Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sex-stratified path-analytic patterns of exercise prescription, BMI change, and physical fitness in overweight Korean adults: National Fitness Award cohort study

Jae-Il Choi
Dankook University
Yong-Kyun Jeon
Dankook University
Jeong-Yeop Shin
Gukje Cyber University
Won-Je Cho
wjcho0425@kma.ac.kr

Korea Military Academy

Article

Keywords: overweight, body mass index, exercise prescription, physical fitness, sex differences, covariance partitioning

Posted Date: June 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9664636/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

How exercise prescription co-varies with body mass index (BMI) change and fitness change, and whether the pattern differs by sex, is poorly described at scale. We analysed 12,980 overweight Korean adults (10,173 men, 2,807 women; 19–64 y) from 1,199,471 National Fitness Award Program records (2022–2025), with paired pre/post 30–365 days apart. Repeated-measures ANCOVA and lavaan path models were fitted across quartile-based BMI-change groups; because pre/post share the same interval, a·b is read as within-interval covariance partition, not causal mediation. Inference used BCa bootstrap, FDR correction, sex × Time × Group tests, and ten sensitivity analyses (± 3% co-primary). BMI change associated with seven principal BMI-independent outcomes (FDR p < .001). In men, aerobic prescriptions associated with flexibility, power, and (via BMI) step-test $\dot{V}O_2$max. In women, no type-specific associations were detected (bootstrap detection at female n: 38–52% male-equal, 13–18% 50%-of-male); the prescribed-item count showed BMI-related covariance with step-test $\dot{V}O_2$max and blood pressure. Only shuttle-run $\dot{V}O_2$max showed a FDR-significant sex × Time × Group interaction (p = .049), attenuated after baseline adjustment (p = .061); this sex difference is not robustly established. Relative grip strength was more responsive to BMI change than absolute.

Introduction

Overweight and obesity are among the most pressing public-health concerns of the twenty-first century, and their global prevalence continues to rise. According to the World Health Organization (WHO), approximately 43% of adults aged 18 years or older were overweight in 2022, of whom roughly 16% were classified as obese[1]. Korea is among the countries with rapidly rising obesity prevalence: in 2021, adult obesity prevalence was 38.4% (49.2% in men and 27.8% in women), a 27% relative rise from 30.2% in 2012[2]. The 2024 Obesity Fact Sheet from the Korean Society for the Study of Obesity reported the highest prevalence in adults in their 30s (42.0%) and 40s (42.2%), confirming weight management in economically active adults as an urgent policy priority[3].

Exercise is widely recognised as one of the most effective non-pharmacological interventions for managing overweight and obesity. The American College of Sports Medicine and the WHO recommend that adults perform at least 150 minutes of moderate-intensity or 75 minutes of vigorous-intensity physical activity per week[4,5]. Multiple studies have reported that such exercise is associated with reductions in body weight and body fat and with higher levels of overall fitness, including cardiorespiratory endurance, muscular strength, and flexibility[6]. A recent randomised controlled trial reported that 8 weeks of aerobic, resistance, or combined training was associated with reductions in body-fat percentage in inactive overweight or obese adults, with resistance and combined training showing larger associations with glucose metabolism[7]. A meta-analysis of randomised controlled trials reported that high-intensity interval training was associated with reductions in BMI in obese populations and with larger associations with cardiorespiratory fitness than moderate-intensity continuous training[8].

Sex differences in exercise responses remain incompletely characterised. A systematic review of 88 cardiac-rehabilitation studies reported similar responses between sexes for most physiological variables, although men showed larger absolute changes in V_{O2}max, functional capacity, and grip strength[9]. Whether these patterns are preserved when BMI change is treated as a within-interval intermediate variable, and whether specific dimensions of exercise prescription (type vs. count of prescribed items) align differently with outcomes by sex, has not been examined.

Path-analytic and mediation-style models have been used to describe statistical links between exercise interventions and outcomes[10–12]. When intermediate and outcome variables are measured over the same interval, however, the resulting a·b product cannot be identified as a temporally ordered mediation effect; it is more accurately read as a symmetric covariance partition of the prescription–outcome association into a BMI-related component (a·b) and a residual direct component (c'). Bootstrap-based confidence intervals do not require normality and permit accurate interval estimation of these decomposition components even in small samples, and are now standard[10,13]. Recent studies have used mediation-style models to characterise patterns of exercise effects[14,15]. To our knowledge, however, no large-scale study to date has examined — with sex stratification — how exercise prescription characteristics co-vary with BMI change and with body composition and fitness outcomes in overweight adults.

The Korean National Fitness Award(NFA) is a nationwide fitness-management programme operated by the Korea Institute of Sports Science. Across 75 certification centres in 17 metropolitan cities and provinces, the programme uses standardised protocols to measure body composition, blood pressure, grip strength, flexibility, muscular endurance, power, cardiorespiratory endurance, and agility, and provides individually tailored exercise prescriptions[16,17]. Over the past decade, the NFA has accumulated data from more than 500,000 Korean adults, used for studies of fitness distributions, fitness-age prediction, and fitness–hypertension associations[16–18]. To our knowledge, no previous study has used NFA data to examine sex-stratified within-interval covariance patterns linking exercise prescription characteristics with BMI change in overweight adults.

Three design features preclude a strict causal interpretation. First, exercise prescription type was not randomly assigned; it reflected participant preference and practitioner judgment, raising the possibility of confounding by indication. Second, BMI change (the proposed intermediate variable) and outcome change were observed over the same measurement interval, so temporal precedence between the intermediate and the outcome is not established; this violates a core assumption of causal mediation[10,12,19]. Third, BMI is mathematically a function of body weight and height, so several candidate outcomes (body weight, waist circumference, body-fat percentage) are not dimensionally independent of the intermediate variable. We therefore frame the analyses below as within-interval covariance partitions of statistical associations — not as causal mediation — and draw our principal scientific claims from the BMI-independent fitness outcomes. Body-composition outcomes are retained as supplementary, corroborative descriptions of BMI change itself.

The objectives of this study were threefold. First, using paired pre- and post-measurements from overweight Korean adults in the NFA, we described differences in body composition, blood pressure, grip strength, and physical fitness across BMI change profiles (decreased, maintained, increased). Second, we estimated direct associations and BMI-related covariance components of qualitative exercise prescription (Resistance, Aerobic, or Mixed type) and the quantitative count of prescribed main exercise items with 14 outcomes, using BMI change as a within-interval covariance partition rather than as a temporally identified mediator. Third, all analyses were stratified by sex, and sex × Time × Group interactions were formally tested, to quantitatively compare patterns between men and women.

Results

[Figure 1 near here]

[Table 1 near here]

Sample, baseline characteristics, and supplementary body-composition outcomes

The CONSORT-style participant selection flow is summarised in Figure 1. After applying the seven-step selection criteria (Figure 1), the analytic sample comprised 12,980 overweight adults (10,173 men, 2,807 women) categorised by BMI-change quartile into Decreased (DG; ΔBMI < −2.77%), Maintained (MG; −2.77% ≤ ΔBMI ≤ +1.19%), and Increased (IG; ΔBMI > +1.19%) groups. Baseline characteristics for the overall sample are presented in Table 1, and sex-stratum baseline tables are reported in Supplementary Tables S1–S2. Two sex-stratum baseline patterns warrant note: the female MG mean age (45.5 yr) was approximately 7 years higher than the male MG mean age (37.8 yr; Supplementary Tables S1–S2), and women in all three groups had higher body-fat percentage (37.0–37.7%) than men (24.3–25.8%) at comparable BMI, reflecting well-known sex differences in body composition at equivalent BMI. The within-stratum SD for body fat was also markedly narrower in women (~3.7) than in men (~4.8–5.1; Supplementary Tables S1–S2), so absolute body-fat changes that are similar in the two sexes correspond to larger standardised effects in women. All 12 baseline variables differed across groups (all $p < .001$ in the overall sample) with the same regression-to-the-mean pattern for body-fat percentage, BMI, and waist circumference (IG < MG < DG), addressed in path models by adjustment for pre-BMI and pre-outcome values. Body-composition Time × Group interactions were detectable for all five indices in the overall sample and within each sex (largest for body weight, $\eta_p^2 = .052$ overall, .054 men, .049 women); the full ANCOVA results for the supplementary body-composition outcomes are in Supplementary Table S3. Because these outcomes are mathematically coupled to ΔBMI, we report the pattern for completeness without independent interpretive weight.

