ORIGINAL ARTICLE

Multidimensional functional fitness response after 12 weeks of multicomponent exercise in older women: Magnitude of change and age-related heterogeneity

Jaime Andrés Laguna Villa. *Faculty of Educational and Sports Science. National School of Sports. Cali, Colombia.*
Bryan Janner Payan Gómez. *Faculty of Educational and Sports Science. National School of Sports. Cali, Colombia.*
Javier Gaviria Chavarro. *Faculty of Basic Sciences. University Santiago de Cali. Cali, Colombia.*
Miguel Ángel Gómez García. *University of Valle. Cali, Colombia.*
Julián Andrés Galeano Virgen. *Faculty of Educational and Sports Science. National School of Sports. Cali, Colombia.*
Isabel Cristina Rojas-Padilla ✉ **.** *Basic Sciences Center. National School of Sports. Cali, Colombia.*

ABSTRACT

Functional fitness is integral to independent living in older age, but the magnitude and age-related variation of functional changes observed after exercise programs remain uncertain. Purpose: To quantify paired changes in six Senior Fitness Test outcomes over 12 weeks and examine whether their magnitudes differed among six age groups of women aged 60–89 years. Methods: In this single-group pre–post study, 150 women completed paired assessments. Mean changes and Cohen's d_z were reported with bias-corrected and accelerated (BCa) bootstrap 95% confidence intervals. Age-group differences in observed change were examined using HC3-robust regression, with Welch analysis of variance (ANOVA) as a sensitivity analysis and Holm adjustment for multiple testing. Results: All six outcomes showed favourable mean temporal differences. The largest standardized difference occurred for the 30-s Chair Stand (+3.89 repetitions; BCa 95% CI [3.47, 4.27]; d_z = 1.55; BCa 95% CI [1.30, 1.84]). No age-group comparison of observed change remained statistically significant after Holm adjustment (adjusted $p \geq .392$). Conclusions: Favourable paired changes were documented, but age-related heterogeneity remains uncertain, particularly in the smallest age strata. Without comparison conditions, changes cannot be attributed causally to the intervention or interpreted as equivalent across age groups.

Keywords: performance analysis, aging, functional capacity, physical activity, physical performance, senior fitness test.

Cite this article as:
Laguna Villa, J. A., Payán Gómez, B. J., Gaviria Chavarro, J., Gómez García, M. A., Galeano Virgen, J. D., & Rojas-Padilla, I. C. (2027). Multidimensional functional fitness response after 12 weeks of multicomponent exercise in older women: Magnitude of change and age-related heterogeneity. *Journal of Human Sport and Exercise, 22*(1), 249-261. https://doi.org/10.55860/8j78np23

✉ **Corresponding author.** *Basic Sciences Center. National School of Sports. Cali, Colombia..*
E-mail: isabelcrojasp@gmail.com

Submitted for publication September 23, 2026.
Accepted for publication November 02, 2026.
Published November 24, 2026.
Journal of Human Sport and Exercise. ISSN 1988-5202.

©Asociación Española de Análisis del Rendimiento Deportivo. Alicante. Spain.

doi: https://doi.org/10.55860/8j78np23

INTRODUCTION

Functional fitness is the physiological capacity required to safely and independently perform the usual activities of daily living. In older adults, it depends on integrating muscle strength, aerobic endurance, flexibility, agility, and dynamic balance—components directly linked to physical independence. The Senior Fitness Test (SFT) assesses these domains in community-dwelling older adults and includes criteria related to maintaining functional independence (Rikli & Jones, 1999, 2013). In Colombia, the Spanish version of the test has demonstrated adequate reliability and reproducibility in adults aged 60 to 90 (Cobo-Mejía et al., 2016).

The progressive decline in these abilities constitutes a significant health problem because it leads to mobility limitations, difficulties with daily activities, and a loss of independence. Its importance increases with an aging population and the concentration of disease and disability in older adults. Prince et al. (2015) estimated that approximately 23% of the global burden of disease was attributable to disorders present in people aged 60 and older and highlighted that, in several age-related conditions, a significant proportion of that burden stems from disability. In this context, maintaining strength, cardiorespiratory fitness, and mobility is important not only for physical performance but also for sustaining independence and participation in daily life (Izquierdo et al., 2021).

Age-related functional decline is not entirely inevitable and can be modified through structured physical exercise. Multicomponent training combines strength, aerobic endurance, balance, mobility, and flexibility, making it an effective approach to address multiple determinants of physical function simultaneously. International recommendations for older adults support this combination and emphasize progression tailored to individual ability (Izquierdo et al., 2021). Furthermore, a systematic review and meta-analysis of 13 studies involving 808 older adults found that this training improves overall physical performance, upper- and lower-body strength, walking speed, and aerobic capacity compared with no training (Labata-Lezaun et al., 2023).

Single-group pre–post evaluations have also documented functional changes in community-dwelling older adults, including a large study in Cali using the Senior Fitness Test (Buriticá-Marín et al., 2023). Other observational exercise cohorts have characterized heterogeneity in trajectories or patterns of change (Ho et al., 2023). These precedents support the descriptive use of repeated assessments but do not provide an untreated counterfactual or establish the efficacy of a given intervention.

Evidence concerning age-related variability in functional change following exercise programs is mixed. Toraman and Sahin (2004) reported favourable changes in adults aged 60–73 and 74–86 years; Arai et al. (2009) did not find consistent age-related associations for many changes after exercise. Gonçalves et al. (2024) reported outcome-specific patterns among 141 older women across three broader age groups. More detailed age stratification can describe observed changes and uncertainty, but it cannot identify age-specific intervention effects without an appropriate comparison group. Therefore, this study aimed to quantify within-person pre–post changes in six Senior Fitness Test outcomes after 12 weeks of multicomponent exercise in women aged 60–89 years and to examine whether these changes differed across six five-year age groups.

