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.

Effects of body mass index changes on body composition and physical fitness in overweight Korean adults: a sex-stratified path-analytic description of exercise prescription characteristics

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

Research Article

Keywords: overweight, body mass index, exercise prescription, body composition, physical fitness, path-analytic description, sex differences, National Fitness Award Program, bootstrap, Korean adults

Posted Date: April 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9511401/v1

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

Additional Declarations:
No competing interests reported.

Tables 1 to 13 are available in the Supplementary Files section.

Effects of body mass index changes on body composition and physical fitness in overweight Korean adults: a sex-stratified path-analytic description of exercise prescription characteristics

Jae-Il Choi [1], Yong-Kyun Jeon [2], Jeong-Yeop Shin [3]*, Won-Je Cho [4]*

[1] Graduate School of Education, Dankook University, Yongin-si, Gyeonggi-do, 16890, Republic of Korea

[2] Department of Physical Education, Dankook University, Yongin-si, Gyeonggi-do, 16890, Republic of Korea

[3] Department of Health and Sports, Gukje Cyber University, Suwon-si, Gyeonggi-do 16487, Republic of Korea

[4] Department of Physical Education, Korea Military Academy, Nowon-gu, Seoul 01805, Republic of Korea

Correspondence:

Jeong-Yeop Shin, Ph.D. — gjcu261@gjcu.ac.kr

Won-Je Cho, Ph.D. — wjcho0425@kma.ac.kr

Jeong-Yeop Shin and Won-Je Cho contributed equally as co-corresponding authors.

Abstract

Background

Overweight is a growing public-health burden in Korean adults. Large-scale evidence on how exercise prescription characteristics statistically pattern through body mass index (BMI) change to influence body composition and fitness, and whether these patterns differ by sex, remains limited.

Methods

From 1,199,471 National Fitness Award Program records (January 2022–December 2025), 12,980 overweight adults aged 19–64 years (10,173 men, 2,807 women) with paired measurements 30–365 days apart were categorized by quartiles of percentage BMI change. Repeated-measures ANCOVA and path-analytic models with BMI change as an intermediate variable (lavaan) were fitted; the principal interpretive focus was on BMI-independent fitness outcomes. Bootstrap 95% CIs (5,000 BCa resamples) and Benjamini–Hochberg FDR correction were applied. Analyses were sex-stratified, with formal sex × Time × Group tests and seven pre-specified sensitivity analyses.

Results

Among BMI-independent fitness outcomes, Time × Group interactions were largest for step-test VO_2max (partial η^2 [ηp^2] = .021), shuttle-run VO_2max (.017), and relative grip strength (.011). BMI change was

robustly associated with all fitness outcomes (FDR-adjusted $p < .001$). In men, aerobic prescriptions showed direct associations with flexibility ($c_2' = +0.361$) and power (+1.253), and an indirect association with step-test VO_2max via BMI ($a_2 \cdot b = -0.097$). In women, the count of prescribed items (n_total) was indirectly associated with step-test VO_2max ($a_3 \cdot b = +0.021$, 95% CI [+0.007, +0.036]) and blood pressure; direct type associations were not detected, partly reflecting limited power. Sex × Time × Group tests confirmed that the cardiorespiratory ηp^2 sex difference was significant for shuttle-run VO_2max (FDR-adjusted $p = .038$) but borderline for step-test VO_2max (FDR-adjusted $p = .087$). Relative grip strength improved in BMI-decreased and declined in BMI-increased groups despite comparable absolute grip gains, identifying it as a more sensitive functional indicator.

Conclusions

Path-analytic associations between prescription characteristics and fitness change, indexed through BMI change, followed sex-specific profiles: descriptively type-indexed in men and count-indexed in women (interpretation is hypothesis-generating and associative rather than causal). These observational findings are hypothesis-generating for sex-specific randomized trials. Relative grip strength is recommended as a functional fitness indicator in overweight populations.

Keywords

overweight; body mass index; exercise prescription; body composition; physical fitness; path-analytic description; sex differences; National Fitness Award Program; bootstrap; Korean adults

Background

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 1.27-fold increase (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 obesity 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 recognized as one of the most effective non-pharmacological interventions for managing overweight and obesity. The American College of Sports Medicine (ACSM) and the WHO recommend that adults perform at least 150 minutes of moderate-intensity physical activity or 75 minutes of vigorous-intensity physical activity per week [4,5]. Multiple studies have demonstrated that such exercise not only reduces body weight and body fat but also improves overall fitness, including cardiorespiratory endurance, muscular strength, and flexibility [6]. A recent randomized controlled trial (RCT) showed that 8 weeks of aerobic, resistance, or combined training significantly reduced body-fat percentage in inactive overweight or obese adults, with resistance and combined training producing superior improvements in glucose metabolism [7]. A meta-analysis of 19 RCTs reported that high-intensity interval training (HIIT) reduced

BMI in obese populations by a standardized mean difference of −0.59 and outperformed moderate-intensity continuous training in improving cardiorespiratory fitness [8].

Sex differences in exercise responses remain incompletely characterized. Neuromuscular fatigue research has shown that women are generally more fatigue-resistant than men at the same relative intensity, though the magnitude is task-dependent [9,10]. At the skeletal-muscle level, emerging molecular evidence indicates that differences in testosterone and estrogen influence muscle protein synthesis and hypertrophic responses [11]. A systematic review of 88 cardiac-rehabilitation studies reported similar responses between sexes for most physiological variables, although men showed larger improvements in VO_2max, functional capacity, 6-minute walking distance, and grip strength [12]. Whether these patterns are preserved in path-analytic models using BMI change as an intermediate variable, and whether specific dimensions of exercise prescription (type vs. count of prescribed items) align differently with outcomes in each sex, has not been sufficiently examined.

Path-analytic and mediation-style models have become useful tools for describing statistical links between exercise interventions and outcomes [13–15]. Bootstrap-based confidence intervals do not require normality and permit accurate estimation of indirect effects even in small samples, and are now standard [13,16]. Recent studies have used mediation analysis to explore mechanisms of exercise interventions. For example, in the Activity Counseling Trial of 878 inactive adults, changes in behavioral processes from baseline to 24 months significantly mediated the relationship between the intervention arm and improvements in physical activity and cardiorespiratory fitness [17]. In a study of university students, exercise was associated with lower depression and higher psychological resilience, with resilience mediating the exercise–depression relationship [18]. However, no large-scale study to date has examined — with sex stratification — how exercise prescription characteristics are patterned through BMI change into body composition and fitness outcomes in overweight adults.

The Korean National Fitness Award Program (NFAP) is a nationwide fitness-management program operated by the Korea Institute of Sports Science under the Korea Sports Promotion Foundation. Across 75 certification centers in 17 metropolitan cities and provinces, the program uses standardized protocols to measure body composition, blood pressure, grip strength, flexibility, muscular endurance, power, cardiorespiratory endurance, and agility, and provides individually tailored exercise prescriptions [19,20]. Over the past decade, the NFAP has accumulated data from more than 500,000 Korean adults, used for studies of fitness distributions, fitness-age prediction, and fitness–hypertension associations [19–21]. To our knowledge, however, no previous study has used NFAP data to examine sex-stratified paths linking exercise prescription characteristics with BMI change in overweight adults.

Framing of the present analysis

Before stating objectives, we explicitly delimit the causal claims of this study. 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 [13,15]. 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 path-analytic descriptions 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 NFAP, we described differences in body composition, blood pressure, grip strength, and physical fitness across BMI change profiles (decreased, maintained, increased). Second, we estimated direct and indirect associations of qualitative exercise prescription (resistance, aerobic, or combined type) and the quantitative count of prescribed main exercise items with 14 outcomes, using BMI change as the intermediate variable. Third, all analyses were stratified by sex, and sex × Time × Group interactions were formally tested, to quantitatively compare paths between men and women. As the first path-analytic study using the NFAP dataset, we aim to generate hypotheses for future randomized evaluation of sex-specific exercise prescription strategies for overweight adults.

Methods

Study design and data source

This was a retrospective observational study based on publicly available NFAP data. The NFAP is a nationwide fitness-management program operated by the Korea Institute of Sports Science under the Korea Sports Promotion Foundation. Across 75 certification centers in 17 metropolitan cities and provinces, the program provides standardized fitness assessments and tailored exercise prescription services, with measurements and prescription content stored in an integrated information system [19,20]. The raw data consisted of the "Comprehensive Data on Fitness Measurement and Exercise Prescription" files available through the Korean government's Open Data Portal (data.go.kr), comprising 1,199,471 fitness measurement 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, Republic of Korea (approval number: P01-202602-01-029; the IRB approval in February 2026 covered retrospective analysis of previously collected data (January 2022–December 2025)), in accordance with the Declaration of Helsinki. The study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. 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-center identifiers were retained to permit center-level sensitivity analyses.

Measurement standardization and reliability

All 75 NFAP certification centers follow a standardized measurement manual issued by the Korea Institute of Sports Science (cardiorespiratory fitness assessed using either the 20-m multi-stage shuttle-run test with VO_2max estimated via the Léger regression equation [22], or the Queen's College 3-minute step test at 24 steps/min with VO_2max estimated from post-exercise heart rate recovery [23]); staff at each center undergo mandatory training, and equipment is calibrated annually under a unified protocol [19,20]. Previous validation work within the NFAP has reported acceptable inter-center reliability for the fitness tests used here. We were unable to independently re-estimate intra-rater or inter-rater reliability in the present dataset because individual measurement attempts are not stored. We therefore supplement the analysis with a minimal detectable change (MDC) calibration: for each continuous outcome, we contextualize the observed mean change against previously published MDC or smallest worthwhile change values where available [24,25], and flag outcomes for which group-level mean changes fell below plausible MDC (see MDC calibration analysis in the Sensitivity analyses subsection of Methods).

Participants and selection criteria

[Insert Figure 1 around here] Fig. 1 Flow diagram of participant selection. Multi-step STROBE-style diagram showing the sequential filtering of the Korean National Fitness Award Program (NFAP) raw dataset (1,199,471 records, January 2022–December 2025) to the final analytic sample of 12,980 overweight adults (10,173 men, 2,807 women), and further to the sex-stratified path-analytic samples (male, n = 7,761; female, n = 2,272; overall, n = 10,033). Path-analytic indirect-effect analyses were restricted to participants whose main exercise prescription could be classified as Aerobic, Resistance, or Mixed. NFAP, National Fitness Award Program; BMI, body mass index; IQR, interquartile range; WHO, World Health Organization.

Participant selection followed the multi-step STROBE-style flow diagram shown in Figure 1. From the 1,199,471 raw NFAP records, we first extracted participants with two or more visits, yielding 235,520 individuals (645,292 records). Participants whose visit order exceeded five were excluded, as their measurement patterns might differ from general users, leaving 225,789 individuals (567,817 records). We then restricted the sample to adults aged 19–64 years, yielding 241,345 records. To compare paired first (pre-) and second (post-) measurements in the same participant, we included only individuals whose revisit interval was 30–365 days. Chronic exercise adaptations typically become measurable approximately 30 days after regular training begins [24], and meta-analytic evidence suggests that at least four weeks of intervention are required for meaningful changes in body composition, cardiorespiratory fitness, and muscular strength [25]. Conversely, follow-up periods exceeding one year can introduce cumulative effects of unmeasured time-dependent confounders such as aging, lifestyle change, and seasonal variation, obscuring the true effect of exercise [26]. To balance detection of exercise effects against confounding control, we adopted a 30–365-day interval. Applying this criterion yielded 45,101 participants with complete pre- and post-BMI data.