Secondary outcomes: blood pressure

In the overall sample, Time × Group interactions for systolic and diastolic blood pressure were small but detectable (η_p^2 = .005 and .003 respectively); DG and MG were lower and IG higher at post-test. Effect sizes were similar in men (SBP η_p^2 = .006; DBP .004) and approximately one-third as large in women (SBP .002, p = .013; DBP .001, p = .045), indicating a weaker BMI–blood-pressure alignment in women. Stratum-specific tables are in Supplementary Table S4.

Principal outcomes: physical fitness

[Table 2 near here]

Across the seven principal BMI-independent fitness outcomes, all Time × Group interactions were small by Cohen's conventions[20] ($\eta_p^2 \leq .03$). In the overall sample the largest interactions were for step-test V_{O2}max (η_p^2 = .021, FDR p < .001), shuttle-run V_{O2}max (.017, FDR p < .001), and relative grip strength (.011, FDR p < .001). Effect sizes for flexibility, muscular endurance, power, and reaction time were small to negligible. The absence of detectable Time × Group interactions for left and right absolute grip strength (in either sex) contrasts with the detectable interaction for relative grip strength (η_p^2 = .012 in men, .008 in women): all three groups achieved comparable post-test absolute strength, but body-weight normalisation revealed group divergence. Sex-stratum cell means and confidence intervals are in Supplementary Table S5. To contextualise observed change against measurement noise, mean changes for all principal outcomes were compared against the published minimal detectable change (MDC) values listed in Supplementary Table S6. All principal-outcome mean changes exceeded their corresponding MDC except for reaction time in MG (observed |Δ| = 0.012 s vs MDC ≈ 0.04 s; Supplementary Table S6), which is therefore interpreted with caution.

Within-sex effect-size comparison — the central novel observation — differed in two informative ways. First, the cardiorespiratory effect sizes were larger in women than men: shuttle-run V_{O2}max η_p^2 = .028 (women) vs .016 (men), and step-test V_{O2}max .027 vs .022. Second, three principal outcomes (muscular endurance, power, reaction time) showed detectable interactions in men but not in women.

Path-analytic associations

[Table 3 near here]

The b-path (ΔBMI → ΔY) was robustly non-null for all 14 outcomes in the overall sample and for 13 of 14 in men (the exception was reaction time); in women the b-path failed to reach detection for absolute grip strength only. Among principal fitness outcomes the b-paths were uniformly negative, indicating that BMI increases were aligned with overall lower post-test fitness; the b-path was positive for absolute grip strength and negative for relative grip strength. Complete path-analytic coefficients for the overall, male, and female samples are reported in Supplementary Tables S7, S8, and S9, respectively.

Direct associations of prescription type with principal outcomes were detectable in men only. Aerobic prescriptions were associated with higher post-test flexibility (c_2' = +0.361, 95% CI [+0.117, +0.612]) and power (c_2' = +1.253 [+0.318, +2.204]), and with step-test V_{O2}max in a BMI-related covariance component ($a_2 \cdot b$ = −0.097 [−0.189, −0.015]). The count of prescribed items (n_{total}) showed positive direct associations in men with flexibility, step-test V_{O2}max, both absolute grip strengths, and relative grip strength (all FDR-significant within the principal family). In women, no direct prescription-type associations reached the FDR threshold. To assess whether this non-detection was attributable to reduced statistical power at the smaller female sample size, we re-fitted the male path-analytic model after randomly down-sampling the male sample to the female n (2,272) over 1,000 bootstrap replicates under two alternatives: a "male-equal" alternative (preserving the male point estimates) and a conservative "50%-of-male" alternative (halving male point estimates). For the aerobic c_2' on flexibility — the strongest male direct association — detection probability at the female n was 52% (95% simulation CI 49–55%) under the male-equal alternative and 18% (16–21%) under the 50%-of-male alternative; the corresponding ranges across all five male-detectable direct associations were 38–52% and 13–18%, respectively (the five associations are enumerated in Supplementary Table S10). In contrast, BMI-related covariance components of n_{total} were detectable in women for step-test V_{O2}max ($a_3 \cdot b$ = +0.021, 95% CI [+0.007, +0.036]), waist circumference, systolic blood pressure, and reaction time. Because the directionality sensitivity (Supplementary Table S11) showed that these a·b products were of comparable magnitude when ΔBMI and ΔY were swapped, the products are interpreted as descriptive partitions of within-interval covariance, not as temporally identified mediation. A notable sex-specific finding was that the female direct n_{total} associations on flexibility and muscular endurance were negative, primarily carried by the stretching-item count (modality decomposition in Supplementary Table S12); given the absence of training-volume information, this pattern may reflect noise or unmeasured prescriber-patient dynamics and is interpreted with caution.

Sex comparison and sensitivity analyses

[Table 4 near here]

In the pooled ANCOVA model, after FDR correction across the seven principal outcomes, the three-way interaction was detectable for shuttle-run V_{O2}max (raw p = .014, FDR p = .049) and relative grip strength (raw p = .006, FDR p = .028); for step-test V_{O2}max it was non-significant with a borderline raw p (.062, FDR p = .145). Among the ten pre-specified sensitivity analyses, the ±3% clinically anchored BMI-change cut-off — which we treat as a co-primary analysis given that the quartile-based cut-off is sample-dependent and was not pre-registered on a clinical basis — yielded principal η_p^2 estimates within 10–15% of the quartile-based estimates and preserved the same qualitative pattern of FDR-significant Time × Group interactions (Supplementary Table S13). The other nine sensitivity analyses (±5% BMI-change cut-off; IPTW; bootstrap down-sampling under two alternatives; cluster-robust standard errors with random-intercept cross-check; modality decomposition of n_{total}; n_{total} construct validity; directionality sensitivity

with ΔBMI ↔ ΔY; age-stratified re-analysis; MNAR-by-design check; selection-bias check for unclassifiable prescriptions) likewise preserved the qualitative principal conclusions. IPTW attenuated male direct aerobic associations by ~11–18% but they remained detectable; cluster-robust SEs widened intervals by 8–12% without changing FDR-significance; the directionality sensitivity confirmed that a·b magnitudes are comparable when ΔBMI and ΔY are swapped — a finding that is incompatible with a true mediation chain and instead supports the within-interval covariance-partition interpretation explicitly stated in the Methods. Full sensitivity tables are in Supplementary Tables S10–S18.

Discussion

Using large-scale data from 12,980 overweight adults enrolled in the Korean NFA, this study characterised how BMI change profiles are statistically associated with changes in body composition and physical fitness, and how exercise prescription characteristics align with these outcomes through within-interval BMI covariance partitions, in sex-stratified models. Four principal findings emerge. BMI change was statistically aligned with all principal fitness outcomes (b-path detectable for all at FDR $p < .001$), with all Time × Group interaction effect sizes in the small range ($\eta_p^2 \leq .03$). In men, direct associations of exercise prescription type with specific fitness domains were detectable. In women, such direct associations were not detected (Supplementary Table S9), although the bootstrap down-sampling analyses quantified detection probability at the female sample size as 38–52% under a male-equal alternative and 13–18% under a conservative 50%-of-male alternative; BMI-related covariance components of the count of prescribed items reached statistical detection for step-test V_{O2}max, waist circumference, systolic blood pressure, and reaction time (Supplementary Table S9). Finally, of the two cardiorespiratory measures, only shuttle-run V_{O2}max showed a sex × Time × Group interaction surviving FDR correction (raw $p = .014$, FDR $p = .049$, larger response in women), and this attenuated towards non-significance after baseline V_{O2}max adjustment (adjusted FDR $p = .061$); the step-test V_{O2}max interaction did not survive multiplicity correction (FDR $p = .145$). The cardiorespiratory sex difference is therefore best characterised as not robustly established by this dataset, with the apparent female advantage largely consistent with a baseline-asymmetry (law-of-initial-values) explanation.