METHODS

Design and participants

A longitudinal, single-group pre–post study was conducted in community-based physical activity programs in Cali, Colombia. The same women were assessed before the program and after 12 weeks. Baseline

 © 2027 ARD Asociación Española

measurements provided an individual reference for describing subsequent change; they did not constitute an untreated control condition. Accordingly, observed temporal differences were evaluated rather than causal intervention effects. The reporting followed recommendations for nonrandomized evaluations and intervention descriptions (Des Jarlais et al., 2004; Hoffmann et al., 2014).

The analytical sample consisted of 150 women aged 60 to 89 years, all with complete pre- and post-measurements. Women aged 60 years or older who were able to understand the instructions and perform the functional tests independently or with minimal assistance, and who provided voluntary participation and informed consent, were included. Exclusion criteria included acute or decompensated medical conditions that contraindicated exercise; severe musculoskeletal, neurological, or cardiovascular limitations that prevented safe performance; simultaneous participation in another structured exercise program; and attendance of less than 85% of the scheduled sessions.

Sample

The sample-size planning calculation used G*Power 3.1 for a two-sided paired *t* test, assuming $d_z = 0.25$, $\alpha = .05$, and 80% power; the corresponding analytical minimum was 128 complete pairs. Assuming 15% attrition yielded a recruitment target of $128/(1 - 0.15) = 150.6$, rounded up to 151 participants. The sample analysed comprised 150 complete pairs: one fewer than the recruitment target but above the minimum for the pooled paired comparison. This calculation was not designed to provide adequate statistical power for six-group comparisons of change, particularly in the older strata (Faul et al., 2009).

Instruments

Functional fitness was assessed using six components of the Senior Fitness Test, a battery of tests developed for community-dwelling older adults and focused on abilities related to physical independence (Rikli & Jones, 1999). The following were recorded: (1) 30-Second Chair Stand, expressed in repetitions, as an indicator of lower-body functional strength; (2) 30-Second Arm Curl, in repetitions, as an indicator of upper-body functional strength; (3) a 2-Minute Step, expressed in the number of valid steps/strides, as an estimate of functional aerobic endurance; (4) Chair Sit-and-Reach from a seated position in a chair, in centimetres, as an indicator of lower-limb flexibility; (5) Back Scratch, bringing the hands together behind the back, in centimetres, as an indicator of upper-limb flexibility; and (6) 8-Foot Up-and-Go (2.44 m), returning, and sitting down, in seconds, as an indicator of agility and dynamic balance. In the flexibility tests, higher values indicate better performance; in the timed test, shorter times indicate better performance. The Spanish version of the SFT has demonstrated adequate reliability among older adults in Colombia (Cobo-Mejía et al., 2016).

Procedure

The 12-week multicomponent program comprised 36 scheduled sessions (three per week), organized into three four-week mesocycles (Table 1). Session content combined mobility, resistance exercise, aerobic/locomotor tasks, balance, agility, and flexibility. Staff prescribed intensity using the Borg CR10 scale (0–10), and the session plans specified exercise selection, work duration, sets, and recovery intervals. Experienced staff monitored exercise techniques and tolerance and modified or suspended activities in response to pain, dizziness, unusual dyspnoea, or excessive fatigue (Izquierdo et al., 2021).

The loading plan followed an undulating weekly pattern. The three mesocycles progressed based on individual adaptation and progressive loading, with training demands adjusted according to participants' tolerance and prescribed perceived exertion. Weekly training volume remained relatively stable, with modest fluctuations rather than a sustained linear increase, while prescribed intensity varied from week to week with an overall progressive trend on the Borg CR10 scale. The initial phase prescribed brief 30-s pauses between

activities. During the main phase, planned recovery intervals between work bouts generally ranged from 1 to 6 min, depending on exercise demands. This combination of fluctuating weekly demands, stable volume, progressive intensity, and planned recovery was intended to manage fatigue and minimize interference with participants' usual daily activities. Table 2 presents a qualitative summary of the loading strategy across the three mesocycles.

Participants were required to attend at least 85% of the 36 scheduled sessions (at least 31 sessions) to be eligible for inclusion in the final analysis.

Table 1. Structure of the three weekly multimodal sessions in the multicomponent program.

Session	Initial phase	Core phase	Final phase	Duration / target Borg CR10
1	Joint mobility: 4 min, Borg 1; 30-s microbreak. Dynamic stretching: 3 min, Borg 2; 30-s microbreak.	Resistance bands: 5 × 1 min, Borg 4; rest 4 min. Bridges: 5 × 1 min, Borg 3; rest 4 min. Dumbbell biceps curls: 6 × 1 min, Borg 4; rest 5 min. Mat abdominal exercises: 6 × 1 min, Borg 5; rest 5 min. Stick presses: 5 × 1 min, Borg 2; rest 4 min.	Head, trunk and limb mobility: 3 min, Borg 0. Stretching: 5 min, Borg 1.	65 min / Borg 4
2	Joint mobility: 2 min, Borg 1; 30-s microbreak. Dynamic stretching: 4 min, Borg 2; 30-s microbreak.	Bench press with 2–3 lb (approximately 0.91–1.36 kg): 5 × 1 min, Borg 4; rest 4 min. Progressive walking: 5 × 1.5 min, Borg 3; rest 4 min. Modified planks: 5 × 1 min, Borg 6; rest 4 min. Steps with 2–3 lb (approximately 0.91–1.36 kg): 4 × 1 min, Borg 5; rest 3 min. Low-impact skipping: 5 × 1 min, Borg 5; rest 4 min.	Head, trunk and limb mobility: 3 min, Borg 0. Stretching: 5 min, Borg 1.	59 min / Borg 5
3	Joint mobility: 2 min, Borg 1; 30-s microbreak. Dynamic stretching: 4 min, Borg 2; 30-s microbreak.	Glute bridge: 6 × 1 min, Borg 7; rest 5 min. Superman exercise on mat: 5 × 1 min, Borg 6; rest 4 min. Lunges: 5 × 1 min, Borg 7; rest 4 min. Abdominal exercises: 5 × 1 min, Borg 6; rest 4 min. Low-impact squats: 4 × 1 min, Borg 7; rest 3 min.	Head, trunk and limb mobility: 3 min, Borg 0. Stretching: 5 min, Borg 1.	60 min / Borg 7