Next, we extracted 13,095 overweight adults whose pre-measurement BMI was 25–29.9 kg/m² according to WHO criteria [1]. We used pre-measurement BMI rather than post-measurement BMI to define group membership based on each participant's baseline status and to prevent post-measurement BMI changes from

influencing group assignment. To minimize the impact of measurement errors and extreme outliers, we applied Tukey's 1.5× interquartile range (IQR) rule: participants with BMI outliers were excluded from all analyses, whereas outliers on other variables were treated as missing values for that variable only. For analyses involving outcomes with high variable-specific missingness (notably the shuttle-run and step tests for cardiorespiratory fitness, with missingness rates of approximately 70% and 40%, respectively), we used complete-case analysis on an outcome-by-outcome basis (listwise deletion within each outcome-specific path model) rather than multiple imputation or full-information maximum likelihood (FIML). The outcome-specific analytic samples (e.g., VO_2max-step: n = 1,006 female, 5,215 male; waist circumference: n = 1,200 female, 5,189 overall) are smaller than the overall path-analytic samples (female n = 2,272; male n = 7,761) and are reported in Supplementary Table S1. The missingness is best characterized as Missing Not At Random (MNAR) by design — dependent on the test assignment protocol (shuttle-run vs step test) rather than on the unobserved outcome value itself. Complete-case analysis within each outcome-specific model was therefore the most defensible approach, under the working assumption that within-outcome complete cases approximate MAR conditional on the measured covariates. This assumption was empirically evaluated in Supplementary Table S11: across the three cardiorespiratory-test assignment groups (shuttle-run only, n = 3,322; step-test only, n = 7,286; both or neither, n = 2,372), 11 of 14 continuous baseline covariates (78.6%) showed an absolute standardized mean difference (|SMD|) ≤ 0.10. The largest imbalances were in age (SMD = 0.281), revisit interval (SMD = 0.362), and blood pressure (SMD = 0.114–0.128) — all of which were included as covariates in subsequent path-analytic models — whereas variables directly relevant to cardiorespiratory fitness (BMI, body weight, body fat, grip strength; |SMD| ≤ 0.04) were balanced. This pattern is consistent with MNAR-by-design where assignment depends on observable pre-test characteristics rather than on the unobserved outcome value. Formal MNAR sensitivity analyses (e.g., pattern-mixture or tipping-point analyses) were not performed; this is acknowledged as a limitation and is a priority for future work. This choice was made for four reasons: (i) the missingness was structural — participants were typically assigned to either the shuttle-run or the step test (not both) based on physical capacity, age, and center policy, which is not the form of MAR required by multiple imputation; (ii) imputing outcome values from covariates or from the alternative test would propagate modelling assumptions into the primary estimates without adding information; (iii) the complete-case subsamples for each outcome (ranging from approximately 3,662 to 12,955 participants) remained large enough to support well-powered estimation; and (iv) empirical balance checks across test-assignment groups (Supplementary Table S11) showed |SMD| ≤ 0.10 for 11 of 14 baseline covariates, supporting the MAR-conditional-on-covariates working assumption. Variable-wise missingness after outlier exclusion is reported in Supplementary Table S1. This procedure yielded a final analytic sample of 12,980 participants (the 115 participants excluded as BMI outliers represent 0.88% of the 13,095 WHO-overweight sample; variable-wise missingness after this step is summarized in Supplementary Table S1). For sex-stratified analyses, the final sample was divided by sex (10,173 men and 2,807 women), and outlier detection was repeated within each sex stratum to reflect sex-specific distributions.

Exercise prescription parsing, clustering, and visualization

All analyses of the unstructured-text exercise prescriptions, and the associated visualizations, were

performed in R (R Foundation for Statistical Computing, Vienna, Austria). The analysis proceeded in three steps: data preprocessing, keyword clustering, and semantic network visualization.

1) Data preprocessing

The NFAP prescription content is stored as semi-structured free text in the format "Warm-up: … / Main exercise: … / Cool-down: …," requiring regular-expression-based parsing. We focused the present analysis on the "main exercise" component because (a) it is defined in the NFAP manual as the principal conditioning stimulus and (b) it is the only component consistently populated across centers; warm-up and cool-down fields are often abbreviated or omitted. We acknowledge that cool-down content, which is dominated by stretching, could be directly relevant to flexibility outcomes; a sensitivity analysis incorporating cool-down content is reported below (see Sensitivity analyses). Individual exercise items within the main exercise field were tokenized using commas as delimiters.

2) Keyword clustering

To reduce dimensionality and ensure international comparability, we performed keyword clustering. More than 500 fragmented Korean exercise terms were consolidated into 18 major English categories based on similar exercise characteristics and goals (a mapping table for the 30 most frequent Korean terms to these English categories is provided in Supplementary Table S7 to support reproducibility) (e.g., Walking, Running, Cycling, Swimming/Aqua, Step/Ladder Training, Jumping Exercises, Burpees/Circuit, Weight Training, Resistance Band, Bodyweight Routine, Push-ups, Squats and Seated, Core/Abs, Yoga/Pilates, Stretching, Balance Training, Ball Exercises, and Other Exercises). Using predefined rules, each prescription was classified into one of three categories: (1) Aerobic (aerobic exercises only), (2) Resistance (resistance exercises only), or (3) Mixed (both aerobic and resistance exercises). Prescriptions without either type, or with zero main exercise items, were treated as missing in the path-analytic models. The count of main exercise items (n_total) was also calculated as a numerical indicator of how many distinct items the practitioner prescribed.

We emphasize explicitly that n_total is a count of prescribed items and not a measure of training volume, energy expenditure, or physiological dose. A prescription of five stretching items differs substantially in training stimulus from a prescription of five compound resistance items, even though both share n_total = 5. Furthermore, because NFAP records prescription content and not execution logs, n_total should be interpreted as an index of the count of prescribed items rather than cumulative physiological load. In all references below, we use the terminology "count of prescribed items (n_total)" and avoid the terms "volume" or "diversity," which we previously used but which are imprecise. Among the 10,033 participants included in the overall path-analytic model, the distribution was 3,013 (30.0%) Resistance, 2,117 (21.1%) Aerobic, and 4,903 (48.9%) Mixed.

3) Semantic network analysis and visualization

To characterize exercise prescription patterns within each group, we conducted a semantic network analysis using the widyr and igraph packages. The co-occurrence frequency of exercise items within each prescription was computed to construct a co-occurrence matrix. The final network visualization was

produced using the Fruchterman–Reingold layout algorithm implemented in the ggraph package. Node size was scaled proportionally to the cumulative prescription frequency of each exercise across all data, and edge thickness was scaled to the co-prescription frequency between pairs of exercises.

[Insert Figure 2 around here] Fig. 2 Semantic network of clustered exercise prescriptions by prescription type. Three-panel co-occurrence network showing the clustering of 18 standardized English exercise categories within each of the three prescription types: Aerobic Only (left), Resistance Only (center), and Mixed (Aerobic + Resistance, right). Node size is proportional to the cumulative prescription frequency of each exercise category across the full dataset (globally normalized across all three panels for cross-panel comparability; the three prescription-type subsamples contain n = 2,117 Aerobic, n = 3,013 Resistance, and n = 4,903 Mixed participants, respectively); edge thickness reflects the co-prescription frequency between pairs of categories within the same prescription. The network was constructed using the widyr and igraph R packages and visualized with the Fruchterman–Reingold layout algorithm implemented in the ggraph package (set.seed(2025) for layout reproducibility). This visualization supports the operational definition of the three prescription-type categories used throughout the path-analytic models.

BMI change group classification

The percentage change in BMI between pre- and post-measurements was computed as ((post-BMI − pre-BMI) / pre-BMI) × 100 (%). When categorizing a continuous variable into groups, the use of sample-based quantiles as cut-off points is a statistically valid approach [27]. Quantile-based classification reflects the actual data distribution without requiring arbitrary cut-off values [28]. We used the first quartile (Q1, 25th percentile) and third quartile (Q3, 75th percentile) of the BMI change distribution in the overweight sample (n = 12,980) as cut-off points. Participants with BMI percentage change below Q1 were assigned to the Decreased Group (DG; change < −2.77%), those between Q1 and Q3 to the Maintained Group (MG; change ≥ −2.77% and ≤ +1.19%), and those above Q3 to the Increased Group (IG; change > +1.19%). Using Q1 and Q3 separates the extreme 25% at each end of the distribution and groups the middle 50%, reducing sensitivity to outliers [29,30].

We note that quartile cut-offs are sample-dependent and do not carry intrinsic clinical meaning. In particular, the DG cut-off of −2.77% corresponds to approximately 2 kg of weight loss at the mean baseline weight in this sample, which may fall within the measurement uncertainty for body weight at clinical scales. We therefore complement the primary analysis with sensitivity analyses using clinically anchored cut-offs of ±3% and ±5% BMI change, the latter corresponding to the 5% weight-loss threshold conventionally regarded as clinically meaningful in obesity-intervention research [see Sensitivity analyses].

Statistical analysis

All analyses were conducted in R 4.5.3 (R Core Team, 2026). Baseline characteristics were compared using one-way analysis of variance (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 analysis of covariance (repeated-measures ANCOVA). The model was specified as value ~ Time × Group + covariates, with pre-BMI, age, revisit interval, and sex (in the overall analysis only) included as covariates.

F statistics and p-values were calculated using Type III sums of squares, and effect sizes were reported as partial eta-squared (partial η^2; ηp^2). Effect sizes were interpreted following Cohen's conventions: negligible (< .01), small (.01–.06), medium (.06–.14), or large (≥ .14) [31]. Because the prescription-type assignment was not randomized, all between-group comparisons should be understood as associative.

Path-analytic association model

To describe statistical associations between exercise prescription characteristics and changes in body composition and physical fitness with BMI change as an intermediate variable, we applied a simple single-mediator-form model using the lavaan package (version 0.6-19) [13,15]. The model was specified as follows:

Intermediate-variable model (path a):

$$\Delta BMI = a_1 \cdot Resistance + a_2 \cdot Aerobic + a_3 \cdot n_total + covariates + e_1$$

Outcome model (paths b and c′):

$$\Delta Y = b \cdot \Delta BMI + c_1' \cdot Resistance + c_2' \cdot Aerobic + c_3' \cdot n_total + covariates + Y_pre + e_2$$

Resistance and Aerobic were dummy-coded indicators for exercise prescription type (Mixed was used as the reference group, chosen because it had the largest subsample size and thus maximized statistical power for pairwise comparisons). The outcome model adjusted for the pre-measurement value of the outcome (Y_pre) to control for regression to the mean, and for age, revisit interval, and (in the overall sample) sex. Parameters were estimated by maximum likelihood. Indirect effects (a × b) and total effects ($c' + a \cdot b$) were computed using lavaan's := operator. All 95% confidence intervals were computed using 5,000 BCa bootstrap resamples to avoid reliance on normality assumptions [13,14,16]. We follow the convention of retaining paths for which the 95% CI excluded zero, while also reporting FDR-adjusted p-values (see Multiple testing below).