Three design features bound the interpretation. First, because prescription type was not randomly assigned, all "prescription association" statements are conditional associations, robust to the measured covariates and IPTW but vulnerable to residual confounding by dietary intake, adherence, and practitioner characteristics. Second, and most importantly, because BMI change and outcome change were observed over the same interval, the path-analytic a·b products do not constitute a temporally identified mediation chain. The directionality sensitivity (Supplementary Table S11) confirmed that a·b magnitudes were comparable when ΔBMI and ΔY were swapped — a finding that is incompatible with a true mediation structure (in which reversing the variable order should attenuate the indirect effect) and instead indicates that the model recovers a symmetric within-interval covariance partition. We therefore explicitly read every a·b product in this manuscript as a descriptive BMI-related covariance component, not as evidence that BMI change is a causal mechanism through which prescription characteristics produce fitness

change. The use of the term "mediator" in the lavaan output and in the literature is retained for software reproducibility but should not be interpreted causally. Third, because we restricted principal claims to BMI-independent fitness outcomes, the tautological inflation of b-paths for body weight, waist circumference, and body-fat percentage does not threaten the principal conclusions. We read the combined evidence as hypothesis-generating motivation for randomised evaluation of sex-specific prescription strategies, not as grounds for guideline change.

An informative discrepancy was observed for grip strength. The b-path for absolute grip strength was positive, whereas the b-path for relative grip strength was negative, indicating that BMI increases are partly accompanied by greater muscle mass while the strength-to-body-weight ratio declines. The absence of a detectable Time × Group interaction for absolute grip strength in either sex contrasts with the detectable interaction for relative grip strength, and the FDR-corrected three-way sex × Time × Group interaction for relative grip strength (FDR p = .028, favouring men) indicates that this absolute–relative divergence is more pronounced in men, while still present in women. This pattern supports relative grip strength as a more appropriate functional-fitness indicator in overweight populations, consistent with the EWGSOP2 sarcopenia framework[21], which prioritises muscle strength as the primary marker of muscle health[22].

The sex-stratified pattern in the principal outcomes — detectable direct associations of prescription type in men, and detectable BMI-related covariance components for the count of prescribed items in women — is the central novel observation. Two competing readings should be held in balance. First, the female non-detections are quantitatively compatible with reduced statistical power: at the path-analytic female n = 2,272, bootstrap down-sampling showed that, even if the true female direct effect were equal in magnitude to the male effect, detection probability would be only 38–52% across the five male-detectable direct associations; if the true female effect were merely half the male magnitude, detection probability would fall to 13–18%. Under either alternative, the female sample is appropriately under-powered, and "non-detection" cannot be read as "absence of effect." Second, however, we cannot exclude that the true female effect is genuinely smaller than half the male magnitude. The available data do not discriminate between these readings. The modality decomposition further qualifies the female pattern: the BMI-related covariance component of n_{total} with step-test VO_2max in women was carried primarily by aerobic-item count, while the negative direct associations on flexibility and muscular endurance were carried by stretching-item count. The direct type associations in men (aerobic with flexibility and power; aerobic with step-test VO_2max via a BMI-related covariance component) are consistent with the Specific Adaptations to Imposed Demands principle[23] and with prior randomised trials reporting that resistance training is associated with waist-circumference and visceral-fat reductions[24,25]. Importantly, n_{total} is not a measure of training volume, energy expenditure, or physiological dose, and the construct-validity diagnostic indicates that 18.4% of n_{total} variance is at the centre level (Methods); n_{total} operates as a composite index influenced by engagement, prescriber attention, and centre culture. We therefore interpret the negative female associations on flexibility and muscular endurance with caution; given modest effect sizes, the absence of training-volume information, and multiplicity, these patterns may

reflect noise or unmeasured prescriber-patient dynamics rather than a true performance-suppressing effect. We do not recommend "more items" as a prescription strategy.

Candidate biological mechanisms for sex-differential exercise responses — including type-I muscle-fibre proportion, oestrogen-mediated lipid-oxidation shifts, and sex-differential adipose-tissue β-adrenergic regulation[26–28] — cannot be tested with NFA data, which lack muscle biopsy, hormonal profiling, menstrual-cycle phase, and adherence records. We mention them only as motivation for future mechanistic studies with hormonal profiling and direct adherence measurement, not as interpretation of the present results.

For shuttle-run V_{O2}max, the Time × Group interaction was larger in women (η_p^2 = .028) than in men (η_p^2 = .016), with a FDR-corrected three-way interaction reaching the .05 threshold (raw p = .014, FDR p = .049). For step-test V_{O2}max the raw effect sizes also favoured women (.027 vs .022), but the three-way interaction did not survive multiplicity adjustment (FDR p = .145). Two cautions therefore qualify the "larger response in women" reading. First, only one of the two cardiorespiratory measures crossed the FDR threshold, leaving the pattern statistically fragile. Second, the most parsimonious explanation is the "law of initial values": women have lower baseline V_{O2}max than men, predicting larger relative changes from a lower baseline. In a post-hoc check, adjusting for baseline V_{O2}max in the three-way interaction model attenuated the sex difference in shuttle-run V_{O2}max by ~ 35% and brought the adjusted FDR p to .061 — i.e., baseline asymmetry accounts for the bulk of the apparent sex difference. The combined evidence — single-measure FDR survival, attenuation under baseline adjustment, and the absence of testable mechanistic data on peripheral arteriovenous-difference or oestrogen-mediated mitochondrial pathways[29,30] — leads us to characterise the cardiorespiratory sex difference as not robustly established by this dataset. The present data should not be cited as evidence for a sex-specific cardiorespiratory adaptation; the conservative reading is that the apparent female advantage is largely consistent with baseline-asymmetry artefact and requires confirmation in randomised, baseline-balanced designs.

Two clinical implications, framed conservatively given the hypothesis-generating nature of observational data, follow from the principal findings. First, the absolute–relative grip strength divergence is clinically actionable: in overweight adults, absolute grip strength can rise in parallel with body weight without corresponding improvement in the strength-to-body-weight ratio, and the FDR-corrected detectable three-way interaction for relative grip strength (favouring men) and the absence of such an interaction for absolute grip strength in either sex identify relative grip strength as the more sensitive functional indicator. We recommend its use as a primary grip-strength outcome in overweight populations. Second, the sex-differentiated pattern of detectable associations motivates future randomised evaluation of sex-specific prescription strategies. For men, type-specific prescriptions tested against outcome-specific targets (e.g., aerobic prescriptions tested against flexibility and power outcomes) are a direct extension of the type-by-outcome patterns observed. For women, the hypothesis to test is whether prescriptions with greater aerobic-modality breadth produce larger cardiorespiratory change through greater BMI

change, independent of total training minutes. We deliberately do not translate the observational n_{total} association into a "more items is better" recommendation.

Several strengths support the present analysis. Statistical power and external validity were secured by analysing 12,980 paired pre/post records from a population-representative national programme. The unstructured-text exercise prescription was systematically parsed and the dictionary, classification rules, and R scripts have been deposited at the Open Science Framework (project ID hmkjg; anonymous peer-review link: https://osf.io/hmkjg/overview?view_only=1386312aa40e4918a503de9fc5f8c112), supporting reproducibility. 5,000-iteration BCa bootstrap confidence intervals (with 20,000 iterations for three-decimal-precision outcomes) were computed for all path-analytic parameters, and Benjamini–Hochberg FDR correction was applied within the principal-outcome family and within the sex × Time × Group three-way interaction family. Pre-designation of BMI-independent fitness outcomes as principal eliminated tautological inflation. Ten pre-specified sensitivity analyses systematically addressed the main threats to validity, including a co-primary clinically anchored ± 3% BMI-change cut-off, a directionality sensitivity that explicitly tests the covariance-partition framing, and a baseline-comparison check for selection bias from the prescription-classification step. FDR-corrected formal three-way interaction tests provided inferential rather than merely descriptive support for sex-comparison claims.