Note. Borg CR10 values represent prescribed perceived-exertion targets (0–10), not measured session averages. Session duration includes planned exercise and recovery periods.

Table 2. Qualitative summary of the loading strategy across the three mesocycles.

Mesocycle	Intensity and loading strategy	Volume and recovery
1 (weeks 1–4)	Initial application of the planned multimodal sessions. Intensity was prescribed using Borg CR10, with demands adjusted according to participants' tolerance.	Weekly volume was planned to remain relatively stable with modest fluctuations. Brief 30-s pauses were used between activities in the initial phase; main-phase recovery intervals depended on exercise demands.
2 (weeks 5–8)	Progression continued according to individual adaptation and progressive-loading principles, while retaining an undulating weekly distribution of training demands. Prescribed intensity contributed to the overall progressive trend.	Weekly volume remained relatively stable rather than increasing linearly. Recovery intervals remained planned according to exercise demands, generally within the 1–6-min range during the main phase.
3 (weeks 9–12)	The progressive-loading strategy continued with intensity prescribed on Borg CR10, preserving week-to-week variation and adjusting demands according to participants' tolerance.	Weekly volume remained relatively stable with planned fluctuations. Recovery between work bouts remained prescribed according to exercise demands, generally within the 1–6-min range during the main phase.

 © 2027 ARD Asociación Española

Statistical analysis

The analysis used paired measurements from 150 participants. Categorical variables were summarized with frequencies and percentages, and continuous outcomes with means, standard deviations and the distributions of individual changes. For each SFT component, Δ was calculated as follow-up minus baseline; positive differences indicate more favourable performance in five tests, whereas negative differences indicate improvement for 8-Foot Up-and-Go.

The mean paired difference and Cohen's d_z (mean difference divided by its standard deviation) were accompanied by bias-corrected and accelerated (BCa) bootstrap 95% confidence intervals based on 10,000 resamples of participants, using the random seed 20261026. Paired two-sided *t*-tests assessed the overall mean difference, and Wilcoxon signed-rank tests assessed distributional sensitivity. Resampling preserved the baseline–follow-up pairing. Bootstrap and nonparametric procedures quantify statistical uncertainty under their assumptions but do not eliminate learning effects, selection bias or time-varying confounding (Efron, 1987).

To address the study objective, the primary between-age-group estimated was the difference in mean observed within-person change (Δ = post − pre). Separate linear models of change were fitted for each functional outcome, treating age group (60–64, 65–69, 70–74, 75–79, 80–84 and 85–89 years) as a categorical explanatory variable. The overall equality of group mean changes was assessed using a five-degree-of-freedom robust Wald *F* test with HC3 heteroscedasticity-consistent covariance, and Welch analysis of variance (ANOVA) was performed as a sensitivity analysis (MacKinnon & White, 1985). The model was specified to describe observed age-stratified change rather than to estimate causal modification of an exercise effect. Age was analysed in the recorded five-year categories. Cluster-robust standard errors were not applied.

Multiplicity control used separate families of six tests: one family for overall paired pre–post differences and one for the six omnibus age-group comparisons of observed change. The Holm sequential procedure was applied within each family (Holm, 1979). All inferential tests were two-sided with α = .05. Participant IDs, paired records, time coding, and age-group membership were checked before the analyses. BCa resampling used 10,000 participant-level samples and a fixed random seed (20261026). The analyses were reproduced in Python using NumPy, SciPy, and statsmodels.

Ethical considerations

The study followed the Declaration of Helsinki and Colombian regulations for research involving human subjects (Ministry of Health Resolution 8430 of 1993; Ministerio de Salud, 1993). The University of Santiago de Cali Ethics Committee approved the protocol on August 11, 2023. All participants signed an informed consent form before the assessments and intervention.

RESULTS

The analysis included 150 women with paired pre- and post-12-week program assessments, yielding 300 measurements. None of the six SFT outcomes had missing value among these participants. The 60–64-year group was the largest (*n* = 51; 34.0%), whereas the 80–84-year group was the smallest (*n* = 12; 8.0%; Table 3).

The six paired outcomes showed favourable mean temporal differences (Table 4 and Figure 1). The 30-s Chair Stand had the largest standardized paired change (+3.89 repetitions; d_z = 1.55); the 2-min Step and

30-s Arm Curl showed intermediate differences ($d_z \approx 0.55$), while flexibility showed smaller differences. The mean 8-Foot Up-and-Go time decreased by 0.69 s. All six paired *t*-tests remained significant after Holm adjustment. These findings do not isolate effects attributable to supervised exercise.

Table 3. Participants by age group.

Age group (years)	*n*	%
60–64	51	34.0
65–69	26	17.3
70–74	28	18.7
75–79	18	12.0
80–84	12	8.0
85–89	15	10.0
Total	150	100.0

Table 4. Overall pre–post change in senior fitness test components.