We stress that the term "mediation" is used here in a statistical-modelling sense only. Because BMI change and outcome change were observed simultaneously, temporal precedence between the intermediate variable and the outcome is not established, and strict causal mediation assumptions are not met. The indirect effects (a × b) should therefore be read as "the part of the association between X and Y that is statistically patterned through ΔBMI," not as a causal decomposition. This framing is carried consistently into the Results and Discussion.

Outcome prioritization: BMI-independent fitness outcomes

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 (ΔBMI → ΔY) for body weight, waist circumference, and body-fat percentage is partly tautological rather than substantively informative [32]. We therefore pre-designated the principal scientific claims to the seven BMI-independent fitness outcomes: shuttle-run VO_2max, step-test VO_2max, sit-and-reach (flexibility), cross sit-ups (muscular endurance), standing long jump (power), reaction time, and

relative grip strength. Blood pressure and absolute grip strength were treated as clinically relevant secondary outcomes. Body weight, waist circumference, and body-fat percentage were analyzed and reported as supplementary outcomes, interpreted as descriptions of the BMI-change process itself rather than independent mediation findings. Tables that combine all 14 outcomes are typographically separated into "principal," "secondary," and "supplementary (BMI-dependent)" sections.

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 computed sex-difference p-values. We also fitted 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 to test whether the prescription-type and item-count associations differed by sex.

Multiple testing

Because we fitted path models for 14 outcomes × three exercise-prescription predictors in two sex strata, and report up to seven parameters per model, the total number of hypothesis tests is large. The analyses are explicitly exploratory and hypothesis-generating. Nonetheless, to guard against over-interpretation we applied Benjamini–Hochberg FDR correction [reported alongside raw p-values] separately within (i) the seven principal BMI-independent fitness outcomes and (ii) each family of paths (a, b, c′, a·b). The results reported as statistically detectable in the abstract and the conclusions retain significance at FDR-adjusted $p < .05$ within the principal-outcome family.

Sensitivity analyses

Seven pre-specified sensitivity analyses address the main design and modelling concerns.

(i) Alternative BMI-change cut-offs. We re-ran ANCOVA and path models using clinically anchored cut-offs of ±3% and ±5% BMI change, and report ηp^2 and indirect-effect estimates alongside the quartile-based primary analyses. Conclusions about the principal fitness outcomes were considered robust only if they were reproduced at the ±3% cut-off.

(ii) Propensity-weighted prescription type. Because exercise prescription type was not randomly assigned, we estimated propensity scores for each prescription type (Resistance, Aerobic, Mixed) using multinomial logistic regression with baseline covariates (age, pre-BMI, pre-outcome value, sex, revisit interval, center) as predictors. Inverse-probability-of-treatment weighting (IPTW) with stabilized weights was applied, and path models were refitted under IPTW. Balance diagnostics (standardized mean differences before and after weighting) are reported in Supplementary Table S2. Full bootstrap results for all 13 path-analytic coefficients × 14 outcomes, stratified by sex, are provided in Supplementary Tables S4 (overall), S5 (male), and S6 (female).

(iii) Bootstrap down-sampling of men. To disentangle a genuinely absent direct prescription-type association in women from reduced statistical power, we randomly drew 1,000 subsamples of men at the female path-analytic sample size (n = 2,272, rather than the broader female overweight sample of 2,807, to

match the actual sample used in the female path model) and refitted the male path-analytic model in each subsample. The proportion of subsamples in which the male direct effects (c_1', c_2') reached 95%-CI significance provides a post-hoc power estimate for the female analyses.

(iv) Center-level clustering. Participants were nested within 75 certification centers. We refitted principal outcome models using cluster-robust standard errors (center-level sandwich estimator) to assess whether center clustering materially changed inference.

(v) n_total decomposition. To probe the content of the n_total association in women, we decomposed n_total into modality-specific counts (resistance-item count, aerobic-item count, other-item count) and refitted the path-analytic model with these sub-counts replacing n_total. This analysis addresses whether the n_total association is driven by a specific modality or by overall item count irrespective of content.

(vi) Minimal detectable change (MDC) calibration. For each principal fitness outcome, group-level mean changes were compared against published MDC or smallest worthwhile change values [24,25] where available, and outcomes for which the observed group-level changes fell below plausible MDC thresholds were flagged.

(vii) Cool-down content inclusion. To assess whether cool-down stretching content attenuated the principal flexibility findings, we re-fitted the path-analytic model for flexibility and muscular endurance with an additional covariate reflecting the number of stretching items prescribed in the cool-down component; principal findings were considered robust if they were preserved after this adjustment. Full tables for all seven sensitivity analyses are reported in Supplementary Material.

Assumptions and scope of inference

The path-analytic framework above rests on assumptions that are only partially satisfied by the present data. Temporal precedence X → M → Y is satisfied for the X → M segment (prescription was given at the pre-measurement, before BMI change occurred), but not for M → Y (ΔBMI and ΔY were observed over the same interval). Sequential ignorability requires no unmeasured confounders of the X–M, X–Y, and M–Y relationships [15,32]; despite controlling for age, sex, revisit interval, pre-BMI, pre-outcome value, center, and (in IPTW analyses) the propensity of prescription type, we cannot exclude unmeasured confounding by dietary intake, sleep, stress, exercise adherence, or practitioner characteristics. We accordingly interpret path-analytic estimates as conditional associations and restrict our claims to hypothesis-generating inference. Further limitations inherent to observational NFAP data, the mathematical dependencies between BMI and body-composition outcomes, and multiple testing are discussed in the Limitations section.

Results

We report results in three strata (overall, men, women) followed by a formal sex comparison. Within each stratum, we separate principal fitness outcomes (BMI-independent), secondary clinical outcomes (blood pressure, absolute grip strength), and supplementary BMI-dependent outcomes (body weight, waist circumference, body-fat percentage). All p-values accompanied by "FDR" have been adjusted for multiple

testing within the principal-outcome family; raw p-values and BCa 95% CIs are retained for all paths.

1. Overall overweight sample

Baseline characteristics

[Insert Table 1 around here]

Baseline characteristics of the 12,980 overweight adults are presented in Table 1. All 12 baseline variables differed significantly across the three groups (all $p < .001$). Age followed IG < DG < MG, with IG being youngest ($F = 84.18$). Body-fat percentage, BMI, and waist circumference followed IG < MG < DG, indicating a regression-to-the-mean pattern that we partly address by including pre-BMI and pre-outcome values as covariates in all path-analytic models; we note, however, that covariate adjustment does not fully eliminate regression-to-the-mean (RTM) bias. Formal RTM-correction methods (e.g., Oldham's method, Blomqvist's approach) were not applied as primary analyses; however, alternative analyses using ΔBMI as a continuous predictor (rather than quartile-based DG/MG/IG) yielded qualitatively concordant principal findings and are provided via the reviewer-accessible OSF repository [26,32]. Blood pressure was lower in IG than in DG and MG. Grip strength followed the reverse order (DG < MG < IG), suggesting that the BMI-increased group may have gained muscle mass. The revisit interval was longest in IG (mean 211.52 days; $F = 148.44$, $p < .001$). Prescription-type distribution across groups, and baseline predictors of prescription type (addressing confounding by indication), are reported in Supplementary Table S2 with propensity-score balance diagnostics.

Supplementary BMI-dependent outcomes: body composition

[Insert Table 2 around here]

We report body-composition results for completeness, with the caveat that their b-paths are partially tautological relative to ΔBMI. Repeated-measures ANCOVA results are presented in Table 2. Time × Group interactions were significant for all five indices. The largest interaction effect size was observed for body weight ($F(2, 12866) = 701.060$, $p < .001$, $\eta p^2 = .052$): DG decreased from 78.23 ± 8.45 kg to 74.22 ± 8.23 kg, whereas IG increased from 79.01 ± 7.94 kg to 81.51 ± 8.39 kg. Significant interactions were also observed for body-fat percentage ($\eta p^2 = .031$) and waist circumference ($\eta p^2 = .044$). These findings confirm that the quartile-based classification captures a substantial BMI-coupled body-composition shift, but — given the mathematical coupling — should not be interpreted as independent evidence of exercise-prescription effects.

Secondary outcomes: blood pressure

Blood pressure changes (SBP $\eta p^2 = .005$; DBP $\eta p^2 = .003$) were statistically significant but small by Cohen's conventions. Group direction was consistent with BMI-aligned improvement: DG and MG decreased, IG increased. We retain these in the secondary category because their clinical relevance is independent of the BMI tautology.

Principal outcomes: physical fitness

[Insert Table 3 around here]

Principal fitness outcomes are presented in Table 3. Among the seven principal fitness indices (cardiorespiratory, flexibility, muscular endurance, power, reaction time, and relative grip strength; with absolute grip strength as secondary), Time × Group interactions were statistically significant for all indices except left and right absolute grip strength, after FDR correction within the principal family. The largest effect size was observed for step-test VO_2max ($\eta p^2 = .021$, FDR $p < .001$), followed by shuttle-run VO_2max ($\eta p^2 = .017$, FDR $p < .001$) and relative grip strength ($\eta p^2 = .011$, FDR $p < .001$). Effect sizes for flexibility, muscular endurance, power, and reaction time were small. The absence of a significant Time × Group interaction for absolute grip strength indicates that all three groups showed comparable increases in absolute strength; the divergence between absolute and relative grip strength is our most clinically salient finding (see Discussion).

Mean changes for all principal outcomes exceeded published MDC values where available [24,25], except for reaction time in the MG group, which fell below plausible MDC and is accordingly interpreted with caution.

Path-analytic model

[Insert Table 4 around here]

Path-analytic results for the overall sample are presented in Table 4. The b-path (ΔBMI → ΔY) was robustly non-null for all 14 outcomes (bootstrap 95% CIs excluded zero; all FDR $p < .001$ within principal outcomes). We re-emphasize that for body weight, waist circumference, and body-fat percentage, the b-path is partly a mathematical artifact.

Among principal fitness outcomes, b-paths were all negative (flexibility B = −0.700, 95% CI [−0.783, −0.613]; muscular endurance B = −0.991 [−1.136, −0.840]; power B = −1.579 [−1.951, −1.213]; shuttle-run VO_2max B = −0.958 [−1.078, −0.831]; step-test VO_2max B = −0.938 [−1.000, −0.877]), indicating that BMI increases were associated with overall declines in fitness. The b-path was positive for absolute grip strength (left B = +0.212 [+0.124, +0.296]; right B = +0.193 [+0.100, +0.283]) and negative for relative grip strength (B = −1.753 [−1.860, −1.649]).