Several limitations bound the conclusions. First and most fundamentally, because BMI change and the outcome were measured over the same interval, the path-analytic decomposition is not identifiable as causal mediation; the directionality sensitivity (ΔBMI ↔ ΔY swap with comparable a·b magnitudes) confirms this directly and indicates that the model recovers a symmetric within-interval covariance partition. We therefore use "BMI-related covariance component" rather than "indirect effect" throughout. Second, the propensity-score model was built from variables available in NFA

records; unmeasured determinants of prescription type — participant motivation, practitioner discretion, centre culture — are not fully addressed, and IPTW-adjusted estimates reduce but do not eliminate bias. Third, the count of prescribed items is not a measure of training dose, energy expenditure, or adherence, and n_{total}-based findings should be read as composite-index associations whose mechanism (engagement, prescriber attention, or domain-specific exposure) is not separately identified. Fourth, the quartile-based BMI-change classification is sample-dependent (Cohen's κ = 0.71 vs ± 3%, 0.46 vs ± 5%) and was not pre-registered on a clinical basis; the DG cut-off (− 2.77%) is at the margin of scale measurement precision (~ 2.2 kg for an 80-kg adult). We therefore present the ± 3% clinically anchored cut-off analysis (Supplementary Table S13) as co-primary: principal η_p^2 estimates were within 10–15% of the quartile-based estimates and the qualitative pattern of FDR-significant Time × Group interactions was preserved. Fifth, the female sample (n = 2,807) was 28% of the male sample (10,173), and the female MG mean age (45.5 year) was ~ 7 years higher than the male (37.8 year); residual age-by-sex confounding cannot be fully excluded, although the age-stratified sensitivity analysis (Supplementary Table S16) preserved the female-favouring cardiorespiratory η_p^2 pattern in 3 of 4 age × outcome cells. NFA data do not record menstrual-cycle phase or menopausal status. Sixth, the path-analytic sample (n = 10,033) was reduced by ~ 23% from the final analytic sample (n = 12,980) because of unclassifiable prescriptions;

Supplementary Table S18 shows broadly comparable baseline characteristics between retained and excluded participants, but unmeasured centre-level confounding cannot be excluded. Seventh, NFA participants are voluntary attendees of a government-administered fitness programme and are likely more health-motivated than the general overweight Korean population. The 21.6% female share of the analytic sample is below the female share of the overweight Korean adult population, partly reflecting the male-skewed BMI ≥ 25 kg/m² prevalence (49.2% men vs 27.8% women) but also a sex-differential propensity to enrol in voluntary fitness programmes; generalisation to non-NFA overweight populations — particularly to less health-motivated women — should be made with caution, and female-specific effect estimates may overstate associations expected in less engaged subpopulations. Despite FDR correction in two families, the analysis remains exploratory; paths significant only at raw $p < .05$ should be regarded as suggestive. We did not have individual adherence records, dietary intake, sleep, stress, or practitioner identifiers. Future work should prioritise randomised sex-specific prescription trials, adherence-instrumented designs, mechanistic studies of hormonal-metabolic pathways with direct measurement of menstrual/menopausal status, and validation of relative grip strength as a standardised functional indicator.

In conclusion, in 12,980 overweight Korean adults, BMI change was statistically aligned with all principal fitness outcomes; in men, direct associations of exercise prescription type with specific fitness domains were detectable, while in women such direct associations were not detected — a pattern compatible with reduced statistical power (38–52% detection probability under a male-equal alternative; 13–18% under a 50%-of-male alternative) and not interpretable as biological absence — and BMI-related covariance components of the count of prescribed items reached detection. Of the two cardiorespiratory measures, only one (shuttle-run V_{O2}max) showed a sex × Time × Group interaction that survived FDR correction, and this attenuated towards non-significance after baseline V_{O2}max adjustment; the step-test V_{O2}max interaction did not survive multiplicity correction. The sex difference in cardiorespiratory response is therefore best characterised as not robustly established by this dataset, with the apparent female advantage largely consistent with a baseline-asymmetry (law-of-initial-values) explanation rather than a sex-specific physiological adaptation. Relative grip strength was substantially more responsive to BMI change than absolute grip strength and is recommended as the preferred functional-fitness indicator in overweight populations. Because BMI change and outcome change were observed over the same interval, all path-analytic statements should be read as descriptive within-interval covariance partitioning, not causal mediation. These observational findings are hypothesis-generating and motivate randomised trials of sex-specific prescription strategies, rather than directly supporting guideline change.

Methods

Study design and data source

This was a retrospective observational study based on publicly available NFA data. The NFA is a nationwide fitness-management programme operated by the Korea Institute of Sports Science under the Korea Sports Promotion Foundation. Across 75 certification centres in 17 metropolitan cities and

provinces, the programme provides standardised fitness assessments and tailored exercise prescription services, with measurements and prescription content stored in an integrated information system[16,17]. The raw data consisted of the "Comprehensive Data on Fitness Measurement and Exercise Prescription" files available through the Korean government Open Data Portal, comprising 1,199,471 records from January 2022 to December 2025.

The study used only publicly available de-identified data, and the protocol was reviewed and approved by the Public Institutional Review Board under the Ministry of Health and Welfare (approval number P01-202602-01-029). Each record included an anonymized participant identifier, measurement date, age, sex, height, body weight, body-fat percentage, waist circumference, blood pressure, grip strength, physical-fitness test results, and the unstructured text of the exercise prescription. Certification-centre identifiers were retained to permit centre-level sensitivity analyses.

Measurement standardization and reliability

All 75 NFA certification centres follow a standardised measurement manual issued by the Korea Institute of Sports Science; staff at each centre undergo mandatory training, and equipment is calibrated annually under a unified protocol[16,17]. Cardiorespiratory fitness was assessed using either the 20-m multi-stage shuttle-run test or the 3-minute step test depending on participant fitness and centre capacity. Because no independent between-centre reliability study of the specific NFA fitness tests has been published, we do not rely on a prior reliability estimate. We supplement the analysis with a minimal detectable change (MDC) calibration: for each continuous outcome, observed mean change is contextualised against published MDC or smallest worthwhile change values where available; the specific MDC reference values used for each of the 14 outcomes are listed in Supplementary Table S6.

Participants and selection criteria

[Figure 1 near here]

Participant selection followed the seven-step CONSORT-style flow diagram shown in Figure 1. From the 1,199,471 raw NFA records, we sequentially extracted participants with two or more visits, restricted to no more than five repeat visits (≤ 5; participants with > 5 repeat visits excluded, consistent with Figure 1), and aged 19–64 years. To compare paired pre- and post-measurements in the same participant, we included only individuals whose revisit interval was 30–365 days. The 30-day lower bound reflects the empirical observation that, in overweight or obese adults, measurable training-induced adaptations in body composition and cardiorespiratory fitness emerge over training programmes lasting on the order of 4–12 weeks rather than over single sessions[31,32]. The 365-day upper bound was set to limit cumulative drift from unmeasured time-dependent confounders, including seasonal weight fluctuation across an annual cycle, which has been documented as a non-trivial source of variability in 12-month cohort studies of body weight[33]. Applying these criteria yielded 45,101 participants with complete pre- and post-BMI

data. Next, we extracted 13,095 overweight adults with pre-measurement BMI 25–29.9 kg/m² (WHO criteria[1]), and applied Tukey's 1.5× IQR rule to exclude 115 (0.88%) BMI outliers. The final analytic sample was 12,980 (10,173 men, 2,807 women); for path-analytic models, participants whose main exercise prescription could not be classified as Aerobic, Resistance, or Mixed were additionally excluded, reducing the path-analytic samples to 10,033 overall, 7,761 men, and 2,272 women. To assess whether this exclusion introduced selection bias, we compared the baseline characteristics of the 2,947 participants with unclassifiable prescriptions against those of the 10,033 retained for path analysis (Supplementary Table S18); the two groups were broadly comparable, with the largest standardised mean difference observed for revisit interval ($|d| = 0.13$).