Outcome	Baseline mean ± *SD*	Follow-up mean ± *SD*	Mean Δ	BCa 95% CI for Δ	d_z	BCa 95% CI for d_z	Holm-adjusted *p*
30-s Chair Stand (rep)	12.24 ± 2.84	16.13 ± 2.12	+3.89	3.47 to 4.27	+1.55	1.30 to 1.84	< .001
30-s Arm Curl (rep)	15.30 ± 2.99	17.22 ± 2.66	+1.92	1.35 to 2.49	+0.53	0.36 to 0.69	< .001
2-min Step (steps)	90.19 ± 14.16	98.84 ± 11.47	+8.65	6.11 to 11.08	+0.55	0.38 to 0.72	< .001
Chair Sit-and-Reach (cm)	1.68 ± 2.15	2.62 ± 2.14	+0.94	0.45 to 1.42	+0.31	0.15 to 0.49	< .001
Back Scratch (cm)	-1.07 ± 2.47	-0.46 ± 2.14	+0.61	0.20 to 1.03	+0.23	0.07 to 0.39	.005
8-Foot Up-and-Go (s)	5.64 ± 1.59	4.95 ± 1.33	-0.69	-0.95 to -0.45	-0.45	-0.61 to -0.28	< .001

Note. Δ = postintervention − preintervention; a negative Δ denotes improvement in 8-Foot Up-and-Go. *SD* = standard deviation; CI = confidence interval; BCa = bias-corrected and accelerated; d_z = Cohen's standardized paired mean difference; rep = repetitions. BCa 95% confidence intervals were estimated using 10,000 participant-level bootstrap resamples (random seed: 20261026). Holm-adjusted *p*-values refer to the family of six two-sided paired *t* tests. The Wilcoxon signed-rank sensitivity analyses supported the direction and statistical significance of the six overall differences.

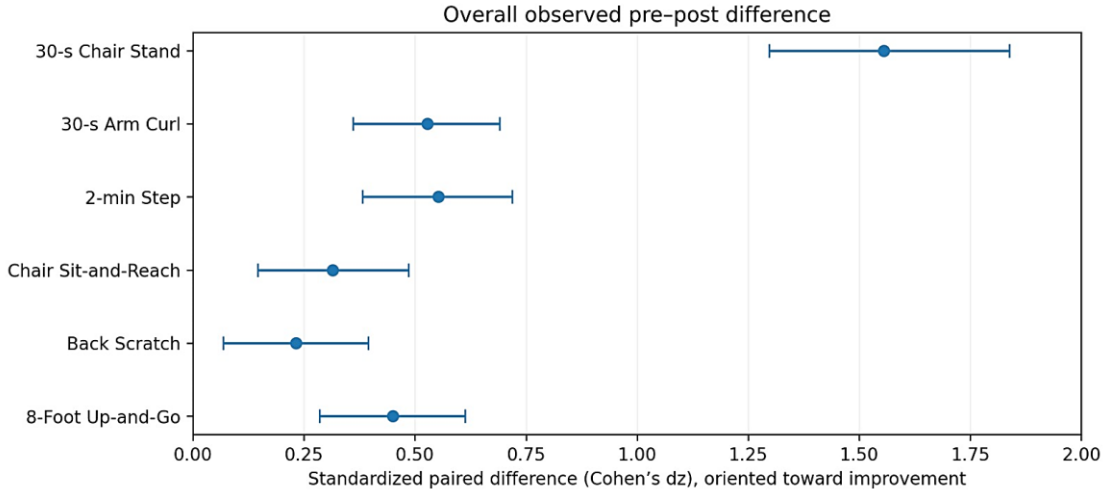


Figure 1. Standardized magnitude of global functional change.

Age-specific raw means changes varied but did not display a consistent monotonic gradient (Table 5 and Figure 2). For 30-s Chair Stand, women aged 85–89 had a mean difference of +4.80 repetitions compared with +3.14 among women aged 70–74. The HC3 age-group omnibus test yielded $p = .068$ for Chair Stand and $p = .065$ for Arm Curl; each increased to $p = .392$ after Holm adjustment. None of the six change-score age-group tests met the multiplicity-adjusted threshold, in agreement with Welch ANOVA. This absence of evidence is not an equivalence test and does not rule out Type II error, particularly in the oldest strata.

 © 2027 ARD Asociación Española

Table 5. Mean change (95% CI) by age group and HC3-robust heterogeneity tests.

Outcome	60–64	65–69	70–74	75–79
30-s Chair Stand (rep)	+3.88 [3.04, 4.73]	+4.04 [3.04, 5.04]	+3.14 [2.18, 4.11]	+4.33 [3.43, 5.24]
30-s Arm Curl (rep)	+0.96 [-0.02, 1.94]	+2.85 [1.34, 4.35]	+2.68 [1.19, 4.17]	+3.17 [1.72, 4.62]
2-min Step (steps)	+8.88 [4.59, 13.18]	+5.92 [-1.03, 12.88]	+10.07 [3.97, 16.17]	+11.17 [2.65, 19.68]
Chair Sit-and-Reach (cm)	+1.01 [0.18, 1.83]	+0.59 [-0.83, 2.00]	+1.77 [0.55, 2.98]	-0.03 [-1.46, 1.40]
Back Scratch (cm)	+0.20 [-0.54, 0.94]	-0.26 [-1.16, 0.64]	+1.02 [0.13, 1.92]	+1.03 [-0.22, 2.29]
8-Foot Up-and-Go (s)	-0.41 [-0.76, -0.05]	-0.58 [-1.06, -0.10]	-0.92 [-1.56, -0.29]	-1.07 [-1.85, -0.28]
Outcome	**80–84**	**85–89**	***p* HC3**	***p* Holm**
30-s Chair Stand (rep)	+3.50 [2.13, 4.87]	+4.80 [4.17, 5.43]	.068	.392
30-s Arm Curl (rep)	+1.83 [-0.14, 3.80]	+0.73 [-1.57, 3.04]	.065	.392
2-min Step (steps)	+4.08 [-4.71, 12.87]	+10.53 [2.41, 18.66]	.752	1.000
Chair Sit-and-Reach (cm)	+1.25 [0.01, 2.49]	+0.69 [-0.80, 2.19]	.482	1.000
Back Scratch (cm)	+1.38 [-0.74, 3.50]	+1.61 [-0.08, 3.29]	.149	.597
8-Foot Up-and-Go (s)	-0.95 [-2.10, 0.20]	-0.77 [-2.09, 0.54]	.534	1.000