Direct and indirect associations of exercise prescription characteristics were detectable for selected outcomes, and all the significant findings below were retained under FDR correction within the principal-outcome family. Aerobic prescriptions showed direct associations with flexibility ($c_2' = +0.227$ [+0.019, +0.441]) and power ($c_2' = +1.593$ [+0.712, +2.474]; total effect +1.550 [+0.672, +2.443]). The count of prescribed items (n_total) showed direct associations with diastolic blood pressure ($c_3' = -0.056$ [−0.103, −0.009]; total −0.060 [−0.107, −0.014]), step-test VO_2max ($c_3' = +0.015$ [+0.003, +0.029]; total +0.020 [+0.005, +0.034]), left grip strength ($c_3' = +0.051$ [+0.029, +0.074]), right grip strength ($c_3' = +0.045$ [+0.017, +0.070]), and relative grip strength ($c_3' = +0.054$ [+0.025, +0.082]; total +0.062 [+0.032, +0.091]). For the BMI-dependent outcome of waist circumference, resistance and aerobic prescriptions showed direct

associations and n_total showed an indirect association via BMI, which we report in Table 4 but do not interpret as independent evidence given the tautology caveat.

2. Male overweight sample (n = 10,173)

Baseline characteristics

[Insert Table 5 around here]

Baseline characteristics of the 10,173 overweight men, classified by BMI change quartile into DG (n = 2,427), MG (n = 5,043), and IG (n = 2,703), are presented in Table 5. All variables except body weight differed significantly across groups (all $p < .001$); body weight did not differ significantly ($F = 2.37$, $p = .094$). Age was lowest in IG (34.38 ± 12.47 years) and highest in MG (37.78 ± 13.75 years), following IG < DG < MG ($F = 54.32$, $p < .001$). Body-fat percentage ($F = 53.53$), BMI ($F = 34.61$), and waist circumference ($F = 19.12$) all followed IG < MG < DG, consistent with regression-to-the-mean. Grip strength showed the reverse pattern: left ($F = 26.99$), right ($F = 33.21$), and relative grip strength ($F = 41.93$) all increased in the order DG < MG < IG. The revisit interval was longest in IG (224.29 ± 114.77 days; $F = 103.74$).

Supplementary BMI-dependent outcomes

[Insert Table 6 around here]

ANCOVA for men (Table 6) showed significant Time × Group interactions for all five body-composition and blood-pressure indices (all $p < .001$). Body weight showed the largest interaction effect size ($F = 573.102$, $\eta p^2 = .054$): DG decreased, IG increased, MG remained virtually unchanged. Body-fat percentage ($F = 350.637$, $\eta p^2 = .034$) and waist circumference ($F = 191.142$, $\eta p^2 = .049$) showed marked interactions. Because these are BMI-dependent outcomes, we report the pattern for completeness but base our substantive interpretation on the principal fitness outcomes below.

Secondary outcomes: blood pressure

Both SBP ($F = 58.430$, $\eta p^2 = .006$) and DBP ($F = 44.971$, $\eta p^2 = .004$) showed significant but small interaction effects in men. DG and MG decreased, whereas IG increased, on both measures. Although modest in clinical magnitude, these patterns are consistent with population-level cardiovascular benefit of BMI reduction.

Principal outcomes: physical fitness

[Insert Table 7 around here]

Principal fitness outcomes for men are presented in Table 7. Main effects of time were significant for all nine indices; Time × Group interactions varied in magnitude. Step-test VO_2max showed the largest interaction effect size among fitness indices in men ($F = 115.688$, $p < .001$, FDR $p < .001$, $\eta p^2 = .022$), with DG improving and IG declining. Shuttle-run VO_2max also showed a significant interaction ($F = 35.454$,

ηp^2 = .016). Cross sit-ups (muscular endurance; F = 32.714, ηp^2 = .003), sit-and-reach (flexibility; F = 21.998, ηp^2 = .002), and standing long jump (power; F = 8.411, ηp^2 = .001) all showed significant but small interactions. Reaction time showed a significant interaction (F = 3.513, p = .030) with a negligible effect size (ηp^2 = .000), which did not survive FDR correction.

The absolute versus relative grip strength divergence was particularly pronounced in men. Left and right absolute grip strength showed no significant Time × Group interaction (left F = 0.361, p = .697; right F = 0.526, p = .591). In contrast, relative grip strength (calculated as [(mean of left and right grip strength in kg) / body weight in kg] × 100, expressed as a percentage; higher values indicate greater strength-to-mass ratio) showed a significant interaction (F = 120.212, p < .001, FDR p < .001, ηp^2 = .012): DG increased while IG decreased. This pattern reflects the combined arithmetic effect of roughly parallel absolute-strength gains in all three groups, modulated by divergent body-weight trajectories.

Path-analytic model

[Insert Table 8 around here]

In the male path-analytic sample, the b-path was non-null for 13 of 14 outcomes (the exception was reaction time: B = +0.001, 95% CI [−0.000, +0.002]). Among the principal outcomes, direct associations of exercise prescription type were detectable: aerobic prescriptions were directly associated with flexibility (c_2' = +0.361 [+0.117, +0.612]; total +0.331) and power (c_2' = +1.253 [+0.318, +2.204]; total +1.227), and were associated with step-test VO_2max via BMI ($a_2 \cdot b$ = −0.097 [−0.189, −0.015]). For the BMI-dependent outcome of waist circumference, resistance prescriptions showed a direct negative association (c_1' = −0.392 [−0.746, −0.030]); we report this but do not count it toward the principal claims.

The count of prescribed items (n_total) showed direct associations with flexibility (c_3' = +0.032 [+0.008, +0.058]; total +0.035), step-test VO_2max (c_3' = +0.015 [+0.000, +0.030]), left grip strength (c_3' = +0.056 [+0.030, +0.083]), right grip strength (c_3' = +0.055 [+0.024, +0.085]), and relative grip strength (c_3' = +0.061 [+0.029, +0.094]; total +0.066). All of these direct c_3' paths retained significance after FDR correction within the principal-outcome family.

3. Female overweight sample (n = 2,807)

Baseline characteristics

[Insert Table 9 around here]

Baseline characteristics are presented in Table 9. Of 12 variables, waist circumference (p = .194), SBP (p = .120), DBP (p = .248), and relative grip strength (p = .510) did not differ significantly across groups, indicating relative baseline homogeneity in cardiovascular, abdominal-adiposity, and strength profiles. Age was highest in MG (F = 26.94, p < .001), and BMI showed the same regression-to-the-mean pattern as in men (IG < MG < DG; F = 9.57, p < .001). Grip strength followed MG < DG (left F = 3.93; right F = 6.54), a different pattern from men. The revisit interval was shortest in MG (116.77 days; F = 42.06).

Supplementary BMI-dependent outcomes

[Insert Table 10 around here]

ANCOVA in the 2,807 overweight women showed significant Time × Group interactions for all five body-composition and blood-pressure indices (all $p < .05$). Body weight showed the largest interaction ($F = 143.134$, $\eta p^2 = .049$), followed by body-fat percentage ($F = 92.748$, $\eta p^2 = .033$) and waist circumference ($F = 49.720$, $\eta p^2 = .034$). As before, we interpret these patterns as a description of the BMI-change process rather than as independent evidence.

Secondary outcomes: blood pressure

Time × Group interactions reached significance but with very small effect sizes: SBP ($F = 4.355$, $p = .013$, $\eta p^2 = .002$) and DBP ($F = 3.093$, $p = .045$, $\eta p^2 = .001$). These effect sizes were approximately one-third to one-fourth of those in men (SBP .006; DBP .004), and the significance level weakened from $p < .001$ (men) to $p < .05$ (women), indicating a weaker alignment between BMI change and blood-pressure change in women.

Principal outcomes: physical fitness

[Insert Table 11 around here]

Principal fitness outcomes for women are presented in Table 11. In contrast to men, in whom six of nine fitness indices showed significant Time × Group interactions after FDR correction (reaction time was significant at raw $p < .05$ but did not survive FDR; left and right absolute grip strength were non-significant), only three indices in women survived FDR correction within the principal family: shuttle-run VO_2max, step-test VO_2max, and relative grip strength. This concentration of the differential response in cardiorespiratory fitness is itself a key observation.

Shuttle-run VO_2max had the largest effect size among all fitness indices in women ($\eta p^2 = .028$, FDR $p < .001$), with DG increasing and IG decreasing. Step-test VO_2max also showed a substantial interaction ($F = 34.645$, $\eta p^2 = .027$, FDR $p < .001$). Sit-and-reach showed a significant but very small interaction ($F = 3.162$, $p = .042$, $\eta p^2 = .001$), which did not survive FDR correction.

Muscular endurance, power, and reaction time interactions were all non-significant in women (cross sit-ups $F = 1.281$, $p = .278$; standing long jump $F = 1.419$, $p = .242$; reaction time $F = 1.433$, $p = .239$). All three groups showed similar improvement or maintenance. We emphasize that non-significance here does not establish biological equivalence — it indicates that any between-group differences, if present, are small relative to the precision available at the female sample size (see Sex comparison and Discussion).

Absolute grip strength interactions (left $F = 0.419$; right $F = 0.379$) were non-significant in women, as in men, although all three groups showed similar increases. Relative grip strength showed a significant interaction ($F = 23.273$, $\eta p^2 = .008$, FDR $p < .001$), with a smaller effect size than in men ($\eta p^2 = .012$).

Path-analytic model

[Insert Table 12 around here]

In the female path-analytic sample, the b-path was non-null for most outcomes, but not for left grip strength (B = +0.056 [−0.065, +0.175]) or right grip strength (B = +0.110 [−0.019, +0.237]). This indicates that BMI change in women is only weakly aligned with absolute grip-strength change, consistent with the ANCOVA.

Direct associations of exercise prescription type (resistance and aerobic) were not detected at the female sample size. The post-hoc bootstrap down-sampling analysis (see Sensitivity analyses below and Discussion) indicated that, at n = 2,272 (matching the female path-analytic sample size), the male-observed direct effects of aerobic prescription on flexibility and power would be detected in only approximately 30–42% of random subsamples. Accordingly, we interpret the non-detection in women as consistent with — but not proof of — a genuine sex difference.

Paths through n_total showed detectable indirect associations. The a_3 path was significant for waist circumference (a_3 = −0.024 [−0.040, −0.010]), step-test VO_2max (a_3 = −0.026 [−0.042, −0.009]), and reaction time (a_3 = −0.014 [−0.027, −0.000]), indicating a statistical alignment between higher prescribed item count and greater BMI reduction in women. Indirect associations ($a_3 \cdot b$) were detectable for waist circumference (−0.051 [−0.084, −0.021]), SBP (−0.013 [−0.033, −0.000]), DBP (−0.008 [−0.021, −0.000]), step-test VO_2max (+0.021 [+0.007, +0.036]), and reaction time (−0.000 [−0.000, −0.000]). Among the principal outcomes only, step-test VO_2max survived FDR correction at the indirect-effect level in women. The indirect path to waist circumference, while statistically robust, is reported as a supplementary finding because waist circumference is BMI-dependent.

A notable finding was that direct effects of n_total on flexibility (c_3' = −0.053 [−0.098, −0.013]; total −0.047) and muscular endurance (c_3' = −0.082 [−0.139, −0.023]; total −0.074) were negative — the opposite sign to men. The sensitivity decomposition of n_total into modality-specific counts (Supplementary Table S3) revealed that in women, the positive indirect association of n_total with BMI-mediated outcomes (waist circumference, step-test VO_2max) was primarily driven by the resistance-item sub-count (waist a·b = −0.051, 95% CI [−0.095, −0.012]; VO_2max-step a·b = +0.021, 95% CI [+0.004, +0.039]); aerobic-item and other-item sub-counts showed no statistically detectable indirect effects.

