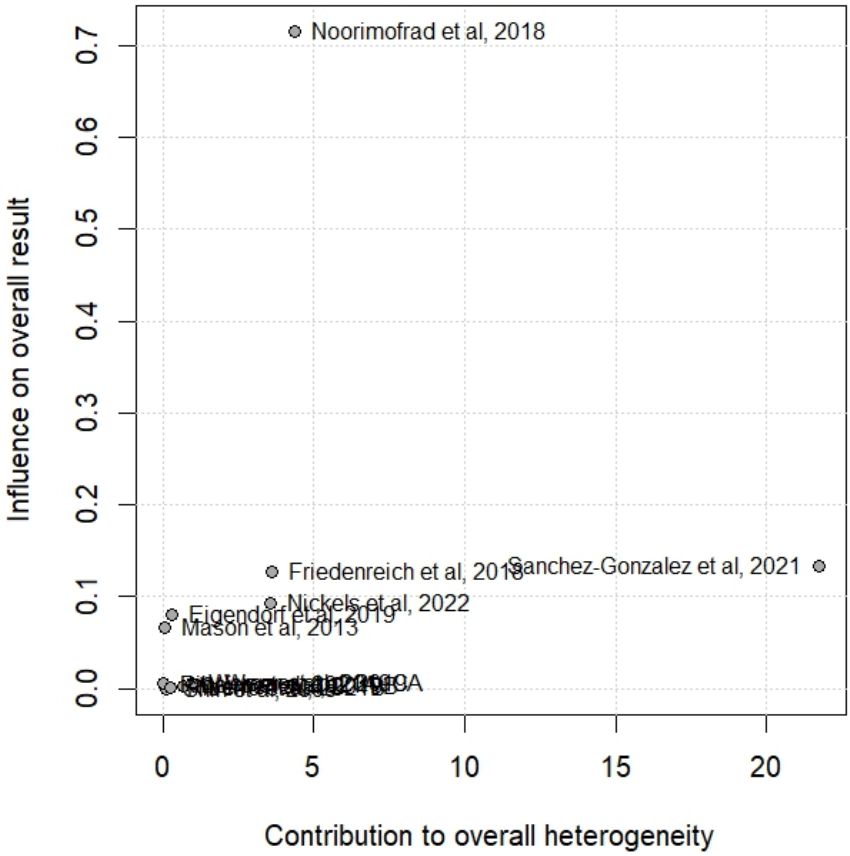

Figure 3: Influential case analysis, produced using the baujat function in the meta package in R