Note. Values are post–pre mean changes and descriptive 95% CIs. CI = confidence interval; HC3 = heteroscedasticity-consistent covariance estimator, type 3; ANOVA = analysis of variance; rep = repetitions. HC3 refers to the heteroscedasticity-consistent global *F* test for six age groups, with *p*-values adjusted for six outcomes using Holm's procedure. Negative changes indicate favourable performance for 8-Foot Up-and-Go. Welch ANOVA *p*-values from the source were .078, .081, .747, .489, .166 and .546, in outcome order.

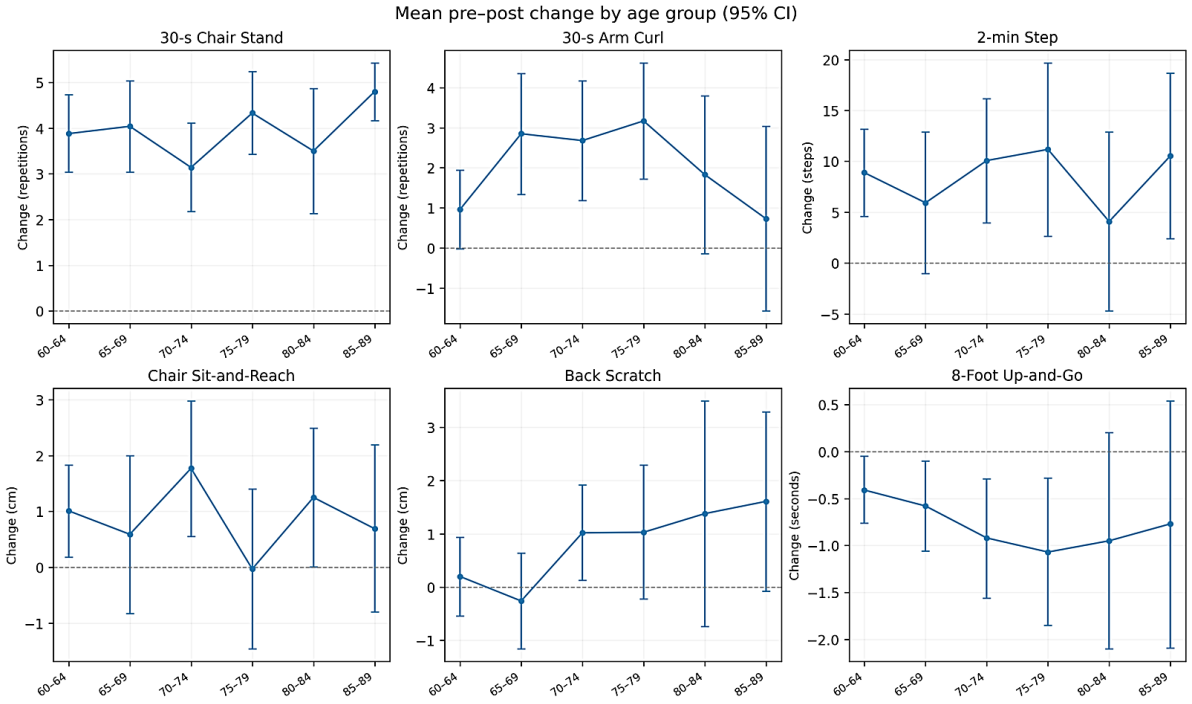


Figure 2. Mean pre–post change by age group.

Baseline functional status varied substantially between age groups before the exercise program; for instance, Chair Stand means were 13.43 repetitions in women aged 60–64 and 7.73 in women aged 85–89. Consequently, interpret observed age-group changes in the context of different starting levels, possible regression to the mean, and the uncertainty created by small age strata. These baseline differences do not justify interpreting age-group contrasts of observed change as causal differences in adaptability.

The proportions of participants whose paired change had a favourable sign were 133/150 (88.7%) for the 30-s Chair Stand, 91/150 (60.7%) for the 30-s Arm Curl, 104/150 (69.3%) for the 2-min Step, 90/150 (60.0%) for

Chair Sit-and-Reach, 89/150 (59.3%) for Back Scratch, and 91/150 (60.7%) for 8-Foot Up-and-Go. These are directional proportions only, not clinically important responder rates: no outcome-specific minimal detectable or clinically important change criterion was used.

Concentrations of favourable values were observed for some outcomes at follow-up. These patterns might reflect score discretization or ceiling-like behaviour; however, a formal ceiling effect relative to test-specific measurement limits was not established. All observations were retained; bootstrap intervals and heteroscedasticity-sensitive procedures were used to assess the stability of descriptive inferences.