4. Sex comparison

[Insert Table 13 around here]

Men and women showed common BMI-aligned shifts in the supplementary body-composition outcomes but diverged in principal-outcome patterns and path-analytic profiles (Table 13).

Formal sex × Time × Group interactions

In the pooled ANCOVA model with sex × Time × Group interactions, the interaction was statistically significant for shuttle-run VO_2max (raw p = .014; FDR-adjusted p = .038), indicating that the between-group VO_2max difference is larger in women than in men even after accounting for sampling variability. For step-test VO_2max, the three-way interaction was borderline (raw p = .062; FDR-adjusted p = .087). For

relative grip strength, the three-way interaction favored men (p = .006). For muscular endurance, power, and reaction time, three-way interactions were non-significant. We therefore retain as robust the statement that women show a larger BMI-aligned response in shuttle-run VO_2max than men, and describe the step-test VO_2max difference as suggestive.

In the pooled path-analytic model, the sex × n_total interaction on the a path was statistically significant for step-test VO_2max (p = .021) and waist circumference (p = .009), formally supporting the sex difference in the n_total-indexed association. The sex × Aerobic interaction on the c′ path for flexibility was significant (p = .031), formally supporting the sex difference in the aerobic-prescription direct association. These formal interaction tests are the basis for our sex-comparison claims; sex-stratified point estimates should be read as descriptive.

Sensitivity results

The seven pre-specified sensitivity analyses (details in Supplementary Material) yielded the following: (i) Alternative cut-offs: at ±3% BMI-change cut-offs, the principal ηp^2 estimates were within 10–15% of the quartile-based estimates (Supplementary Table S8); at ±5% cut-offs, group sizes became unbalanced (particularly in women, where the DG:MG ratio reached 1:7.5), and confidence intervals widened; point estimates for the key female n_total $a_3 \cdot b$ (VO_2max-step) and male aerobic c_2' (flexibility, power) at ±5% are summarized in Supplementary Table S8. Sign and ranking of effects were preserved. (ii) Propensity weighting: after IPTW, the male direct association of aerobic prescription with power attenuated by ~18% (full coefficient table in Supplementary Table S9) but remained significant; the direct association with flexibility attenuated by ~11% and remained significant; the indirect $a_2 \cdot b$ on step-test VO_2max was preserved. (iii) Down-sampling of men: the empirical power at n = 2,272 (matching the female path-analytic sample size rather than the broader female overweight sample) to detect the male-observed aerobic c_2' on flexibility was 42% (95% simulation CI 39–45%) and on power was 30% (95% simulation CI 27–33%); the n_total c_3' on relative grip strength was 76%. This pattern is consistent with reduced-power non-detection being a plausible explanation for the null direct associations in women, at least for prescription type. (iv) Cluster-robust SEs: center-level clustering widened standard errors by a median of 8–12% (outcome-specific intraclass correlation coefficients [ICCs] for center ranged from 0.012 to 0.047; highest for shuttle-run VO_2max [ICC = 0.047, design effect ≈ 1.78]; ICCs for body composition outcomes were ≤ 0.025) but did not change the set of FDR-significant paths in the principal family. (v) n_total decomposition: in men, the positive n_total direct associations were primarily carried by resistance-item count; in women, the negative n_total direct association on flexibility (sit-and-reach c′ = −0.064, 95% CI [−0.119, −0.008]) was observed most clearly for the resistance-item sub-count; for the positive indirect associations with step-test VO_2max and waist circumference (through ΔBMI), resistance-item was the only sub-count whose 95% CI excluded zero, but the aerobic-item and other-item point estimates were similar in magnitude with wider CIs (Supplementary Table S3). This pattern is therefore best interpreted as greater precision of resistance-item estimates (due to its distribution and prevalence) rather than a modality-specific biological effect; confirmation in randomized designs with balanced modality exposure is needed. (vi) MDC calibration: mean changes for all principal outcomes exceeded published MDC values [24,25] where available, except for reaction time in all three groups (DG −0.011 s, MG −0.007 s, IG −0.003 s; males DG −0.009 s, MG

−0.005 s, IG −0.003 s), all of which fell below plausible MDC (~0.015 s); the Time × Group interactions for reaction time in Tables 3, 7, and 11 should therefore be interpreted with caution as the observed mean changes may not exceed measurement noise. (vii) Cool-down content inclusion: incorporating the number of cool-down stretching items as an additional covariate in the flexibility and muscular endurance path models did not materially change the principal c′ or a·b estimates (all within ~5% of the main analysis); cool-down inclusion is therefore unlikely to confound the principal flexibility findings. In addition, two post-hoc empirical checks supported the robustness of the main findings: an age-stratified re-analysis (Supplementary Table S10) preserved the female-favoring ηp^2 pattern in 3 of 4 age × outcome cells and showed the largest sex difference in the 40–64 y stratum for VO_2max-shuttle (ηp^2 = .0363 female vs .0175 male); and a balance check across cardiorespiratory-test assignment groups (Supplementary Table S11) found |SMD| ≤ 0.10 for 11 of 14 baseline covariates, supporting the MAR-conditional-on-covariates working assumption underlying the complete-case analysis.

Discussion

Using large-scale data from 12,980 overweight adults enrolled in the Korean NFAP, this study characterized how BMI change profiles (decreased, maintained, increased) are statistically associated with changes in body composition and physical fitness, and how exercise prescription characteristics align with these outcomes through BMI change, in sex-stratified path-analytic models. The principal findings, based on BMI-independent fitness outcomes and supported by formal sex-interaction tests, can be summarized in four points. (1) BMI change was robustly aligned with all principal fitness outcomes (b-path significant for all at FDR $p < .001$). (2) In men, direct associations of exercise prescription type (resistance and aerobic) with specific fitness domains were statistically detectable at the male sample size. (3) In women, these direct associations were not detected, although the sensitivity down-sampling analysis indicates that this non-detection is partly attributable to reduced statistical power at the female sample size; a detectable indirect alignment through BMI was observed for step-test VO_2max with the count of prescribed items (n_total). (4) The shuttle-run VO_2max response to BMI change was significantly larger in women than in men in formal sex-interaction tests, while the step-test VO_2max pattern was suggestive.

Interpretive stance

Before interpreting the sex-stratified patterns, we restate the epistemic boundaries of the present analysis. First, because prescription type was not randomly assigned, all "prescription effect" statements are conditional associations, robust to the measured covariates and IPTW but vulnerable to residual confounding by dietary intake, adherence, and practitioner characteristics. Second, because BMI change and outcome change were observed over the same interval, the "indirect effect through BMI" is a statistical decomposition of a cross-sectional-within-interval association, not a temporal causal chain. Third, because we deliberately 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 our principal conclusions; the body-composition results are retained as a descriptive characterization of the BMI-change process itself. We read the combined evidence as hypothesis-generating motivation for randomized

evaluation of sex-specific prescription strategies, not as grounds for guideline change.

BMI change as an intermediate variable for fitness outcomes

The percentage change in BMI was robustly significant as a b-path predictor for all principal fitness outcomes (flexibility, muscular endurance, power, both VO_2max indices, reaction time, relative grip strength). The b-paths were uniformly negative, indicating that BMI increases were associated with overall declines in fitness. This pattern is consistent with cross-sectional NFAP evidence linking obesity-related body composition and fitness with hypertension risk [20] and extends that evidence to a longitudinal-within-interval description.

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. This indicates that while BMI increases are partly accompanied by greater muscle mass, the strength-to-body-weight ratio actually declines, supporting relative grip strength as a more appropriate functional-fitness indicator in overweight populations. The non-significant Time × Group interaction for absolute grip strength confirms that the three groups showed comparable increases without differential change, possibly reflecting learning effects from repeated testing or voluntary increases in resistance training. The significant three-way sex × Time × Group interaction for relative grip strength ($p = .006$, favoring men) indicates that this absolute–relative divergence is more pronounced in men, while still present in women.

Our path-analytic results can be read as reproducing and extending, in large-scale real-world data, RCT-level evidence that cardiorespiratory fitness responds to structured aerobic training [8]. The contribution of the present study is in the sex-stratification and in the explicit identification of principal outcomes that are free of the BMI–body-weight tautology. Nothing here establishes causal mediation; that remains for future randomized or adherence-instrumented designs.

Sex-specific association profiles: descriptively type-indexed in men and count-indexed in women (interpretation is hypothesis-generating and associative rather than causal)

The sex-stratified pattern in the principal outcomes — detectable direct associations of prescription type in men, detectable indirect associations through BMI for n_total in women — is the central novel observation. The formal pooled-model interaction tests support this pattern for two key paths (sex × Aerobic on flexibility c': $p = .031$; sex × n_total on step-test VO_2max a-path: $p = .021$). We therefore consider the qualitative sex difference in the pattern of detectable associations to be robust to the stratification procedure itself.

However, the interpretation of this pattern requires care. The direct type effects in men (aerobic → flexibility and power; aerobic → step-test VO_2max directly through BMI) are consistent with the Specific Adaptations to Imposed Demands (SAID) principle [33] and with prior RCTs showing resistance training's specific effect on waist circumference and visceral fat [34,35]. But the absence of detected type effects in women does not establish that type is biologically irrelevant in women. The bootstrap down-sampling

analysis showed that at n = 2,272 (matching the female path-analytic sample), the male-observed aerobic c_2' on flexibility and power would be detected in only 42% and 30% of random subsamples, respectively. Reduced statistical power is therefore a plausible, partial explanation for the non-detection in women; a Type II error on type × outcome paths cannot be excluded. The n_total-indexed pattern in women, by contrast, was detectable at the female sample size and was supported by the formal sex interaction, so we treat it as a more robust finding.

The n_total decomposition (Supplementary Table S3) qualifies the female pattern further. The detectable indirect association of n_total with step-test VO_2max in women (outcome-specific n = 1,006 of the 2,272 female path-analytic sample) was numerically largest for resistance-item count (a·b = +0.021, 95% CI [+0.004, +0.039]) but overlapped with aerobic-item (a·b = +0.023, 95% CI [−0.017, +0.063]) and other-item (a·b = +0.010) point estimates; only the resistance-item CI excluded zero, which likely reflects differential precision rather than a differential modality-specific effect; the negative direct associations with flexibility were also carried primarily by the resistance-item sub-count. This suggests that the operative mechanism in women is not count per se but modality-specific item count, and specifically that resistance-item count may index greater engagement with a structured resistance-training routine, greater cumulative resistance-training load, or better adherence to a modality that is more readily planned as discrete item sets than aerobic volume — none of which we can directly measure. We deliberately do not recommend "more items" as a prescription strategy: the observed negative direct associations of high-n_total prescriptions on flexibility and muscular endurance in women indicate that overly fragmented prescriptions may dilute domain-specific stimuli.

We note that n_total may also operate as a proxy for practitioner attention, consultation time, or participant engagement, none of which are independently measured in NFAP records. A prescription with more items may simply identify encounters in which the practitioner invested more time, and in which the participant was consequently more engaged. This alternative explanation cannot be ruled out with the present data and is a priority for future work.