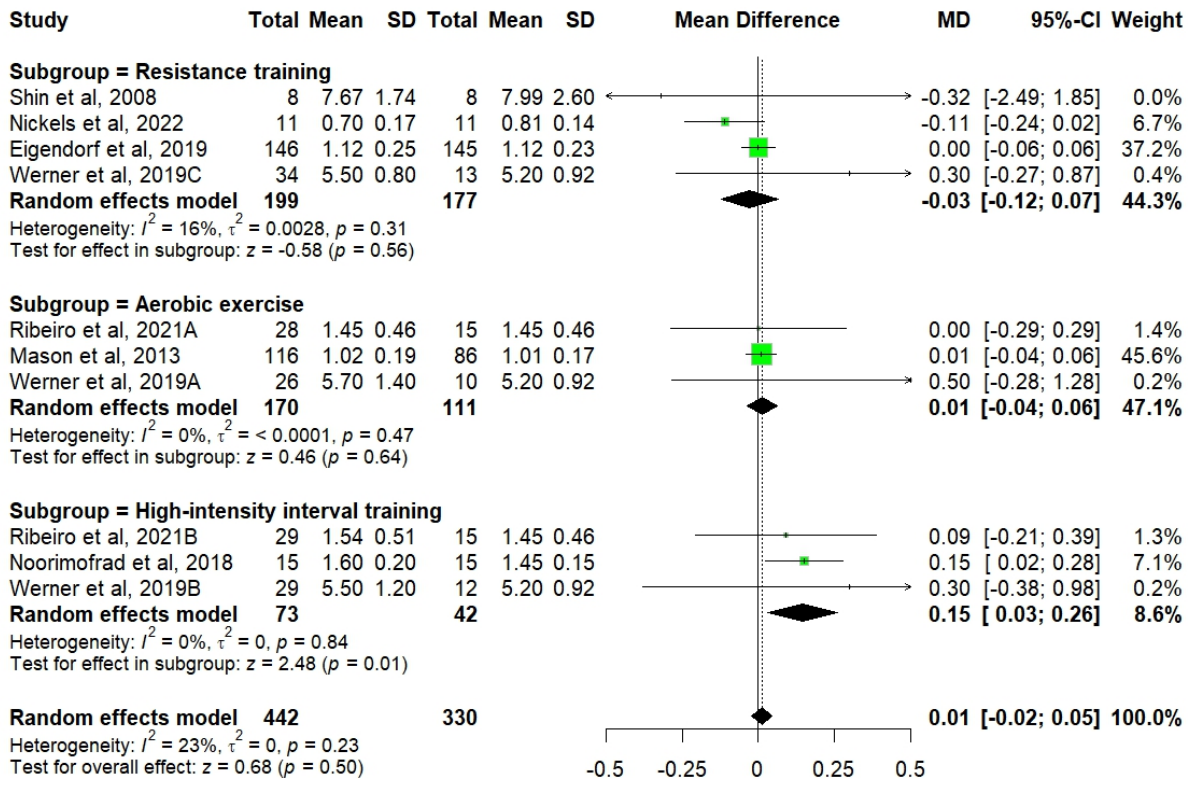

Figure 4: RT vs AE vs HIIT subgroup analysis, produced using the forest function in the meta package in R

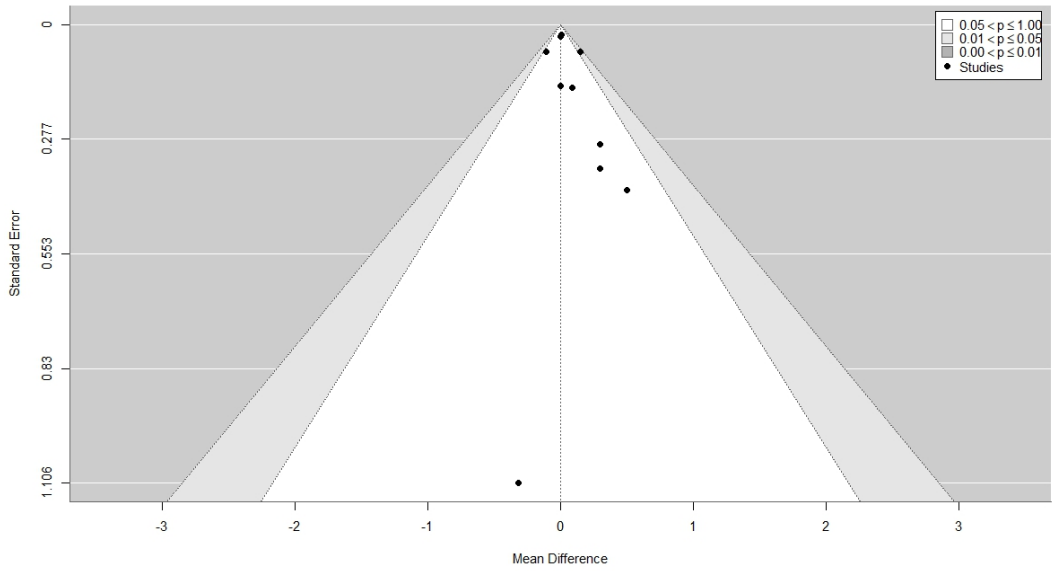

Figure 5: Funnel plot including the 10 studies that remained after the omission of 2 studies, produced using the funnel function in the metafor package in R

5 Limitations

The most notable limitation is the choice of outcome measure for each study. The use of MDs assumes that outcome measures are on the same scale across all studies[11]. In two studies, the authors chose to include TL measures that used kb pairs[12][13], even when there was T/S ratio data available, wheras for all other studies, they used T/S ratios [8][9][14][15][16][17][18]. This inconsistency should have been addressed, either by converting all outcome measures to T/S ratios or by opting to use SMDs to account for the different units. The authors also used the post-intervention means, rather than change scores (differences between the pre- and post-intervention means). Change scores can be more *"efficient and powerful than comparison of final values, as it removes a component of between-person variability from the analysis for understanding the actual effect of the intervention"*[11].