DISCUSSION

The principal finding was a favourable paired temporal difference in each of the six SFT outcomes between baseline and 12 weeks. The largest standardized change was observed in the 30-s Chair Stand, whereas flexibility outcomes showed smaller changes. The findings align with evidence from controlled studies of multicomponent exercise, but concordance across studies does not establish that the intervention caused the observed differences in an uncontrolled cohort (Labata-Lezaun et al., 2023; Toraman et al., 2004).

The change in the 30-s Chair Stand warrants careful interpretation: an increase of 3.89 repetitions and d_z = 1.55 represent a substantial standardized paired difference in a community program. Because d_z uses the standard deviation of individual changes as its denominator, it cannot be directly compared with standardized effects between independent groups. Familiarity with test instructions, improved pacing, confidence, and movement strategy could have contributed to the repeated test performance. Selection associated with the ≥85% attendance criterion, baseline differences, and regression to the mean are further possible explanations. The study did not include an untreated control group or additional baseline assessments to quantify familiarization; consequently, the study cannot separate physiological adaptation from learning and other temporal influences. A large d_z does not in itself establish clinical importance.

All six SFT outcomes were performance-based field tests administered on two occasions. Repetition of the same procedures could have improved test execution independently of training adaptation. This possible learning effect is particularly important when interpreting the large Chair Stand difference and is not resolved by bootstrapping, HC3 estimation or the use of alternative hypothesis tests.

The descriptive use of repeated functional assessments is not unique to the present study. Buriticá-Marín et al. (2023) described SFT changes within a community-based cohort in Cali; Chang et al. (2018) employed a single-group pre–post design to examine physical function after whole-body vibration in older adults with sarcopenia; and Ho et al. (2023) investigated heterogeneity in functional change in an exercise cohort. Their populations and interventions differ from those examined in the present study, and their publications do not support causal interpretation. Rather, these studies illustrate the legitimate descriptive scope of single-group longitudinal assessments while leaving open the counterfactual question of what would have occurred without the program.

The intermediate paired changes in the 30-s Arm Curl and 2-min Step, along with the smaller changes in flexibility, broadly align with the outcome-specific patterns described in the synthesis by Toosi et al. (2026). However, standardized effects from controlled trials and d_z from the present paired cohort are not numerically interchangeable. Previous research has documented improvements in strength, walking performance, and aerobic capacity after multicomponent training (Labata-Lezaun et al., 2023), including in frail older

 © 2027 ARD Asociación Española

populations (Cadore et al., 2014); those findings provide physiological context, not causal identification for the present estimates.

The findings can also be contextualized using evidence from frail older adults. In a meta-analysis of 28 randomized trials involving 4,857 participants, Yang et al. (2024) reported improvements in muscle strength, gait speed, balance, the Short Physical Performance Battery, and Timed Up and Go (TUG) following multicomponent exercise. Their subgroup analysis also identified favourable results for interventions lasting approximately 12 weeks. These findings are relevant to functional mobility, although TUG and the SFT 8-Foot Up-and-Go have different protocols and should not be treated as identical outcomes.

Similarly, Wang et al. (2024), in a systematic review and meta-analysis of 19 studies involving community-dwelling older adults with frailty, reported improvements in physical function, including Timed Up and Go performance. These external findings provide a broader comparison for the present mobility results, but differences in test protocol and participant characteristics prevent direct comparison of effect magnitudes.

The second objective was to examine whether the magnitude of observed paired change differed across six age strata. None of the HC3-robust omnibus age-group tests remained statistically significant after Holm adjustment, with Welch analyses yielding the same general inference. Interpret this finding in the context of mixed previous evidence. Toraman and Sahin (2004) found no statistically significant difference in absolute functional change between broad age groups of 60–73 and 74–86 years. Arai et al. (2009) found that most changes after a three-month program were not correlated with age, although grip strength and functional reach showed weak negative correlations. By contrast, Gonçalves et al. (2024) reported outcome-specific differences after a ten-month multicomponent program, with several measures largely maintained among women aged 80 years or older. Differences in study duration, age strata, outcomes, and initial function limit direct comparability. Moreover, the 80–84- and 85–89-year groups included only 12 and 15 participants, respectively. Thus, the nonsignificant results do not establish equivalent change across ages, and type II error remains plausible.

Interpretation of age-specific observed changes also depends on initial function. For example, baseline Chair Stand means differed between the 60–64-year (13.43 repetitions) and 85–89-year (7.73 repetitions) strata; their observed baseline ranges scarcely overlapped. Unadjusted mean change was retained as the descriptive estimate specified by the objective. An analysis of covariance (ANCOVA) of follow-up scores conditional on baseline would target a different, model-dependent contrast and may be sensitive to poor baseline overlap, nonparallel slopes, and assumptions about pre-existing age groups (van Breukelen, 2013). Neither model alone establishes causal differences in adaptation, and the scientific estimate not statistical significance, should determine the choice between them.

Change-score regression operationalizes the present question about differences in observed temporal change; nevertheless, it remains vulnerable to baseline-related mechanisms, measurement error, and unmeasured time-varying influences. Change-score analyses should not be interpreted as estimating causal effects from observational data (Tennant et al., 2022). Studies evaluating differential training efficacy require an appropriate comparison condition, better representation of older age strata, and, where feasible, repeated baseline measurements and continuously recorded age.

BCa intervals, HC3 covariance estimation, Holm correction and Welch/Wilcoxon sensitivity analyses improve the precision and robustness of descriptive inference under their respective assumptions; none replaces an untreated counterfactual. HC3 accommodates heteroscedasticity but does not have a correct correlation

among participants within shared exercise groups. Neither cluster-adjusted inference nor multivariable adjustment for clinical or behavioural characteristics was performed in the reported models.