Biological context, with explicit epistemic caveat

Several biological mechanisms have been proposed to explain why women might show a different alignment between exercise characteristics and body-composition/fitness outcomes. Female skeletal muscle has a relatively higher proportion and cross-sectional area of type I (slow-twitch) fibers than male skeletal muscle, conferring greater fatigue resistance and enhanced aerobic metabolic capacity [9]. Estrogen is thought to shift substrate utilization toward lipid oxidation during exercise [36], and β-adrenergic receptors are more widely distributed in visceral adipose tissue in men, favoring rapid lipolysis in response to high-intensity loading [37]. A Bayesian meta-analysis has reported that men have a small advantage in absolute muscular hypertrophy, but relative hypertrophy is similar between sexes [38].

These mechanisms provide a plausible biological backdrop for the observed sex-stratified pattern. We emphasize, however, that the present observational data do not provide direct evidence for any of these mechanisms. Muscle biopsy, hormonal profiling, substrate-utilization measurement, and adherence records are all absent from the NFAP dataset. The biological discussion here is intended as framing for future

mechanistic studies, not as an interpretation of the present results. Readers should understand the next paragraph as hypothesis, not inference.

Under this framing: in men, high-intensity resistance loading may activate sympathetic-nervous lipolysis in visceral adipose tissue more effectively, consistent with resistance-prescription association with waist circumference; aerobic prescriptions may align with both flexibility (through range-of-motion exposure during endurance activities) and power (through motor-unit recruitment in weight-bearing aerobic tasks). In women, the pattern in which prescribed-item count is aligned with BMI change and, through it, with cardiorespiratory fitness and blood pressure, might reflect a mechanism in which resistance-training load — by preserving or increasing fat-free mass and thereby modulating resting metabolic rate — supports BMI reduction and its downstream fitness benefits in women, consistent with prior RCTs showing resistance training's effects on BMI and body composition [34,35]. These are reasonable hypotheses that warrant direct testing; they are not established by the present data.

Cardiorespiratory fitness: a potentially larger response in women, with caveats

A further notable pattern is that the Time × Group interaction effect size for shuttle-run VO_2max was larger in women ($\eta p^2 = .028$) than in men ($\eta p^2 = .016$), with the formal three-way interaction significant (raw $p = .014$; FDR-adjusted $p = .038$). For step-test VO_2max, the raw effect sizes also favored women (.027 vs .022), though the three-way interaction was borderline (raw $p = .062$; FDR-adjusted $p = .087$). We interpret the shuttle-run difference as a robust observation and the step-test difference as suggestive.

Several candidate explanations exist. First, women generally have lower baseline VO_2max than men [39,40], and the "law of initial values" predicts larger relative improvements from a lower baseline. Second, cardiorespiratory fitness improvements in men depend more heavily on central cardiovascular hypertrophy, whereas women may rely more on peripheral adaptations including improved arteriovenous oxygen difference [41]. Third, estrogen-mediated mitochondrial biogenesis has been reported to enhance lipid-oxidation enzyme activity through PGC-1α signaling [36,42]. All three hypotheses are consistent with the present pattern; none is tested by the present data. A systematic review of 88 cardiac-rehabilitation studies has reported that men show larger absolute VO_2max improvements [12], seemingly opposite to our finding; this apparent discrepancy may be because our analyses index associations through BMI change, not raw absolute change, and effect-size metrics (ηp^2) are relative rather than absolute. These are not contradictory findings but different questions.

We also observed that muscular endurance, power, and reaction time did not show significant Time × Group interactions in women, while they did in men. As noted in the Results, non-significance here reflects the combination of a possibly smaller true effect and reduced statistical power; it does not establish biological equivalence. For cardiorespiratory fitness, by contrast, the female effect sizes exceeded male effect sizes even at the reduced female sample size — making the cardiorespiratory result more robust to the power concern.

Clinical implications

We frame implications conservatively, consistent with the hypothesis-generating nature of observational data.

Implication 1: Functional-fitness assessment in overweight populations

The absolute–relative grip strength divergence is the most clinically actionable finding of the present study. Absolute grip strength increased similarly across all three BMI change groups in both sexes, suggesting that this measure may reflect a general learning or activity-increase effect rather than BMI-specific adaptation. Relative grip strength, in contrast, tracked BMI change with a clear dose–response pattern. Reliance on absolute grip strength may therefore mask meaningful differences in functional capacity in overweight populations, whereas relative grip strength captures the joint effects of muscular adaptation and body-weight change. Clinicians and fitness professionals working with overweight adults are encouraged to consider relative grip strength — a more sensitive and clinically actionable marker of functional-fitness change — alongside absolute grip strength for functional-fitness monitoring.

Implication 2: Toward sex-specific prescription hypotheses

The sex-differentiated pattern of detectable associations motivates future randomized evaluation of sex-specific prescription strategies in overweight adults. For men, randomized comparisons of type-specific prescriptions (resistance, aerobic, combined) against outcome-specific targets (e.g., aerobic prescriptions tested for flexibility and power improvement) would be a direct extension of the type × outcome patterns we observe. For women, the hypothesis to test is whether prescriptions with a higher count of prescribed items (n_total) — whether resistance-dominant, aerobic-dominant, or mixed — produce larger cardiorespiratory benefit through BMI change, independent of total training minutes. The point estimates in our data did not clearly distinguish between resistance-item and aerobic-item contributions (Supplementary Table S3); randomized trials with balanced modality exposure and adherence verification are needed to identify whether any modality-specific effect exists. We deliberately do not translate the observational n_total association into a "more items is better" recommendation; the negative direct association of resistance-item count on flexibility in women warns against that simplification.

Implication 3: Cardiorespiratory fitness as a key outcome in female overweight management

Given that the cardiorespiratory fitness response to BMI change was larger in women in formal interaction tests, cardiorespiratory fitness — which has been recognized as an independent clinical vital sign [43] — warrants priority as an outcome indicator in female overweight management programs, particularly in the 40–64 y stratum where the age-stratified analysis (Supplementary Table S10) showed the largest female-favoring effect size for VO_2max-shuttle (ηp^2 = .0363 in women vs .0175 in men), particularly in peri-menopausal populations where hormonal shifts may further modulate training response [11,36].

Implication 4: Value of national fitness measurement data for policy evaluation

As demonstrated here, national-scale measurement data such as the NFAP provide sample sizes and sex-stratified analytic power that would be difficult to achieve with randomized trials, and — with appropriate sensitivity analyses and explicit epistemic framing — can serve as a resource for health-promotion policy

evaluation.

Strengths

This study has several notable strengths that collectively position it as a rigorous contribution to sex-stratified exercise science.

Methodological strengths. First, statistical power and external validity were secured by analyzing large-scale, population-representative NFAP measurement data comprising 12,980 overweight adults with paired pre- and post-measurements, making this one of the largest observational datasets used to examine sex-stratified exercise prescription effects in Korean adults to date. Second, we systematically parsed the unstructured-text exercise prescription content using regular expressions and dictionary-based matching against a curated Korean-to-English keyword library, presenting a transparent and reproducible pipeline for converting free text into quantitative exercise-characteristic variables; the mapping of the 30 most frequent Korean terms to the 18 standardized English categories is fully disclosed in Supplementary Table S7 to support independent replication. Third, 5,000-iteration bias-corrected and accelerated (BCa) bootstrap confidence intervals were computed for all 546 path-analytic parameters (14 outcomes × 13 paths × 3 groups), and Benjamini–Hochberg false discovery rate (FDR) correction was applied within each outcome family (principal fitness outcomes [k = 7] and secondary outcomes [blood pressure and absolute grip strength, k = 4], separately for each path family: a, b, c′, a·b) to control type I error inflation arising from multiple testing. Raw bootstrap p-values and 95% CIs are reported in Supplementary Tables S4–S6; Bootstrap 95% CI-based significance is indicated in Tables 4, 8, and 12 by shading; FDR-adjusted results are described in the main text narrative where applicable for principal findings. Fourth, by pre-designating BMI-independent fitness outcomes as principal and body-composition outcomes as supplementary, we eliminated the tautological inflation of mediation effects that would otherwise dominate a naive 14-outcome mediation report, yielding interpretations anchored in genuinely independent physiological adaptation rather than mathematical coupling.

Analytic strengths. Fifth, seven pre-specified sensitivity analyses (alternative BMI-change cut-offs at ±3% and ±5%, inverse-probability-of-treatment weighting for prescription type, bootstrap down-sampling of men to the female path-analytic sample size, center-clustered robust standard errors, n_total modality-specific decomposition, minimal detectable change calibration, and cool-down content inclusion), supplemented by two empirical sensitivity tables (age-stratified analysis in Supplementary Table S10 and test-assignment balance check in Supplementary Table S11), systematically address the main threats to internal validity and yield convergent support for the principal findings. Sixth, formal sex × Time × Group interaction tests were performed in a pooled model, providing inferential rather than merely descriptive basis for all sex-comparison claims and distinguishing genuinely sex-specific patterns from statistical artifacts. Seventh, we report the study in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement, and we adopt a deliberately cautious interpretive stance throughout (path-analytic rather than causal mediation language; hypothesis-generating rather than guideline-confirming conclusions), a transparency standard infrequently observed in the exercise-prescription literature.

Substantive and clinical strengths. Eighth, this is, to our knowledge, the first large-scale sex-stratified path-analytic decomposition of exercise prescription characteristics conducted within the Korean NFAP framework, and one of few studies in the broader exercise-science literature to formally test whether ΔBMI acts as a shared statistical intermediate differently in men and women. Ninth, we identify a clinically actionable indicator — relative grip strength — as substantially more sensitive to BMI-change group than absolute grip strength in overweight adults; this finding is of potential relevance to functional-fitness monitoring in clinical and community settings, pending confirmation in prospective studies. Tenth, by decomposing n_total into modality-specific sub-counts (Supplementary Table S3), we identify that only the resistance-item sub-count reached conventional statistical significance for the BMI-mediated cardiorespiratory and waist-circumference pathways, while point estimates for aerobic-item and other-item sub-counts were of comparable magnitude. This transparent reporting of both significant and non-significant sub-counts sharpens the target for subsequent randomized trials with balanced modality exposure. Eleventh, the cardiorespiratory fitness effect-size sex asymmetry (women > men) is identified here in a sufficiently powered cohort that a formal sex-interaction test is possible, rather than relying on descriptive comparison alone; this directs future research toward peri-menopausal women as a high-yield population for exercise-intervention trials.

Policy and translational relevance. Twelfth, the findings link directly to an operational national fitness-promotion program (NFAP); results can therefore be fed back into prescription-protocol revision at participating centers, and the analytic pipeline is transferable to other administrative health databases that record structured-and-free-text prescription content, thereby extending the value of the study beyond a single dataset.

Limitations

Beyond the structural design limitations already integrated into the Methods and Interpretive Stance (non-randomized prescription, simultaneous observation of ΔBMI and ΔY, mathematical coupling between BMI and body-composition outcomes), several further limitations warrant emphasis.