There are also significant limitations in the findings and concluding remarks. In light of the sizeable degree of heterogeneity, a more appropriate conclusion would have been that there is not enough evidence to suggest that exercise interventions significantly impact TL, and that the effectiveness of such interventions is likely to vary in the presence of other factors. In relation to the sub-group analysis, the authors claimed that HIIT *"significantly increases the length of telomeres"*. This claim is weak at best. As shown in Figure 4, while the HIIT subgroup results in a significant pooled effect size (p = 0.01, 95% CI [0.03, 0.26]), there are only three studies in this sub-group. Two of the three CIs for the individual studies overlap with zero and the lower bound of the other study is also very close to 0. The authors should have emphasized that this sub-group consisted of a limited number of studies and wide CIs, and that further research is needed before drawing definitive conclusions about the effect of HIIT on TL.

6 Extensions

As an extension and to address limitations; the meta-analysis was conducted again, this time using change scores and only T/S ratios as the outcome measure. Where CIs were provided for means, these were converted to SE and then SD using

$$SE = \frac{\text{Upper} - \text{Lower}}{2 \times 1.96}$$

and

$$SD = SE \times \sqrt{n}$$

SDs were then used to calculate an SD for pre- and post-intervention change using the *pooled standard deviation for change scores* formula:

$$SD_{\Delta} = \sqrt{SD_{pre}^2 + SD_{post}^2 - 2 \times r \times SD_{pre} \times SD_{post}}$$

where SD_{pre} and SD_{post} are the SDs of the pre- and post-intervention scores, respectively. r is the correlation coefficient between the pre- and post-intervention scores. This correlation coefficient r was not reported, therefore 0.5 was used as a default. In hindsight, r should have been imputed[11].The results of the new meta-analysis using change scores can be found in Figure 6.

We are 95% confident that the mean effect is between the CI of [-0.02, 0.07]. We are also 95% confident that the true effect for any randomly selected study, including those not in the meta-analysis[19], will be within the PI of [-0.10, 0.15]. As the intervals cover 0, we can answer the original research question by stating that there is not enough evidence to suggest that exercise has an effect on TL. However, sub-group analyses did provide some support that there may be beneficial effects for RT and HIIT on TL. See Appendix.

A sensitivity analysis was conducted to determine the best estimator between the the DerSimonian-Laird (DL) method and REML, which are both appropriate estimators for MD. The DL increased the upper bound of the CI and PI by 0.01, but produced no changes to I^2 values. The REML estimator was likely a good choice as it is generally unbiased and efficient[20]. Another funnel plot was also created, this time only with the HIIT sub-group to show that, clearly, a) more studies are required, and b) the asymettry is likely due to the number of studies and small study effects. See Figure 7. An Egger's test was not performed as there were <10 studies[7].

Finally, as an extension, meta-regressions were conducted to assess the whether different moderator variables, including subgroups and the average age of participants, accounted for some of the residual heterogeneity found in the analysis. Given the small number of studies available, these moderators were not combined into a single meta-regression. When using age as a moderator; 55.25% (I^2) of the total variability in effect sizes between studies was due to real differences, rather than random chance alone. This residual heterogeneity (I^2) reduced to 21.99% when using subgroups as the moderator instead of age. The meta-regression results for subgroups can be found in Table 1. Although the QE test suggested that this residual heterogeneity is insignificant ($p = 0.63$), it is still important to interpret the subgroup results cautiously. For instance, instead

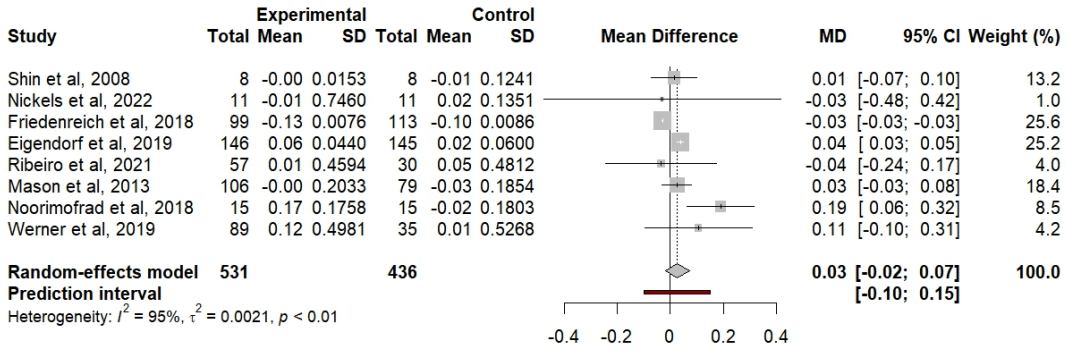

Figure 6: Forest plot for meta-analysis using change scores, produced with forest function using the meta package in R