Several limitations must be considered. First, because all women received the program and were evaluated twice, changes due to test familiarization, regression to the mean, secular or seasonal influences, and concurrent changes in habitual activity cannot be separated from program-related adaptation. Second, we analysed age in five-year categories, and the small numbers of women in the oldest groups limited precision; the sample-size calculation addressed the pooled paired comparison, not age-specific heterogeneity. Third, differences in baseline function and limited overlap in some outcomes constrain conditional age comparisons, while raw change-score comparisons remain descriptive. Fourth, the primary models did not assess potential within-site dependence or residual confounding by health, medication use, nutrition, socioeconomic circumstances, or habitual activity. Finally, selection based on ≥85% attendance limits generalization to women with lower participation, men, institutionalized populations, and older adults with marked frailty.

Study strengths include 150 women with complete paired observations on six standardized SFT outcomes, an age range spanning 60–89 years, and a transparent analysis emphasizing magnitudes, confidence intervals, multiplicity control, and heteroscedasticity-sensitive sensitivity checks. The 150 paired observations exceeded the analytical minimum of 128, although the rounded recruitment target was 151. The sample-size plan did not ensure adequate power to detect modest differences in change across six unequally sized age groups.

CONCLUSIONS

Favourable paired temporal differences were observed across six Senior Fitness Test outcomes in 150 community-dwelling women aged 60–89 years after a 12-week multicomponent exercise program. Chair Stand exhibited the largest standardized difference. Without a comparison condition or measurement-familiarization control, these findings do not provide evidence of intervention efficacy.

Across the six age groups, comparisons of observed changes provided no statistically robust evidence of heterogeneity after adjustment for multiple testing. This does not establish equivalent functional change, rule out type II error, or demonstrate that training effects are independent of chronological age. The results should remain framed as descriptive observations. Future controlled studies should assess test familiarization, collect continuous age and baseline clinical data, document individual program exposure, and recruit sufficient participants in the oldest strata to estimate age-related effect modification with greater precision.

AUTHOR CONTRIBUTIONS

All authors meet the criteria for authorship in accordance with established ethical guidelines. Contributions are specified according to the CRediT (Contributor Roles Taxonomy) as follows:

Conceptualisation: Javier Gaviria Chavarro, Isabel Cristina Rojas-Padilla. Methodology: Javier Gaviria Chavarro, Isabel Cristina Rojas-Padilla. Formal analysis: Javier Gaviria Chavarro, Isabel Cristina Rojas-Padilla. Investigation: Javier Gaviria Chavarro, Isabel Cristina Rojas-Padilla, Julián Andrés Galeano Virgen, Bryan Janner Payan Gómez. Data curation: Javier Gaviria Chavarro, Isabel Cristina Rojas-Padilla, Julián Andrés Galeano Virgen, Bryan Janner Payan Gómez. Writing – original draft: Jaime Andrés Laguna Villa, Bryan Janner Payan Gómez, Javier Gaviria Chavarro, Miguel Ángel Gómez García, Julián Andrés Galeano Virgen, Isabel Cristina Rojas-Padilla. Writing – review & editing: Jaime Andrés Laguna Villa, Bryan Janner

 © 2027 ARD Asociación Española

Payan Gómez, Javier Gaviria Chavarro, Miguel Ángel Gómez García, Julián Andrés Galeano Virgen, Isabel Cristina Rojas-Padilla. Supervision: Javier Gaviria Chavarro, Isabel Cristina Rojas-Padilla.

All authors have critically reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

AI USE DISCLOSURE

In accordance with current publishing ethics and transparency recommendations, artificial intelligence (AI) tools were used solely to assist with translation and language editing, with the aim of improving clarity and readability. No AI tools were used in the generation of scientific content, including the study design, data collection, analysis, interpretation of results, or the formulation of conclusions. The authors retain full responsibility for the content of the manuscript and confirm its originality, integrity, and accuracy.

REFERENCES

Arai, T., Obuchi, S., Kojima, M., Nishizawa, S., Matsumoto, Y., & Inaba, Y. (2009). The relationship between age and change in physical functions after exercise intervention: Trainability of Japanese community-dwelling older elderly. *Journal of the Japanese Physical Therapy Association*, *12*(1), 1–8. https://doi.org/10.1298/jjpta.12.1

Buriticá-Marín, E. D., Daza-Arana, J. E., Jaramillo-Losada, J., Riascos-Zuñiga, A. R., & Ordoñez-Mora, L. T. (2023). Effects of a physical exercise program on the physical capacities of older adults: A quasi-experimental study. *Clinical Interventions in Aging*, *18*, 273–282. https://doi.org/10.2147/CIA.S388052

Cadore, E. L., Casas-Herrero, A., Zambom-Ferraresi, F., Idoate, F., Millor, N., Gómez, M., Rodriguez-Mañas, L., & Izquierdo, M. (2014). Multicomponent exercises including muscle power training enhance muscle mass, power output, and functional outcomes in institutionalized frail nonagenarians. *Age*, *36*(2), 773–785. https://doi.org/10.1007/s11357-013-9586-z

Chang, S.-F., Lin, P.-C., Yang, R.-S., & Yang, R.-J. (2018). The preliminary effect of whole-body vibration intervention on improving the skeletal muscle mass index, physical fitness, and quality of life among older people with sarcopenia. *BMC Geriatrics*, *18*, Article 17. https://doi.org/10.1186/s12877-018-0712-8

Cobo-Mejía, E. A., Ochoa González, M. E., Ruiz Castillo, L. Y., Vargas Niño, D. M., Sáenz Pacheco, A. M., & Sandoval-Cuellar, C. (2016). Confiabilidad del Senior Fitness Test versión en español, para población adulta mayor en Tunja-Colombia [Reliability of Senior Fitness Test version in Spanish for older people in Tunja-Colombia]. *Archivos de Medicina del* Deporte, *33*(176), 382–386.