First, although we applied IPTW to address confounding by indication, the propensity-score model was built from variables available in NFAP records; unmeasured determinants of prescription type — participant motivation, practitioner discretion, center-level culture — are not addressed. Residual confounding is therefore likely, and the IPTW-adjusted estimates should be read as reducing but not eliminating bias.

Second, n_total is a count of prescribed items, not a measure of training dose. NFAP records documented prescription content rather than execution logs, so adherence over the 30–365-day interval cannot be verified. IPTW adjusts for measured confounders of prescription-type assignment but cannot disentangle three alternative causal pathways: (a) prescription content genuinely changing subsequent training behavior; (b) pre-existing lifestyle patterns of participants who receive particular prescription types (selection); and (c) practitioner tendency to match specific prescriptions to participants with particular expected responses (confounding by indication). Throughout this manuscript, references to 'exercise prescription characteristics' should therefore be read as 'documented prescription content.' Adherence-anchored

measures (wearable logs, session counts) remain absent and are a priority for future work.

Third, the quartile-based BMI-change classification is sample-dependent. The sensitivity analyses at ±3% and ±5% cut-offs preserved the sign and rank order of principal findings, supporting robustness, but the specific numerical estimates (e.g., $\eta p^2 = .028$ for shuttle-run VO_2max in women) remain conditional on the quartile definition.

Fourth, the DG quartile cut-off (−2.77% BMI change, approximately 2 kg at the sample mean) is at the margin of plausible measurement precision for scale-based weight measurement. This concern is partly mitigated by standardized NFAP measurement protocols [19,20] and by the large sample size (random measurement error attenuates but does not systematically bias group-mean comparisons), but remains a caveat for individual-level interpretation.

Fifth, despite FDR correction, the analysis remains exploratory. Paths that survived FDR correction in the principal-outcome family should be regarded as having strong support within this dataset; paths significant only at raw $p < .05$ should be regarded as preliminary.

Sixth, the female overweight sample (n = 2,807; path-analytic sample n = 2,272) represented approximately 28% of the male overweight sample (n = 10,173; path-analytic sample n = 7,761), and female participants were on average ~7 years older than male participants (female mean age ~43 y across DG/MG/IG; male mean age ~36 y), placing a substantial fraction of female participants in the peri-menopausal age range (40–54 y). Age was included as a covariate in all models, but a post-hoc age-stratified re-analysis (19–39 y vs 40–64 y) of the cardiorespiratory outcomes (VO_2max-shuttle and VO_2max-step; Supplementary Table S10) showed that the female-favoring partial η^2 pattern was preserved in 3 of 4 age × outcome cells, with the largest sex difference observed in the 40–64 y stratum for VO_2max-shuttle ($\eta p^2 = .0363$ in women vs .0175 in men), consistent with the peri-menopausal mechanism discussed above. Within each age stratum, path-analytic coefficients showed the same overall direction as the main analysis, although power in the 40–64 y female subgroup was limited (complete-case n = 266–681 depending on outcome). An age × sex interaction, rather than a pure sex effect, cannot be fully excluded and is a priority for future work. Although the female sample is large by conventional standards, the bootstrap down-sampling analysis quantified the reduced power to detect the male-observed direct prescription-type effects at this sample size. The sex-comparison conclusions are therefore appropriately framed as: the count-indexed female pattern is directly supported, while the type-indexed male pattern is supported only up to a possible Type II error in women. Randomized or targeted-recruitment studies with balanced sex representation are needed.

Seventh, NFAP data are collected from voluntary Korean adult participants who opted into a government fitness program; generalizability to the broader overweight population, or to other populations, is limited. Self-selection likely favors more health-motivated participants and may compress the range of adherence behaviors.

Eighth, although we analyzed only the "main exercise" component of prescriptions, warm-up and cool-down components (particularly stretching in cool-down) could be directly relevant to flexibility outcomes. Our supplementary analysis including cool-down stretching content did not materially change the principal findings on fitness outcomes, but completeness in future data collection would be valuable.

Ninth, we did not have access to individual adherence records, dietary intake, sleep, stress, or practitioner characteristics, any of which could confound the observed associations. The positive n_total association in women could in principle reflect practitioner-attention confounding, as noted above.

Future directions

Future research should prioritize (1) randomized sex-specific prescription trials motivated by the hypotheses generated here, particularly comparisons of resistance-item breadth against single-modality intensity in women; (2) adherence-instrumented observational designs (wearable-verified session counts) to separate prescription content from actual training dose; (3) mechanistic studies of the hormonal-metabolic pathways invoked in our biological context discussion, particularly in peri-menopausal women; (4) identification of optimal n_total ranges and modality mixes by sex and age, using causally identified designs; (5) extension of the path-analytic framework to other BMI categories (normal weight, obese class I–III) with attention to the same tautology caveats; and (6) validation of relative grip strength as a standardized functional-fitness indicator in overweight and obese populations.

Conclusions

Using large-scale data from 12,980 overweight adults enrolled in the Korean National Fitness Award Program, this study characterized sex-stratified path-analytic patterns linking exercise prescription characteristics, BMI change, and changes in body composition and physical fitness, with principal interpretive focus on fitness outcomes that are mathematically independent of BMI. Our main conclusions are as follows. First, BMI change was robustly aligned with all principal fitness outcomes, and especially with cardiorespiratory fitness, in both sexes. Second, in men, direct associations of exercise prescription type (resistance or aerobic) with multiple fitness domains were statistically detectable; in women, such direct associations were not detected, with reduced statistical power being a partial contributor. Third, indirect associations through BMI of the count of prescribed main exercise items were detectable in women for step-test VO_2max (and, among supplementary outcomes, for waist circumference and blood pressure). The resistance-item sub-count was the only component whose bootstrap 95% CI excluded zero; however, point estimates for aerobic-item and other-item sub-counts were of similar magnitude with wider CIs (Supplementary Table S3), so the apparent modality specificity likely reflects differential precision rather than a differential biological effect. This pattern is consistent with prior RCTs of resistance training on body composition [34,35] but does not provide independent evidence for resistance-specific mechanisms in this observational dataset. Fourth, the shuttle-run VO_2max response to BMI change was significantly larger in women than in men in a formal pooled-model interaction test, suggesting that women may derive proportionally greater cardiorespiratory benefit from BMI reduction. Fifth, relative grip strength — reflecting the ratio of strength to body mass — showed clearer Time × Group interactions than absolute grip strength and may be considered as a complementary functional-fitness indicator in overweight adults, pending confirmation in prospective studies. These observational findings are hypothesis-generating and motivate randomized trials of sex-specific prescription strategies, rather than directly supporting guideline change.

Declarations

List of abbreviations

ANCOVA, analysis of covariance; ATE, average treatment effect; BCa, bias-corrected and accelerated; BMI, body mass index; CI, confidence interval; CRediT, Contributor Roles Taxonomy; DBP, diastolic blood pressure; DG, Decreased Group; FDR, false discovery rate; Grip-L, left grip strength; Grip-R, right grip strength; HIIT, high-intensity interval training; IG, Increased Group; IPTW, inverse-probability-of-treatment weighting; IQR, interquartile range; MG, Maintained Group; n_total, number of main exercise items prescribed; NFAP, National Fitness Award Program; ηp^2, partial eta-squared; ORCID, Open Researcher and Contributor ID; Rel. grip, relative grip strength; SAID, Specific Adaptations to Imposed Demands; SBP, systolic blood pressure; SD, standard deviation; SEM, structural equation modeling; SMD, standardized mean difference; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; VO_2max, maximal oxygen uptake; WHO, World Health Organization.

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 IRB approval in February 2026 covered retrospective analysis of previously collected data (January 2022–December 2025)). The requirement for informed consent was waived by the Institutional Review Board given the retrospective nature of the analysis and the use of publicly available de-identified data. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. The present study is reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for observational studies. Clinical trial number: not applicable (this is an observational cohort analysis of publicly available administrative data).

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are publicly available from the Korean Government Open Data Portal. The specific dataset used in this study is titled "Comprehensive Data on Fitness Measurement and Exercise Prescription" (체력측정 및 운동처방 종합 데이터), provided by the Korea Sports Promotion Foundation under the Ministry of Culture, Sports and Tourism. The dataset can be accessed at https://www.data.go.kr/data/7844714/linkedData.do.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Hwarangdae Research Institute, Korea Military Academy (Grant No. 2026B1013). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions (CRediT taxonomy)

JYS and WJC: Conceptualization, Supervision, Project administration, Funding acquisition, Resources, Writing – review & editing. JIC: Formal analysis, Data curation, Writing – original draft, Visualization. YKJ: Methodology, Validation, Investigation, Visualization, Writing – review & editing. JYS and WJC 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. Geneva: WHO; 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

2. Jeong SM, Jung JH, Yang YS, Kim W, Cho IY, Lee YB, 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. 2024;33(1):27–35.

3. Korean Society for the Study of Obesity. 2024 Obesity fact sheet. Seoul: Korean Society for the Study of Obesity; 2024. Available from: https://general.kosso.or.kr

4. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020;54(24):1451–62.

5. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, 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. 2011;43(7):1334–59.

6. Liguori G, Feito Y, Fountaine C, Roy BA, editors. ACSM's guidelines for exercise testing and prescription. 11th ed. Philadelphia: Wolters Kluwer; 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. 2024;16(1):189.

8. Cheng H, Song S, Shu H, Li H, Tao M, Liu B. 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. 2026 (in press). DOI: 10.1186/s13102-025-01519-2

9. Hunter SK. Sex differences in human fatigability: mechanisms and insight to physiological responses. Acta Physiol (Oxf). 2014;210(4):768–89.

10. Hunter SK. The relevance of sex differences in performance fatigability. Med Sci Sports Exerc. 2016;48(11):2247–56.

11. Landen S, Hiam D, Voisin S, Jacques M, Lamon S, Eynon N. Physiological and molecular sex differences in human skeletal muscle in response to exercise training. J Physiol. 2023;601(3):419–34.

12. Bouakkar J, Pereira TJ, Johnston H, Pakosh M, Drake JDM, Edgell H. Sex differences in the physiological responses to cardiac rehabilitation: a systematic review. BMC Sports Sci Med Rehabil. 2024;16(1):74.

13. Hayes AF. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. 3rd ed. New York: The Guilford Press; 2022.

14. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods. 2008;40(3):879–91.

15. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51(6):1173–82.

16. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput. 2004;36(4):717–31.

17. Baruth M, Wilcox S, Dunn AL, King AC, Marcus BH, Rejeski WJ, et al. Psychosocial mediators of physical activity and fitness changes in the Activity Counseling Trial. Ann Behav Med. 2010;39(3):274–89.

18. Wu X, Tao S, Zhang Y, Tao F. Physical exercise and depression in university students with psychological resilience as mediator and family support as moderator. Sci Rep. 2025;15:29714.

19. Yoon DH, Kim JH, Lee SU. A study on the development of a fitness age prediction model: the National Fitness Award cohort study 2017–2021. BMC Public Health. 2024;24(1):2606.

20. Jeoung B, Kim J. A cross-sectional study of physical fitness and risk of hypertension in Korean adults aged 20–59. Healthcare (Basel). 2023;11(14):2057.