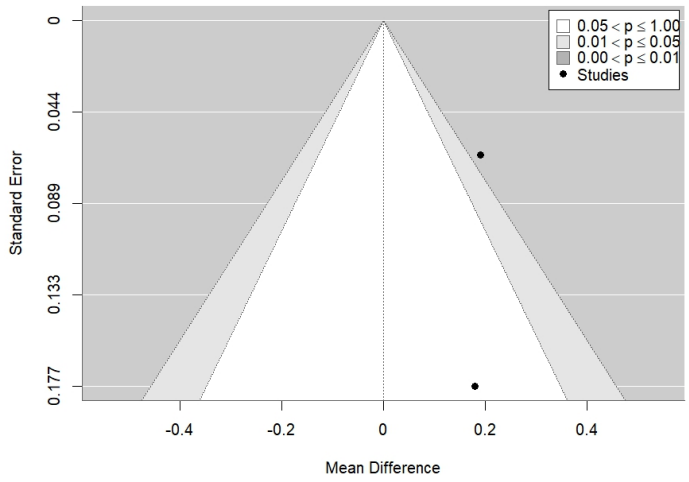

Figure 7: Funnel plot for HIIT studies showing asymmetry, likely caused by insufficient number of studies and small study effects

of relying on (I^2) as an absolute measure of heterogeneity, some authors have argued for the use of PIs to show how effect sizes vary across populations[21]. In this case, it is important to note that our PIs were large and covered 0, which could have been due to the lack of statistical power or the possibility of no true effect.

7 Conclusion

We were able to replicate the key findings of our meta-analysis. While we agree with the main conclusions, we caution that statements about the significant effect sizes should not be generalized across subgroups without further scrutiny.

It's important to note that the results from the meta-regressions are likely influenced by the small number of studies included in the analysis. With a limited dataset, regression models can easily identify patterns or relationships, fit the data and potentially exaggerate the explanatory power of the moderators[22]. With more data, the effects of HIIT might not have been as pronounced or consistent. As a result, any causal interpretations should be made cautiously.

Table 1: Mixed-effects model results (k = 10; tau^2 estimator: REML)

Statistic	**Value**			
I^2 *(residual heterogeneity)*	21.99%			
R^2 *(amount of heterogeneity accounted for)*	75.25%			
Model Results	**Estimate**	**SE**	**95% CI (Lower)**	**95% CI (Upper)**
Intercept (aerobic exercise)	-0.0133	0.0189	-0.0504	0.0238
Subgroup: High-intensity interval training	0.2020	0.0670	0.0707	0.3333*
Subgroup: Resistance training	0.0459	0.0283	-0.0096	0.1014

Note: *Significant at $p < 0.01$.

References

[1] J. Sánchez-González, J. Sánchez-Rodríguez, S. Varela-Rodríguez, R. González-Sarmiento, C. Rivera-Picón, R. Juárez-Vela, C. Tejada-Garrido, J. Martín-Vallejo, and V. Navarro-López, "Effects of Physical Exercise on Telomere Length in Healthy Adults: Systematic Review, Meta-Analysis, and Meta-Regression," *JMIR Public Health Surveill.*, vol. 10, p. e46019, 2024. doi: 10.2196/46019.

[2] J. Lin, D. L. Smith, K. Esteves, and S. Drury, "Telomere length measurement by qPCR – Summary of critical factors and recommendations for assay design," *Psychoneuroendocrinology*, vol. 99, pp. 271–278, 2019.

[3] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," *PLoS Med.*, vol. 6, no. 7, p. e1000097, 2009. doi: 10.1371/journal.pmed.1000097.

[4] J. IntHout, J. P. A. Ioannidis, and G. F. Borm, "The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method," *BMC Med Res Methodol*, vol. 14, p. 25, 2014. doi: 10.1186/1471-2288-14-25.

[5] D. C. Hoaglin, D. C., "Misunderstandings about Q and 'Cochran's Q test' in meta-analysis," *Statist. Med.,* vol. 35, pp. 485–495, 2016. doi: 10.1002/sim.6632.