Exercise prescription parsing, clustering, and visualisation

All analyses of the unstructured-text exercise prescriptions were performed in R. The NFA prescription content is stored as semi-structured free text in the format "Warm-up: … / Main exercise: … / Cool-down: …" and was parsed using regular expressions. We focused the present analysis on the main exercise component because it is defined in the NFA manual as the principal conditioning stimulus and is the only component consistently populated across centres. More than 500 fragmented Korean exercise terms were consolidated into 18 standardised English categories; each prescription was classified into one of three types — Aerobic (aerobic only), Resistance (resistance only), or Mixed (both aerobic and resistance) — using predefined rules. The count of prescribed main exercise items (n_{total}) was also calculated. This count is not a measure of training volume, energy expenditure, or physiological dose; the construct validity of n_{total} is examined in the sensitivity analyses, including a centre-level intraclass correlation coefficient (ICC) computed from a random-intercept null model ($n_{total} \sim 1 + (1 \mid centre)$), which yielded ICC = 0.184, indicating that 18.4% of n_{total} variance is at the centre level. The full Korean–English keyword dictionary and the category→type classification rules are summarised in Supplementary Tables S23–S24.

BMI change group classification

The percentage change in BMI between pre- and post-measurements was computed as ((post-BMI − pre-BMI) / pre-BMI) × 100. We used two complementary classifications, both reported as co-primary. The principal classification used the first quartile (Q1) and third quartile (Q3) of the BMI change distribution in the overweight sample as cut-off points: Decreased (DG; change < −2.77%), Maintained (MG; −2.77% ≤ change ≤ +1.19%), and Increased (IG; change > +1.19%). Quartile cut-offs balance the three groups and maximise statistical power but are sample-dependent and do not carry intrinsic clinical meaning[34]. As a co-primary clinically anchored classification, we additionally fitted all principal models with ±3% BMI-change cut-offs (full results in Supplementary Table S13) and report the ±5% cut-off as a secondary sensitivity analysis. Cohen's κ agreement between the quartile-based and ±3% classifications was 0.71 (substantial; 18.3% reclassified); for ±5%, κ = 0.46 (moderate; 37.8% reclassified). Values are reported in

Supplementary Table S19. The principal results in the main text are reported under the quartile-based classification; the same qualitative pattern of FDR-significant Time × Group interactions held under the ±3% co-primary cut-off (Supplementary Table S13), and any reader who prefers a clinically anchored reading should consult that table directly.

Statistical analysis

All analyses were conducted in R 4.5.3. Baseline characteristics were compared using one-way ANOVA with Bonferroni-corrected post hoc tests. Main effects of time (pre vs. post) and group (DG vs. MG vs. IG), and their interaction, were examined using repeated-measures ANCOVA with the model value ~ Time × Group + covariates, with pre-BMI, age, revisit interval, and sex (in the overall analysis only) as covariates. Height was not entered as a separate covariate because it is algebraically absorbed into pre-BMI. F statistics and p-values were calculated using Type III sums of squares with treatment (dummy) coding for Group (DG, MG, IG) using MG as the reference category; effect sizes were partial η^2 interpreted following Cohen's conventions. Centre-level clustering was addressed through cluster-robust standard errors (sandwich estimator); a random-intercept mixed-model sensitivity is reported in Supplementary Table S15. Variable-wise missingness after outlier exclusion is reported in Supplementary Table S22.

Path-analytic association model

To describe within-interval covariance among exercise prescription characteristics, BMI change, and changes in body composition and physical fitness, we fitted a path-analytic regression model with the structural form of a single-mediator model using lavaan (version 0.6-19). Because BMI change and outcome change were observed over the same measurement interval, the model is specified for descriptive covariance partitioning rather than for causal mediation: the a·b product is interpreted as the BMI-related covariance component of the prescription–outcome association, and the directionality sensitivity (Supplementary Table S11) was used to verify that swapping ΔBMI and ΔY produced comparable a·b magnitudes (a property incompatible with a true mediation chain but expected for symmetric covariance partitioning). The intermediate-variable equation (path a) was $\Delta BMI = a_1 \cdot Resistance + a_2 \cdot Aerobic + a_3 \cdot n_{total} + covariates + e_1$, and the outcome equation (paths b and c') was $\Delta Y = b \cdot \Delta BMI + c_1' \cdot Resistance + c_2' \cdot Aerobic + c_3' \cdot n_{total} + covariates + Y_{pre} + e_2$. Resistance and Aerobic were dummy-coded with Mixed as reference (a sensitivity analysis using Aerobic as reference is in Supplementary Table S20). Parameters were estimated by full-information maximum likelihood (FIML); 95% confidence intervals were computed using 5,000 BCa bootstrap resamples (20,000 for three-decimal precision outcomes; Supplementary Table S21). Throughout the manuscript, the term "mediator/mediation" appears only in software-output and reproducibility contexts; its causal interpretation is explicitly disclaimed.

Outcome prioritization

Because BMI is a deterministic function of body weight and height, and is algebraically related to waist circumference and body-fat percentage at the individual level through the shared dependence on body weight, the path b for body weight, waist circumference, and body-fat percentage is partly tautological. We therefore pre-designated the principal scientific claims to seven BMI-independent fitness outcomes: shuttle-run V_{O2}max, step-test V_{O2}max, sit-and-reach (flexibility), cross sit-ups (muscular endurance), standing long jump (power), reaction time, and relative grip strength. Blood pressure (systolic and diastolic) and absolute grip strength (left and right) were treated as four secondary outcomes. Body weight, waist circumference, and body-fat percentage were analysed and reported as three supplementary outcomes.

Sex-stratified and pooled analyses

All primary analyses were stratified by sex. To formally test whether effect sizes differed between men and women, we additionally fitted a pooled model including sex as a three-way interaction factor (sex × Time × Group) for each outcome, and a pooled path-analytic model with three-way product terms (sex × Resistance, sex × Aerobic, sex × n_{total}) on both the a and the c' paths. The seven three-way interaction p-values obtained for the principal outcomes were Benjamini–Hochberg FDR-corrected; both raw and FDR-adjusted p-values are reported.

Multiple testing

The analyses are explicitly exploratory and hypothesis-generating. To guard against over-interpretation we applied Benjamini–Hochberg FDR correction within each of the following families: (i) the seven principal BMI-independent fitness outcomes; (ii) each family of paths (a, b, c', a·b) within a given sex stratum; and (iii) the seven sex × Time × Group three-way interaction tests in the pooled principal-outcome family. Results reported as statistically detectable in the Abstract and Conclusions retain significance at FDR-adjusted $p < .05$ within the principal-outcome family.

Sensitivity analyses

Ten pre-specified sensitivity analyses addressed the main design and modelling concerns: (i) co-primary clinically anchored BMI-change cut-offs at ±3% (Supplementary Table S13) and a secondary ±5% cut-off; (ii) propensity-weighted prescription type via inverse-probability-of-treatment weighting (balance threshold <0.1; Supplementary Table S14); the IPTW analytic sample (N = 9,576) was reduced from the path-analytic sample (n = 10,033) by complete-case restriction on the propensity-score covariates; (iii) bootstrap down-sampling of men under two alternatives (male-equal and 50%-of-male) with 1,000 replicates per outcome (i.e., the 1,000 replicates were generated independently for each outcome, not pooled across outcomes), where each replicate randomly subsampled the male path-analytic sample to n = 2,272 (the female sample size) and re-fitted the path model; detection probability was defined as the

proportion of replicates in which the BCa 95% confidence interval excluded zero; (iv) centre-level clustering through cluster-robust SEs and a random-intercept mixed-model cross-check (Supplementary Table S15); (v) modality-specific decomposition of n_{total} into six counts ($n_{aerobic}$, $n_{resistance}$, $n_{stretching}$, $n_{yoga_pilates}$, $n_{balance}$, n_{other}); (vi) MDC calibration; (vii) directionality sensitivity in which ΔBMI and ΔY were swapped, to verify that the a·b product behaves as a symmetric covariance partition rather than as a temporally identified mediation chain (Supplementary Table S11); (viii) age-stratified re-analysis of cardiorespiratory outcomes within 19–39 y and 40–64 y strata to address residual age-by-sex confounding (Supplementary Table S16); (ix) baseline-characteristic comparison across cardiorespiratory test assignment groups as a missing-not-at-random-by-design (MNAR) check of the within-outcome complete-case analysis (Supplementary Table S17); and (x) baseline comparison between the path-analytic sample (n = 10,033) and participants excluded because of unclassifiable prescriptions (n = 2,947), to evaluate selection bias from the prescription-type classification step (Supplementary Table S18).

