Meta-Analysis Statistical Methods

For meta-analysis of continuous outcome data such as physical measures (activities of daily living, muscle strength, appendicular lean mass), mobility (such as gait speed, timed up and go test, chair sit & stand test, balance test, short physical performance battery), frailty, diet quality and quality of life, we utilized change from baseline to immediate post-treatment data (means, standard deviations) for both intervention and control groups to generate the summary measures of effect in the form of standardized mean difference (SMD). The SMD was used as a summary statistic because the studies in this systematic review often assessed the same outcome measured in a variety of ways (i.e. physical function measured as stair climb, balance test, gait speed, chair rise repetition, sit-to-stand test, SPPB, gait speed, TUG test etc.). In this situation, it was necessary to standardize the results of the studies before they could be compared across studies or combined in a quantitative synthesis. SMDs were estimated using Hedges' g approach (also known as bias corrected effect size). The SMDs of 0.2-0.5, 0.5-0.8, and >0.8 are considered as small, medium, and large effects respectively.

We used a random effects multi-level meta-analytic approach to account for dependency between effect sizes i.e. the correlation between effect sizes due to multiple measures or submeasures of same outcome with-in a study or comparison of multiple interventions to a single control group. In such cases, multiples measures and comparisons from same study are nested within studies first and variance in observed effect sizes is decomposed into sampling variance, with-in study variance and between-study variance to account for intracluster (or intraclass) correlation in the true effects. For pooling of mobility performance measures, the direction of effect was adjusted to ensure consistency of desirable outcome responses i.e. reduction in gait speed measured in seconds reflects a better outcome, whereas, an increase in gait speed measured in meter / second reflects a positive outcome. Similarly, a reduction in sit-to-stand test (seconds), 5 Chair repetition test (seconds) and timed up and go test (seconds) is desirable, whereas, an increase in one-leg standing test (seconds) and Short Physical Performance Battery (SPPB) score reflects a positive outcome.

Appendix 3, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

For dichotomous outcomes of interest such frailty assessed using criteria cut-offs, we utilized number of events at post-intervention to generate the summary measures of effect in the form of risk ratio (RR) using DerSimonian and Laird random effects models with Mantel-Haenszel method.

The Cochran's Q (α=0.05) was employed to detect statistical heterogeneity and I^2 statistic to quantify the magnitude of statistical heterogeneity between studies where I^2 30% to 60% represents moderate and I^2 60% to 90% represents substantial heterogeneity across studies. Publication bias was assessed using funnel plots. The funnel plots are based on multi-level meta-analysis where multiple effect estimates were included for some studies and since sample size would be similar across the estimates from the same study, the resulting standard error would be the same. For consistency and to avoid confusion, the format of funnel plots was kept consistent with that of standard reporting in existing literature. All analyses were performed using R software (metafor, and dmetar packages) and GRADEpro Guideline Development Tool software packages.

Appendix 3, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.