[6] B. Baujat, C. Mahé, J. P. Pignon, and C. Hill, "A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials". *Statist. Med.*, vol. 21, pp. 2641-2652, 2002. doi: 10.1002/sim.1221

[7] J. A. C. Sterne, A. J. Sutton, J. P. A. Ioannidis, N. Terrin, D. R. Jones, J. Lau, J. Carpenter, G. Rücker, R. M. Harbord, C. H. Schmid, J. Tetzlaff, J. J. Deeks, J. Peters, P. Macaskill, G. Schwarzer, S. Duval, D. G. Altman, D. Moher, and J. P. T. Higgins, "Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials," *BMJ*, vol. 343, no. 7818, pp. 302-307, 2011. doi: 10.1136/bmj.d4002.

[8] J. L. Sánchez-González, J. L. Sánchez-Rodríguez, J. Martín-Vallejo, A. Martel-Martel, and R. González-Sarmiento, "Effects of physical exercise on cognition and telomere length in healthy older women," *Brain Sci.*, vol. 11, no. 11, p. 1417, 2021.

[9] C. M. Friedenreich, Q. Wang, N. S. Ting, D. R. Brenner, S. M. Conroy, J. B. McIntyre, A. Mickle, K. S. Courneya, and T. Beattie, "Effect of a 12-month exercise intervention on leukocyte telomere length: Results from the ALPHA Trial," *Cancer Epidemiol.*, vol. 56, pp. 67-74, 2018.

[10] W. Viechtbauer, "Conducting Meta-Analyses in R with the metafor Package," *J. Stat. Softw.*, vol. 36, no. 3, pp. 1-48, 2010. doi: 10.18637/jss.v036.i03.

[11] J. P. T. Higgins and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions*, 1st ed., vol. 1. Wiley, 2008.

[12] Y. A. Shin, J. H. Lee, W. Song, and T. W. Jun, "Exercise training improves the antioxidant enzyme activity with no changes of telomere length." *Mechanisms of ageing and development*, vol. 129, no. 5, pp. 254-260, 2008.

[13] C. M. Werner, A. Hecksteden, A. Morsch, J. Zundler, M. Wegmann, J. Kratzsch, J. Thiery, et al., "Differential effects of endurance, interval, and resistance training on telomerase activity and telomere length in a randomized, controlled study," *Eur. Heart J.*, vol. 40, no. 1, pp. 34-46, 2019.

[14] M. Nickels, S. Mastana, M. Denniff, V. Codd, and E. Akam, "Pilates and telomere dynamics: A 12-month longitudinal study," *J. Bodywork Movement Ther.*, vol. 30, pp. 118-124, 2022.

[15] J. Eigendorf, A. Melk, S. Haufe, D. Boethig, D. Berliner, A. Kerling, M. Kueck, et al., "Effects of personalized endurance training on cellular age and vascular function in middle-aged sedentary women," *Eur. J. Prev. Cardiol.*, vol. 26, no. 17, pp. 1903-1906, 2019.

[16] V. B. Ribeiro, D. C. C. Pedroso, G. S. Kogure, I. P. Lopes, B. A. Santana, H. C. D. Souza, R. A. Ferriani, R. T. Calado, C. L. M. Furtado, and R. M. dos Reis, "Short-term aerobic exercise did not change telomere length while it reduced testosterone levels and obesity indexes in PCOS: A randomized controlled clinical trial study," *Int. J. Environ. Res. Public Health*, vol. 18, no. 21, p. 11274, 2021.

[17] C. Mason, R.-A. Risques, L. Xiao, C. R. Duggan, I. Imayama, K. L. Campbell, A. Kong, et al., "Independent and combined effects of dietary weight loss and exercise on leukocyte telomere length in postmenopausal women," *Obesity*, vol. 21, no. 12, pp. E549-E554, 2013.

[18] S. Noorimofrad and K. Ebrahim, "The effect of high intensity interval training on telomere length and telomerase activity in non-athlete young men," *J. Basic Res. Med. Sci.*, vol. 5, no. 2, pp. 1-7, 2018.

[19] J. IntHout, J. P. A. Ioannidis, M. M. Rovers, and J. J. Goeman, "Plea for routinely presenting prediction intervals in meta-analysis," *BMJ Open*, vol. 6, no. 7, p. e010247, 2016.