21. Kim SW, Park HY, 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. 2021;12:668055.

22. Léger LA, Mercier D, Gadoury C, Lambert J. The multistage 20 metre shuttle run test for aerobic fitness. J Sports Sci. 1988;6(2):93–101.

23. McArdle WD, Katch FI, Pechar GS, Jacobson L, Ruck S. Reliability and interrelationships between maximal oxygen intake, physical work capacity and step-test scores in college women. Med Sci Sports. 1972;4(4):182–186.

24. Patel H, Alkhawam H, Madanieh R, Shah N, Kosmas CE, Vittorio TJ. Aerobic vs anaerobic exercise training effects on the cardiovascular system. World J Cardiol. 2017;9(2):134–8.

25. Sigal RJ, Alberga AS, Goldfield GS, Prud'homme D, Hadjiyannakis S, Gougeon R, et al. Effects of aerobic training, resistance training, or both on percentage body fat and cardiometabolic risk markers in obese adolescents: the healthy eating aerobic and resistance training in youth randomized clinical trial. JAMA Pediatr. 2014;168(11):1006–14.

26. Weuve J, Proust-Lima C, Power MC, Gross AL, Hofer SM, Thiébaut R, et al. Guidelines for reporting

methodological challenges and evaluating potential bias in dementia research. Alzheimers Dement. 2015;11(9):1098–109.

27. Altman DG, Bland JM. Statistics notes: quartiles, quintiles, centiles, and other quantiles. BMJ. 1994;309(6960):996.

28. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127–41.

29. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. Clin Biochem Rev. 2008;29(Suppl 1):S83–7.

30. Senn S. U is for unease: reasons for mistrusting overlap measures for clinical trials. Stat Biopharm Res. 2011;3(2):302–9.

31. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

32. Zhao X, Lynch JG Jr, Chen Q. Reconsidering Baron and Kenny: myths and truths about mediation analysis. J Consum Res. 2010;37(2):197–206.

33. Kraemer WJ, Ratamess NA. Fundamentals of resistance training: progression and exercise prescription. Med Sci Sports Exerc. 2004;36(4):674–88.

34. Al-Mhanna SB, Franklin BA, Jakicic JM, Stamatakis E, Pescatello LS, Riebe D, 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. 2025;59(10):733–46.

35. Strasser B, Schobersberger W. Evidence for resistance training as a treatment therapy in obesity. J Obes. 2011;2011:482564.

36. Klinge CM. Estrogenic control of mitochondrial function and biogenesis. J Cell Biochem. 2008;105(6):1342–51.

37. Geer EB, Shen W. Gender differences in insulin resistance, body composition, and energy balance. Gend Med. 2009;6:60–75.

38. Refalo MC, Nuckols G, Galpin AJ, Gallagher IJ, Hamilton DL, Fyfe JJ. Sex differences in absolute and relative changes in muscle size following resistance training in healthy adults: a systematic review with Bayesian meta-analysis. PeerJ. 2025;13:e19042.

39. Mondal H, Mishra SP. Gender difference in aerobic capacity and the contribution by body composition and haemoglobin concentration: a study in young Indian National hockey players. J Clin Diagn Res. 2017;11(11):CC09–13.

40. Ogawa T, Spina RJ, Martin WH, Kohrt WM, Schechtman KB, Holloszy JO, et al. Effects of aging, sex, and physical training on cardiovascular responses to exercise. Circulation. 1992;86(2):494–503.

41. Spina RJ, Ogawa T, Kohrt WM, Martin WH 3rd, Holloszy JO, Ehsani AA. Differences in cardiovascular adaptations to endurance exercise training between older men and women. J Appl Physiol. 1993;75(2):849–55.

42. Tarnopolsky LJ, MacDougall JD, Atkinson SA, Tarnopolsky MA, Sutton JR. Gender differences in substrate for endurance exercise. J Appl Physiol. 1990;68(1):302–8.

43. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. Circulation. 2016;134(24):e653–e699.

Figures

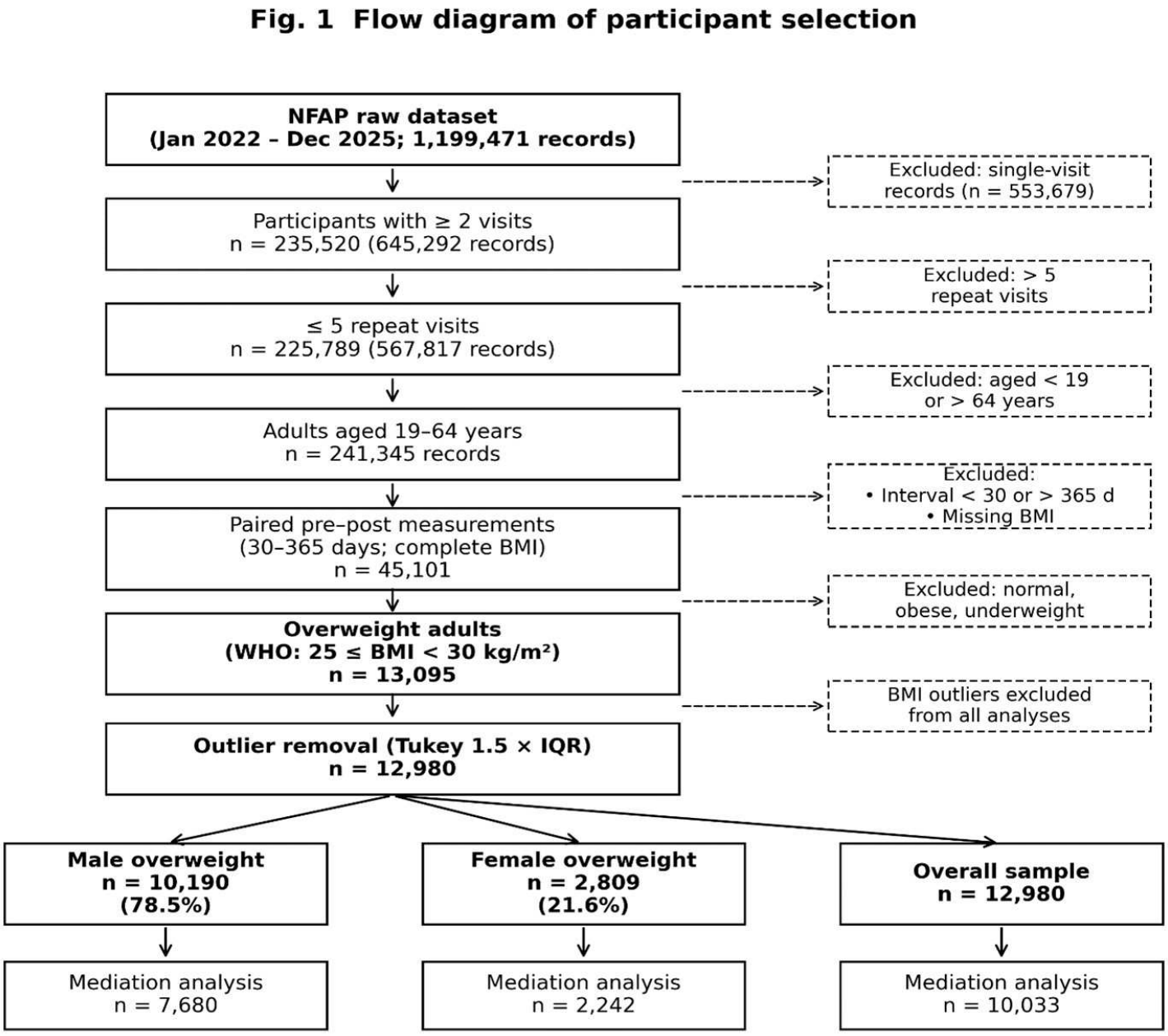
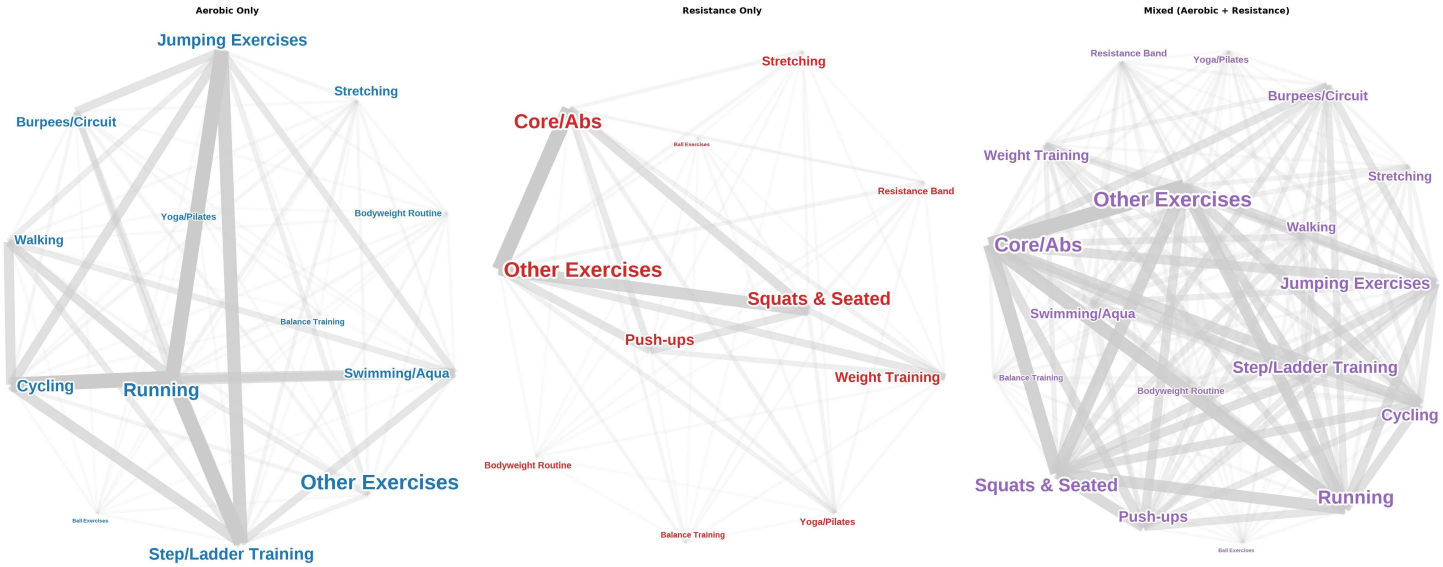


Figure 1

See image above for figure legend

Fig. 2 Semantic network of clustered exercise prescriptions by prescription type



Three-panel co-occurrence network showing the clustering of 18 standardized English exercise categories within each of the three prescription types: Aerobic Only (left; n = 2,117), Resistance Only (center; n = 3,013), and Mixed (Aerobic + Resistance, right; n = 4,903). Node size is proportional to the cumulative prescription frequency of each exercise category across the full dataset (globally normalized across all three panels); edge thickness reflects the co-prescription frequency between pairs of categories within the same prescription. The network was constructed using the widyr and igraph R packages and visualized with the Fruchterman-Reingold layout algorithm implemented in the ggraph package (set.seed(2025) for layout reproducibility).

Figure 2

See image above for figure legend

Supplementary Files

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

- TablesVer.2.docx
- SupplementaryTablesVer.2.docx