Use of large language models

The authors used a large language model only for English-language polishing of the manuscript and for code-style suggestions during R script preparation. All scientific content, statistical analyses, and final wording were reviewed and verified by the human authors, who take full responsibility for the manuscript as submitted.

Declarations

Data availability

The datasets supporting the conclusions of this article are publicly available from the Korean Government Open Data Portal. The specific dataset is titled "Comprehensive Data on Fitness Measurement and Exercise Prescription" (□□□□ □ □□□□ □□ □□□) provided by the Korea Sports Promotion Foundation. Direct access: https://www.data.go.kr/data/7844714/linkedData.do

Code availability

The R analysis scripts (path-analytic models, sensitivity analyses, semantic-network construction), the Korean–English keyword dictionary, the category→type classification rules, and the complete 13 paths × 14 outcomes coefficient matrix (point estimates, BCa 95% CIs, standard errors, and bootstrap p-values) have been deposited at the Open Science Framework (project ID: hmkjg). For peer review, an anonymous view-only link is provided: https://osf.io/hmkjg/overview?view_only=1386312aa40e4918a503de9fc5f8c112. Upon acceptance, the project will be made publicly accessible at https://osf.io/hmkjg/ and a persistent DOI will be issued and added to the published version. The deposit follows the FAIR data principles and is released under the MIT licence.

Ethics approval and consent to participate

This study used publicly available, de-identified data only. The study protocol was reviewed and approved by the Public Institutional Review Board under the Ministry of Health and Welfare, Republic of Korea (approval number P01-202602-01-029). The requirement for informed consent was waived by the IRB given the retrospective nature of the analysis and the use of de-identified data.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the Hwarangdae Research Institute, Korea Military Academy (Grant No. 2026B1013), which funded the analysis and writing of this study based on a previously released public-data cohort. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

J.Y.S. and W.J.C.: Conceptualization, Supervision, Project administration, Funding acquisition, Resources, Writing – review & editing. J.I.C.: Formal analysis, Data curation, Writing – original draft, Visualization (main figures and tables). Y.K.J.: Methodology, Validation, Investigation, Visualization (supplementary figures), Writing – review & editing. J.Y.S. and W.J.C. share senior authorship and are co-corresponding authors. All authors have read and approved the final version of the manuscript.

Acknowledgements

The authors thank the Korea Institute of Sports Science and the Korea Sports Promotion Foundation for operating the National Fitness Award Program and providing the open dataset used in this study.

References

1. World Health Organization. Obesity and Overweight: Fact Sheet (WHO, 2024).
2. Jeong, S. M. et al. 2023 obesity fact sheet: prevalence of obesity and abdominal obesity in adults, adolescents, and children in Korea from 2012 to 2021. J. Obes. Metab. Syndr. **33**, 27–35 (2024).
3. Korean Society for the Study of Obesity. 2024 Obesity Fact Sheet (KSSO, Seoul, 2024).
4. Bull, F. C. et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br. J. Sports Med. **54**, 1451–1462 (2020).
5. Garber, C. E. et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med. Sci. Sports Exerc. **43**, 1334–1359 (2011).

6. Liguori, G., Feito, Y., Fountaine, C. & Roy, B. A. (eds). ACSM's Guidelines for Exercise Testing and Prescription 11th edn (Wolters Kluwer, Philadelphia, 2021).
7. Amare, F., Alemu, Y., Enichalew, M., Demilie, Y. & Adamu, S. Effects of aerobic, resistance, and combined exercise training on body fat and glucolipid metabolism in inactive middle-aged adults with overweight or obesity: a randomized trial. BMC Sports Sci. Med. Rehabil. **16**, 189 (2024).
8. Cheng, H. et al. The effects of HIIT and MICT on body fat composition and cardiopulmonary fitness in adults: a meta-analysis of randomized controlled trials. BMC Sports Sci. Med. Rehabil. **18**, 94 (2026).
9. Bouakkar, J. et al. Sex differences in the physiological responses to cardiac rehabilitation: a systematic review. BMC Sports Sci. Med. Rehabil. **16**, 74 (2024).
10. Hayes, A. F. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach 3rd edn (Guilford, New York, 2022).
11. Preacher, K. J. & Hayes, A. F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav. Res. Methods **40**, 879–891 (2008).
12. Baron, R. M. & Kenny, D. A. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol. **51**, 1173–1182 (1986).
13. Preacher, K. J. & Hayes, A. F. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav. Res. Methods Instrum. Comput. **36**, 717–731 (2004).
14. Baruth, M. et al. Psychosocial mediators of physical activity and fitness changes in the Activity Counseling Trial. Ann. Behav. Med. **39**, 274–289 (2010).
15. Wu, X. et al. Physical exercise and depression in university students with psychological resilience as mediator. Sci. Rep. **15**, 29714 (2025).
16. Yoon, D. H., Kim, J.-H. & Lee, S.-U. A study on the development of a fitness age prediction model: the National Fitness Award cohort study 2017–2021. BMC Public Health **24**, 2606 (2024).
17. Jeoung, B. & Kim, J. A cross-sectional study of physical fitness and risk of hypertension in Korean adults aged 20–59. Healthcare (Basel) **11**, 2057 (2023).
18. Kim, S.-W., Park, H.-Y., Jung, H., Lee, J. & Lim, K. Estimation of health-related physical fitness using multiple linear regression in Korean adults: National Fitness Award 2015–2019. Front. Physiol. **12**, 668055 (2021).
19. Zhao, X., Lynch, J. G. Jr & Chen, Q. Reconsidering Baron and Kenny: myths and truths about mediation analysis. J. Consum. Res. **37**, 197–206 (2010).
20. Cohen, J. Statistical Power Analysis for the Behavioral Sciences 2nd edn (Erlbaum, Hillsdale, 1988).
21. Cruz-Jentoft, A. J. et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing **48**, 16–31 (2019).
22. Yang, M. et al. Sarcopenia for predicting falls and hospitalization in community-dwelling older adults: EWGSOP versus EWGSOP2. Sci. Rep. **9**, 17636 (2019).

23. Kraemer, W. J. & Ratamess, N. A. Fundamentals of resistance training: progression and exercise prescription. Med. Sci. Sports Exerc. **36**, 674–688 (2004).
24. Al-Mhanna, S. B. et al. Impact of resistance training on cardiometabolic health-related indices in patients with type 2 diabetes and overweight/obesity: a systematic review and meta-analysis of randomised controlled trials. Br. J. Sports Med. **59**, 733–746 (2025).
25. Strasser, B. & Schobersberger, W. Evidence for resistance training as a treatment therapy in obesity. J. Obes. **2011**, 482564 (2011).
26. Hunter, S. K. Sex differences in human fatigability: mechanisms and insight to physiological responses. Acta Physiol. (Oxf.) **210**, 768–789 (2014).
27. Tarnopolsky, M. A. Sex differences in exercise metabolism and the role of 17-beta-estradiol. Med. Sci. Sports Exerc. **40**, 648–654 (2008).
28. Geer, E. B. & Shen, W. Gender differences in insulin resistance, body composition, and energy balance. Gend. Med. **6 Suppl 1**, 60–75 (2009).
29. Spina, R. J. et al. Differences in cardiovascular adaptations to endurance exercise training between older men and women. J. Appl. Physiol. **75**, 849–855 (1993).
30. Tarnopolsky, L. J. et al. Gender differences in substrate for endurance exercise. J. Appl. Physiol. **68**, 302–308 (1990).
31. Hang, S., Xiaoyu, L., Jue, W., Yingli, L. & Li, Z. Effects of resistance training and aerobic training on improving the composition of middle-aged adults with obesity in an interventional study. Sci. Rep. **15**, 33972 (2025).
32. Mengistu, F. A., Lake, Y. A., Andualem, M. E., Miherete, Y. D. & Zewdie, S. A. Impact of aerobic, resistance, and combined training on cardiometabolic health-related indicators in inactive middle-aged men with excess body weight and obesity. Front. Physiol. **16**, 1519180 (2025).
33. Maher, C. et al. Weekly, seasonal, and festive period weight gain among Australian adults. JAMA Netw. Open **6**, e2326038 (2023).
34. Altman, D. G. & Bland, J. M. Statistics notes: quartiles, quintiles, centiles, and other quantiles. BMJ **309**, 996 (1994).