[20] W. Viechtbauer, "Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model," *J. Educ. Behav. Stat.*, vol. 30, no. 3, pp. 261-293, 2005. doi: 10.3102/10769986030003261.

[21] M. Borenstein, J. P. T. Higgins, L. V. Hedges, and H. R. Rothstein, "Basics of meta-analysis: I2 is not an absolute measure of heterogeneity," *Res. Synth. Methods*, vol. 8, no. 1, pp. 5-18, 2017. doi: 10.1002/jrsm.1230.

[22] S. G. Thompson, and J. P. T. Higgins, "How should meta-regression analyses be undertaken and interpreted?," *Statist. Med.*, vol. 21, pp. 1559-1573, 2002.

8 Appendix

Table 2: Summary of edits to study sample sizes by authors

Study	Original study design	Changes
Ribeiro, et al.[16]	Participants: 87 Group 1 - Aerobic: 28 Group 2 - HIIT: 29 Group 3 - Control: 30	No change to number of participants Study subdivided into: Ribeiro A - Aerobic: 28, Control: 15 Ribeiro B - HIIT: 29, Control: 15
Mason, et al.[17]	Participants: 439 Group 1 - Dietary weight loss: 118 Group 2 - Aerobic: 117 Group 3 - Diet + Aerobic: 117 Group 4 - Control: 87	Removed 235 participants (Groups 1 and 3) Removed 1 participants from Group 2 Removed 1 participant from Group 4 Participants: 202 Aerobic: 116 Control: 86
Werner, et al.[13]	Participants: 124 Group 1 - Aerobic: 26 Group 2 - HIIT: 29 Group 3 - Resistance: 34 Group 4 - Control: 35	No change to number of participants Study subdivided into: Werner A - Aerobic: 26, Control: 10 Werner B - HIIT: 29, Control: 12 Werner C - Resistance: 34, Control: 13

Table 3: Comparison of authors' outcome measures, our outcome measures, and issues

Study	Authors' Mean ± SD	Our Mean ± SD	Issues
Shin, et al. [12]	Experimental: 7.67 ±1.74; Control: 7.99 ±2.6	Experimental: -0.0003 ±0.02; Control: -0.0144 ±0.12	Used kb as the measure of telomere length. Used post results instead of change from pre- to post-intervention.
Nickels, et al. [14]	Experimental: 0.70 ± 0.17; Control: 0.81 ± 0.14	Experimental: -0.013 ±0.75; Control: -0.018 ±0.14	Used post results instead of change from pre- to post-intervention.
Friedenreich, et al. [9]	Experimental: 0.87 ± 0.22; Control: 0.92 ± 0.28	Experimental: -0.13 ±0.008; Control: -0.1 ±0.009	Used post results instead of change from pre- to post-intervention.
Eigendorf, et al. [15]	Experimental: 1.12 ± 0.23; Control: 1.12 ± 0.25	Experimental: 0.058 ±0.044; Control: 0.2 ±0.06	Used post results instead of change from pre- to post-intervention..
Ribeiro, et al. A [16]	Experimental: 1.45 ± 0.46; Control: 1.45 ± 0.46	Experimental: 0.015 ±0.46; Control: 0.05 ±0.48	Used post results instead of change from pre- to post-intervention. Did not pool results of IAT and CAT groups, wrongly assigning IAT to HIIT.
Ribeiro, et al. B [16]	Experimental: 1.54 ± 0.51; Control: 1.45 ± 0.46	As above	As above.
Mason, et al.[17]	Experimental: 1.02 ± 0.19; Control: 1.01 ± 0.17	Experimental: -0.002 ±0.203; Control: -0.027 ±0.185	Used post results instead of change from pre- to post-intervention.
Noorimofrad, et al.[18]	Experimental: 1.6 ± 2.00; Control: 1.45 ± 0.46	Experimental: 0.17 ±0.176; Control: -0.02 ±0.18	Used post results instead of change from pre- to post-intervention.
Werner, et al. A[13]	Experimental: 5.7 ± 1.4; Control: 5.2 ± 0.92	Experimental: 0.17 ±0.439; Control: 0.01 ±0.527	Used post results instead of change from pre- to post-intervention. Appear to have used kb as the measure of telomere length.
Werner, et al. B[13]	Experimental: 5.5 ± 1.2; Control: 5.2 ± 0.92	Experimental: 0.19 ±0.425; Control: 0.01 ±0.527	As above.
Werner, et al. C[13]	Experimental: 5.5 ± 0.8; Control: 5.2 ± 0.92	Experimental: 0.01 ±0.6; Control: 0.01 ±0.527	As above.
Sanchez-Gonzalez, et al.[8]	Experimental: 4.81 ± 3.16; Control: 1.81 ± 2.34	Experimental: 2.51 ±2.86; Control: -2.07 ±5.27	Strange results comparably.