Des Jarlais, D. C., Lyles, C., Crepaz, N., & TREND Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, *94*(3), 361–366. https://doi.org/10.2105/AJPH.94.3.361

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, *82*(397), 171–185. https://doi.org/10.1080/01621459.1987.10478410

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Gonçalves, A. K., da Silva, P. C., Griebler, E. M., da Silva, W. A., Sant Helena, D. P., Possamai, V. D., & Martins, V. F. (2024). Multicomponent physical program: Effects on physical fitness of older women of different age groups. *Research Quarterly for Exercise and Sport*, *95*(3), 759–765. https://doi.org/10.1080/02701367.2024.2306285

Ho, V., Chan, Y. H., & Merchant, R. A. (2023). Patterns of improvement in functional ability and predictors of responders to dual-task exercise: A latent class analysis. *Frontiers in Public Health*, *10*, Article 1069970. https://doi.org/10.3389/fpubh.2022.1069970

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ, 348*, Article g1687. https://doi.org/10.1136/bmj.g1687

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Izquierdo, M., Merchant, R. A., Morley, J. E., Anker, S. D., Aprahamian, I., Arai, H., Aubertin-Leheudre, M., Bernabei, R., Cadore, E. L., Cesari, M., Chen, L.-K., de Souto Barreto, P., Duque, G., Ferrucci, L., Fielding, R. A., García-Hermoso, A., Gutiérrez-Robledo, L. M., Harridge, S. D. R., Kirk, B., . . . Fiatarone Singh, M. (2021). International exercise recommendations in older adults (ICFSR): Expert consensus guidelines. *The Journal of Nutrition, Health & Aging*, *25*(7), 824–853. https://doi.org/10.1007/s12603-021-1665-8

Labata-Lezaun, N., González-Rueda, V., Llurda-Almuzara, L., López-de-Celis, C., Rodríguez-Sanz, J., Bosch, J., Vicente-Rodríguez, G., Gorczakowska, D., Araluze-Arizti, P., & Pérez-Bellmunt, A. (2023). Effectiveness of multicomponent training on physical performance in older adults: A systematic review and meta-analysis. *Archives of Gerontology and Geriatrics*, *104*, Article 104838. https://doi.org/10.1016/j.archger.2022.104838

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Ministerio de Salud. (1993, October 4). *Resolución número 8430 de 1993: Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud* [Resolution No. 8430 of 1993: Scientific, technical, and administrative standards for health research]. https://minsalud.gov.co/Normatividad_Nuevo/RESOLUCION%208430%20DE%201993.pdf

Prince, M. J., Wu, F., Guo, Y., Gutierrez Robledo, L. M., O'Donnell, M., Sullivan, R., & Yusuf, S. (2015). The burden of disease in older people and implications for health policy and practice. *The Lancet*, *385*(9967), 549–562. https://doi.org/10.1016/S0140-6736(14)61347-7

Rikli, R. E., & Jones, C. J. (1999). Development and validation of a functional fitness test for community-residing older adults. *Journal of Aging and Physical Activity*, *7*(2), 129–161. https://doi.org/10.1123/japa.7.2.129

 © 2027 ARD Asociación Española

Rikli, R. E., & Jones, C. J. (2013). Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. *The Gerontologist*, *53*(2), 255–267. https://doi.org/10.1093/geront/gns071

Tennant, P. W. G., Arnold, K. F., Ellison, G. T. H., & Gilthorpe, M. S. (2022). Analyses of "change scores" do not estimate causal effects in observational data. *International Journal of Epidemiology*, *51*(5), 1604–1615. https://doi.org/10.1093/ije/dyab050

Toosi, M., Daneshmandi, H., Mottaghitalab, M., & Zarei, H. (2026). The effects of multicomponent training on functional fitness in older adults: A systematic review and meta-analysis. *Journal of Aging and Physical Activity*. Advance online publication. https://doi.org/10.1123/japa.2025-0070

Toraman, N. F., Erman, A., & Agyar, E. (2004). Effects of multicomponent training on functional fitness in older adults. *Journal of Aging and Physical Activity*, *12*(4), 538–553. https://doi.org/10.1123/japa.12.4.538

Toraman, N. F., & Sahin, G. (2004). Age responses to multicomponent training programme in older adults. *Disability and Rehabilitation*, *26*(8), 448–454. https://doi.org/10.1080/09638280310000166 3012

van Breukelen, G. J. P. (2013). ANCOVA versus CHANGE from baseline in nonrandomized studies: The difference. *Multivariate Behavioral Research*, *48*(6), 895–922. https://doi.org/10.1080/00273171.2013.831743

Wang, H., Liang, Y., Wu, H., Shi, H., & Hou, R. (2024). Effect of multicomponent exercise intervention in community dwelling frail elderly: A systematic review and meta-analysis. *Archives of Gerontology and Geriatrics*, *126*, Article 105543. https://doi.org/10.1016/j.archger.2024.105543

Yang, X., Li, S., Xu, L., Liu, H., Li, Y., Song, X., Bao, J., Liao, S., Xi, Y., & Guo, G. (2024). Effects of multicomponent exercise on frailty status and physical function in frail older adults: A meta-analysis and systematic review. *Experimental Gerontology*, *197*, Article 112604. https://doi.org/10.1016/j.exger.2024.112604



This work is licensed under a Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0 DEED).