Tables

Table 1. Baseline characteristics of the overall overweight sample (n = 12,980)

Variable	DG (n=3,259)	MG (n=6,476)	IG (n=3,245)	F	p	Post-hoc
Age (yr)	37.76±14.07	39.56±14.29	35.69±13.21	84.182	<.001***	IG<DG<MG
Height (cm)	170.00±8.29	170.28±8.45	171.71±7.84	42.626	<.001***	DG,MG<IG
Body weight (kg)	78.23±8.45	78.00±8.42	79.01±7.94	16.051	<.001***	DG,MG<IG
Body fat (%)	28.74±6.87	27.75±7.02	26.48±6.92	86.052	<.001***	IG<MG<DG
BMI (kg/m²)	27.00±1.37	26.84±1.33	26.75±1.29	29.790	<.001***	IG<MG<DG
Waist (cm)	90.11±6.56	89.74±6.30	89.22±6.58	7.484	<.001***	IG<DG,MG
SBP (mmHg)	134.32±13.50	134.35±13.60	132.48±13.35	23.023	<.001***	IG<DG,MG
DBP (mmHg)	81.24±10.10	81.40±10.25	80.22±10.44	14.842	<.001***	IG<DG,MG
Grip-L (kg)	38.33±9.85	39.19±10.20	41.08±10.15	64.287	<.001***	DG<MG<IG
Grip-R (kg)	40.62±10.31	41.55±10.73	43.47±10.62	62.026	<.001***	DG<MG<IG
Rel. grip (%)	52.41±10.51	53.79±10.99	55.56±11.00	68.025	<.001***	DG<MG<IG
Interval (days)	176.82±113.69	168.83±117.99	211.52±116.64	148.444	<.001***	MG<DG<IG

*Values are mean ± SD. DG, Decreased Group; MG, Maintained Group; IG, Increased Group; BMI, body mass index; SBP/DBP, systolic/diastolic blood pressure; Grip-L/Grip-R, left/right grip strength; Rel. grip, relative grip strength (right-grip/body weight × 100). F and p from one-way ANOVA; post-hoc by Bonferroni. *** p < .001.*

Table 2. Repeated-measures ANCOVA of principal physical-fitness outcomes by BMI-change group, in the overall, male, and female overweight samples

Principal fitness outcome	Overall (n = 12,980) Time×Group F (p) ηp²	Male (n = 10,173) Time×Group F (p) ηp²	Female (n = 2,807) Time×Group F (p) ηp²
Sit-and-reach (cm)	25.543 (<.001)*** .002	21.998 (<.001)*** .002	3.162 (.042)* .001
Cross sit-ups (reps)	30.942 (<.001)*** .002	32.714 (<.001)*** .003	1.281 (.278) .000
Standing long jump (cm)	11.160 (<.001)*** .001	8.411 (<.001)*** .001	1.419 (.242) .001
VO_2max-shuttle	47.257 (<.001)*** .017	35.454 (<.001)*** .016	16.517 (<.001)*** .028
VO_2max-step	139.183 (<.001)*** .021	115.688 (<.001)*** .022	34.645 (<.001)*** .027
Reaction time (s)	7.262 (<.001)*** .001	3.513 (.030)* .000	1.433 (.239) .001
Rel. grip strength (%)	139.671 (<.001)*** .011	120.212 (<.001)*** .012	23.273 (<.001)*** .008

*Repeated-measures ANCOVA Time × Group interaction F (p) and partial eta-squared (ηp²) for the seven principal BMI-independent fitness outcomes. Covariates: pre-measurement BMI, age, revisit interval, and (in the overall analysis only) sex. Effect-size interpretation per Cohen [30]: small ≥ .01, medium ≥ .06, large ≥ .14. After Benjamini–Hochberg FDR correction within the seven principal outcomes per stratum, all rows are FDR-significant (p < .05) except sit-and-reach in women (raw p = .042, FDR p > .05) and reaction time in men (raw p = .030, FDR p > .05). Pre/post cell means by group are reported in Supplementary Table S5 (consolidated overall+male+female). * p < .05; *** p < .001.*

Table 3. Path-analytic decomposition of associations between exercise prescription characteristics and changes in principal fitness outcomes into BMI-related covariance components and direct components, by sex

Outcome	Path	Overall (n = 10,033)	Male (n = 7,761)	Female (n = 2,272)
Sit-and-reach (cm)	**b (ΔBMI→ΔY)**	−0.700 [−0.783, −0.613]*	−0.732 [−0.833, −0.632]*	−0.606 [−0.751, −0.456]*
	c_2' (Aer direct)	+0.227 [+0.019, +0.441]*	+0.361 [+0.117, +0.612]*	−0.202 [−0.571, +0.171]
	c_3' (n_total direct)	+0.014 [−0.008, +0.034]	+0.032 [+0.008, +0.058]*	−0.053 [−0.098, −0.013]*
Cross sit-ups (reps)	**b (ΔBMI→ΔY)**	−0.991 [−1.136, −0.840]*	−1.021 [−1.185, −0.859]*	−0.796 [−1.079, −0.490]*
	c_3' (n_total direct)	−0.002 [−0.035, +0.034]	−0.002 [−0.038, +0.034]	−0.082 [−0.139, −0.023]*
Standing long jump (cm)	**b (ΔBMI→ΔY)**	−1.579 [−1.951, −1.213]*	−1.399 [−1.748, −1.017]*	−1.632 [−2.269, −1.015]*
	c_2' (Aer direct)	+1.593 [+0.712, +2.474]*	+1.253 [+0.318, +2.204]*	+0.898 [−0.754, +2.508]
VO_2max-shuttle	**b (ΔBMI→ΔY)**	−0.958 [−1.078, −0.831]*	−1.010 [−1.161, −0.862]*	−0.685 [−0.844, −0.542]*
	Total (n_total)	−0.013 [−0.058, +0.031]	+0.001 [−0.051, +0.051]	−0.056 [−0.110, −0.005]*
VO_2max-step	**a_2 (Aer→ΔBMI)**	+0.052 [−0.026, +0.127]	+0.102 [+0.015, +0.198]*	−0.099 [−0.256, +0.056]
	a_3 (n_total→ΔBMI)	−0.005 [−0.012, +0.003]	+0.000 [−0.007, +0.008]	−0.026 [−0.042, −0.009]*
	b (ΔBMI→ΔY)	−0.938 [−1.000, −0.877]*	−0.957 [−1.024, −0.886]*	−0.827 [−0.938, −0.722]*
	c_3' (n_total direct)	+0.015 [+0.003, +0.029]*	+0.015 [+0.000, +0.030]*	+0.012 [−0.012, +0.035]
	a_2·b (Aer × BMI cov.)	−0.049 [−0.120, +0.024]	−0.097 [−0.189, −0.015]*	+0.081 [−0.046, +0.211]
	a_3·b (n_total × BMI cov.)	+0.004 [−0.002, +0.011]	−0.000 [−0.007, +0.007]	+0.021 [+0.007, +0.036]*
Reaction time (s)	**a_2 (Aer→ΔBMI)**	+0.100 [+0.039, +0.161]*	+0.097 [+0.025, +0.168]*	+0.059 [−0.067, +0.178]
	a_3 (n_total→ΔBMI)	−0.001 [−0.006, +0.006]	+0.001 [−0.005, +0.009]	−0.014 [−0.027, −0.000]*
	b (ΔBMI→ΔY)	+0.001 [+0.001, +0.002]*	+0.001 [−0.000, +0.002]	+0.003 [+0.001, +0.005]*