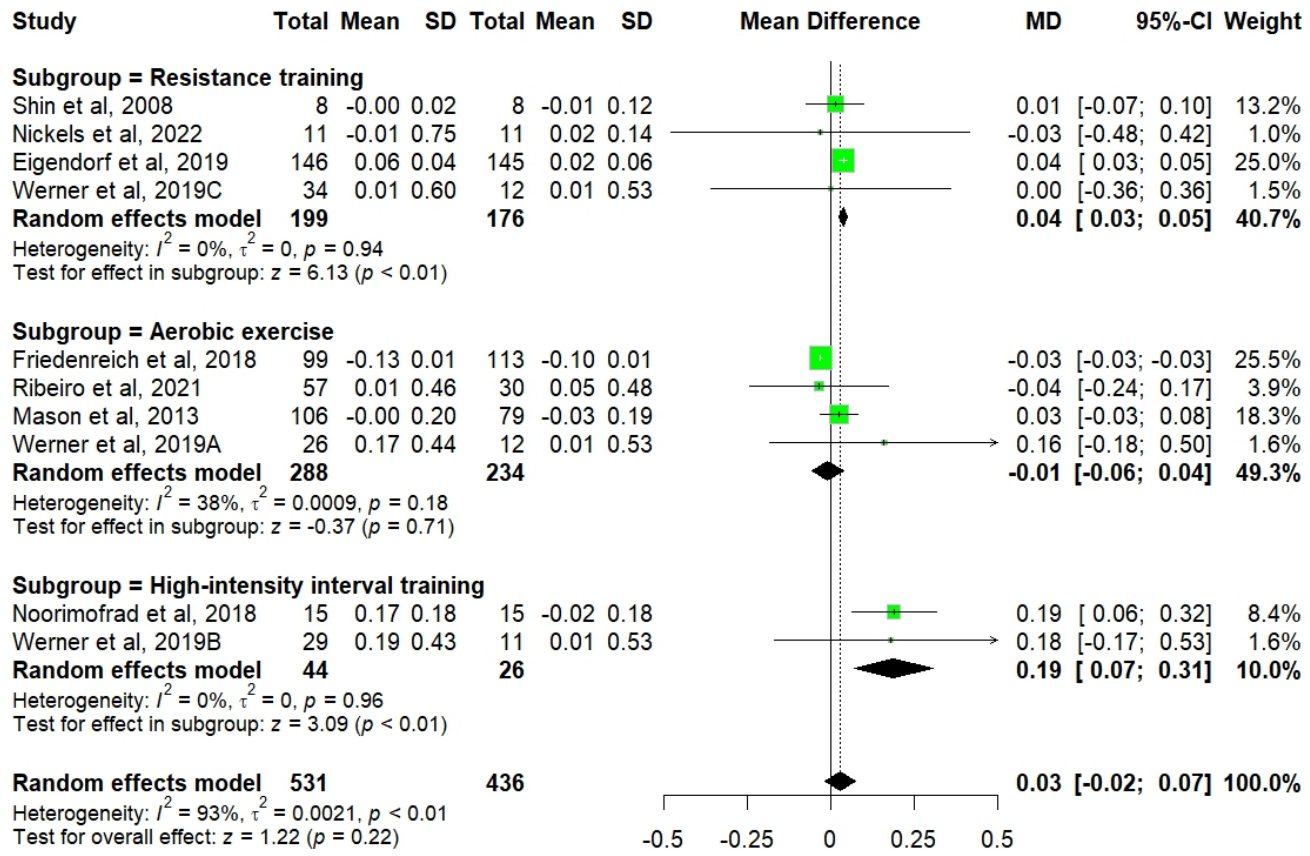

Figure 8: Forest plot for subgroup analysis using mean changes between pre- and post-intervention, produced with forest function using the meta package in R