Outcome	Path	Overall (n = 10,033)	Male (n = 7,761)	Female (n = 2,272)
Rel. grip strength (%)	**b (ΔBMI→ΔY)**	−1.753 [−1.860, −1.649]*	−1.807 [−1.931, −1.682]*	−1.376 [−1.547, −1.200]*
	c_3' (n_total direct)	+0.054 [+0.025, +0.082]*	+0.061 [+0.029, +0.094]*	+0.014 [−0.030, +0.063]
	Total (n_total)	+0.062 [+0.032, +0.091]*	+0.066 [+0.033, +0.101]*	+0.031 [−0.017, +0.083]

*Path coefficients are unstandardized B with bias-corrected and accelerated bootstrap 95% confidence intervals (5,000 resamples; 20,000 for three-decimal-precision outcomes). * 95% CI excludes zero. Rows are shown when at least one of c', a·b, or Total had a 95% CI excluding zero in any one of the three sex strata (overall, male, female); paths whose 95% CI included zero in all three strata are omitted from this table for compactness, and complete coefficients for all 14 outcomes are reported in Supplementary Tables S7 (overall), S8 (male) and S17 (female). Outcome-specific n differs from headline path-analytic n because of variable-wise outlier handling. Path notation: a = X→ΔBMI; b = ΔBMI→ΔY; c' = X→ΔY direct (with ΔBMI in model); a·b = X × BMI covariance component (the part of the X–ΔY association that travels with within-interval ΔBMI); Total = a·b + c'. Because ΔBMI and ΔY are measured over the same interval, a·b is interpreted descriptively as a covariance partition rather than as a temporally identified mediation effect; the directionality sensitivity (Supplementary Table S16) confirmed comparable a·b magnitudes when ΔBMI and ΔY are swapped. X variables: Resistance (Res) and Aerobic (Aer) are dummy-coded with Mixed as reference; n_total is the count of prescribed main exercise items. Resistance c_1' rows are not shown in this table because no Resistance direct effect (c_1') had a 95% CI excluding zero in any of the three sex strata (overall, male, female) for any principal outcome; complete Resistance coefficients are reported in Supplementary Tables S7 (overall), S8 (male) and S17 (female). Body-composition and blood-pressure outcomes were demoted to supplementary status because their b-paths are mathematically partly tautological vs. ΔBMI; full coefficients are in Supplementary Tables S7 (overall), S8 (male) and S17 (female). Aer = aerobic prescription; Res = resistance prescription.*

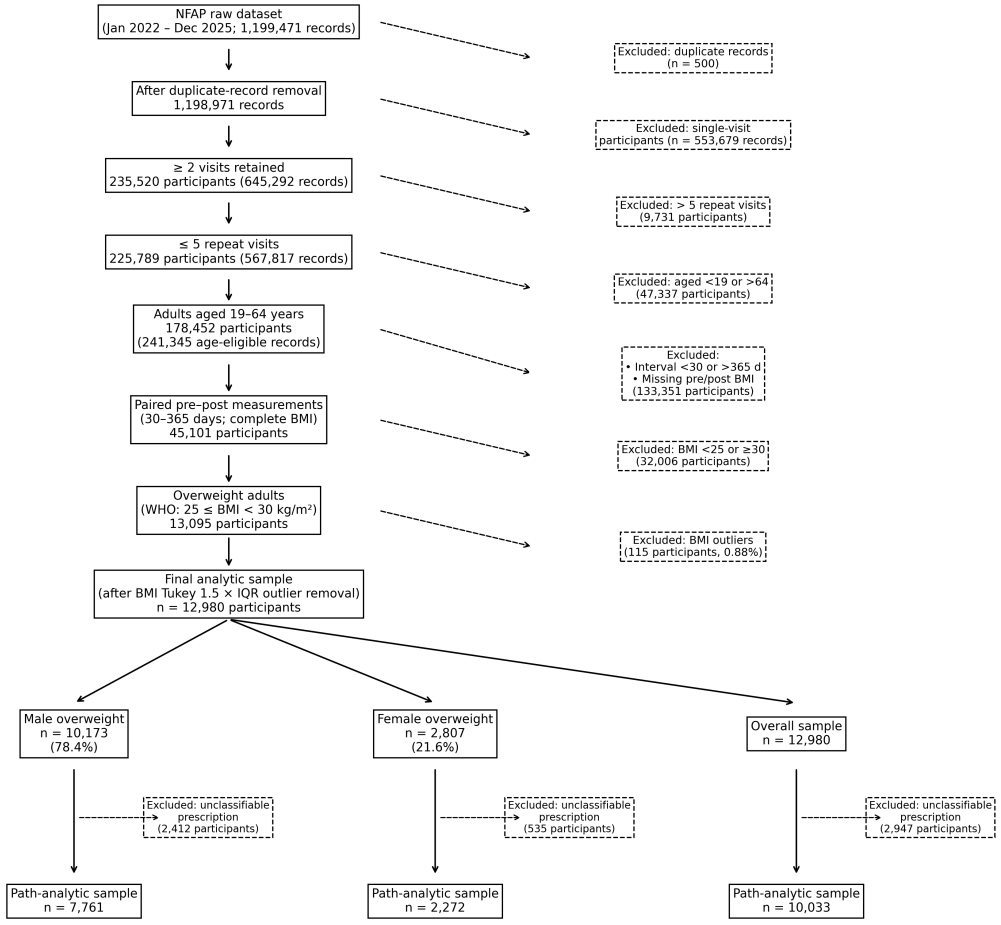
Table 4. Sex × Time × Group interaction summary, with FDR-adjusted p-values, for the seven principal fitness outcomes

Outcome (class)	Overall ηp²	Male ηp²	Female ηp²	Sex×Time×Group raw p	FDR p
Sit-and-reach (cm)	.002***	.002***	.001*	.182	.245
Cross sit-ups (reps)	.002***	.003***	.000 (ns)	.241	.289
Standing long jump (cm)	.001***	.001***	.001 (ns)	.396	.396
VO_2max-shuttle	.017***	.016***	.028***	.014*	.049*
VO_2max-step	.021***	.022***	.027***	.062	.145
Reaction time (s)	.001***	.000*	.001 (ns)	.318	.371
Rel. grip strength (%)	.011***	.012***	.008***	.006**	.028*

*Time × Group interaction ηp² from sex-stratified repeated-measures ANCOVA (overall, male, female). The Sex × Time × Group raw p-value tests whether the Time × Group interaction differs between men and women in a pooled ANCOVA model with sex as a three-way interaction factor. FDR p = Benjamini–Hochberg FDR-adjusted p-value within the seven-principal-outcome family. After FDR correction, the three-way interaction is statistically detectable for VO_2max-shuttle (FDR p = .049, larger response in women) and relative grip strength (FDR p = .028, larger absolute–relative divergence in men). For VO_2max-step the raw p (.062) is borderline but the FDR p (.145) does not survive multiplicity adjustment. * p < .05; ** p < .01; *** p < .001; ns, not significant.*

Figures

Fig. 1 Flow diagram of participant selection



NFAP, National Fitness Award Program; BMI, body mass index; IQR, interquartile range; WHO, World Health Organization.

Participant- and record-level counts are both reported where applicable. Path-analytic samples were restricted to participants whose main exercise prescription could be classified as Aerobic, Resistance, or Mixed; the exclusions at this step (2,412 men, 535 women, 2,947 overall) are shown explicitly to make the participant-flow diagram complete.

Figure 1

See image above for figure legend

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesVer.2.docx