

Sex-Specific Performance Patterns in IFBB Fitness Challenge: A Multi-Year Analysis of Competitive Output Under Standardized Loading Conditions

Cristian Băltărețu¹, Bogdan Costache², Adrian Iacobini³, Virgil Tudor¹

¹ National University of Physical Education and Sport, Bucharest, Romania

² Bucharest University of Economic Studies (ASE), Bucharest, Romania

³ CSU ASE Bucharest, Romania

Abstract

Competitive Fitness Challenge uses sex-specific prescribed loads within a standardized six-station circuit, creating an unusual model for examining whether competition design attenuates sex differences in repetition-based performance. The purpose of this study was to characterize sex-specific performance patterns across multiple international competition seasons and to determine whether male-female output relationships differed between competitive standards and exercise stations. A retrospective repeated cross-sectional analysis was conducted using official International Federation of Bodybuilding and Fitness competition results from 2021 to 2026. The primary longitudinal analysis examined senior Gold and Silver winning scores, while detailed station-level analyses used complete recent European Championship datasets. Female-to-male performance ratios were calculated for total scores and individual stations. Across the observation period, total-score sex differences narrowed substantially compared with the inaugural season, particularly at Silver level. In recent competitions, mean total output was nearly equivalent between men and women, despite clear station-specific differences. Upper-body bodyweight pulling and pressing stations generally retained a male advantage, whereas several externally loaded lower-body and whole-body stations produced female repetition outputs comparable with or greater than male outputs under the sex-specific load prescriptions. These findings indicate that total-score parity can coexist with distinct sex-specific performance profiles at the station level. The study suggests that standardized sex-specific loading can attenuate differences in aggregate competitive output without eliminating task-specific physiological and mechanical differences. The results provide an initial evidence base for future load calibration, athlete profiling, competition design, and sex-specific training research in Fitness Challenge.

Keywords: Fitness Challenge; sex differences; competitive performance; load scaling; functional fitness.

1. Introduction

Sex differences in athletic performance are well established, but their magnitude is not constant across sports or tasks. Adult men typically demonstrate greater absolute muscular strength, power, speed, and maximal aerobic capacity than women of comparable age and training status, whereas women may display lower fatigability during selected submaximal tasks. The resulting performance gap is therefore strongly dependent on the relative contribution of strength, power, endurance, body mass, movement economy, and fatigue resistance to the competitive task (Hunter et al., 2023; Hunter and Senefeld, 2024).

This task dependence has direct implications for sports that use sex-specific competition prescriptions. In weight-class, functional-fitness, and hybrid events, women and men may perform the same movement pattern while using different external loads, implement heights, or repetition requirements. Such scaling is intended to create an appropriate competitive challenge for both divisions, yet equivalent aggregate scores do not necessarily indicate equivalent physiological demands. Instead, similar outcomes may emerge because the competition structure compensates for expected differences in absolute strength and power while leaving other characteristics, such as local muscular endurance and technical efficiency, less strongly differentiated.

Fitness Challenge, governed internationally by the International Federation of Bodybuilding and Fitness, provides a particularly useful model for studying this issue. The discipline is based on a six-station circuit in which athletes seek to complete the maximum number of technically valid repetitions during two-minute work periods, separated by standardized transition periods. Competitive standards include Bronze, Silver, and Gold, with exercise difficulty and external resistance prescribed by level and sex (International Federation of Bodybuilding and Fitness, 2024a; International Federation of Bodybuilding and Fitness, 2025b). Current rules describe the discipline as a performance-based sport focused on strength-exercise stations and repetition output rather than physique assessment.

The structure is scientifically interesting because men and women compete in separate divisions under standardized but sex-specific load prescriptions. For example, recent Gold competition has included identical movement patterns for both sexes but different absolute external loads on selected stations, whereas several bodyweight stations remain mechanically similar in format. This creates an opportunity to examine whether sex-specific loading produces comparable total competitive output and, simultaneously, whether task-specific sex differences remain visible at individual stations.

Research in CrossFit and high-intensity functional training provides relevant comparative evidence. Schlegel and Křehký (2022) reported that men outperformed women in most events of the 2021 CrossFit Games despite sex-specific scaling, with the magnitude of the gap varying considerably according to event characteristics. Mangine, Grundlingh and Feito (2023a) similarly demonstrated significant sex differences across most CrossFit Open workouts, while their longitudinal analysis of repeated workouts showed that men and women did not improve identically over time (Mangine, Grundlingh and Feito, 2023b). More recent work indicates that pacing, movement type, relative resistance, and competition experience influence functional-fitness performance patterns (Mangine et al., 2024).

However, these findings cannot be transferred directly to Fitness Challenge. Unlike competition formats characterized by changing or undisclosed workouts, Fitness Challenge is organized around a constrained station structure, official judging standards, and standardized work intervals. This relative standardization provides a stronger basis for comparing male and female repetition outputs over time, while the progressive evolution of exercise selection and regulation requires careful interpretation.

The purpose of the present study was therefore to characterize sex-specific performance patterns in international Fitness Challenge competition across the 2021-2026 period and to examine the extent to which standardized sex-specific loading attenuates differences in aggregate competitive output. A secondary purpose was to determine whether station-level sex patterns differ according to movement type and competitive standard.

2. Literature Review

2.1 Biological basis of sex differences in performance

Biological sex influences athletic performance through interacting anatomical, endocrine, cardiovascular, hematological, and neuromuscular mechanisms. The American College of Sports Medicine consensus statement identifies post-pubertal differences in muscle mass, muscle cross-sectional area, hemoglobin concentration, cardiac size, maximal oxygen transport, and androgen exposure as major contributors to the male advantage observed in strength-, power-, speed-, and endurance-dependent performance (Hunter et al., 2023). Complementary endocrine and performance syntheses emphasize the central role of androgen-mediated changes in muscle and oxygen-transport capacity and the persistence of sport-specific performance gaps after puberty (Handelsman et al., 2018; Joyner et al., 2025). Hunter and Senefeld (2024) similarly concluded that the magnitude of the sex difference is performance-domain specific rather than uniform.

Absolute force and power differences are especially relevant to tasks involving substantial external resistance. In highly trained functional-fitness athletes, D'Hulst et al. (2024) reported clear male-female differences in absolute cardiorespiratory, critical-power, and force measures, even within a highly selected competitive population. Ponce-García et al. (2025) likewise identified significant sex differences in anaerobic power among CrossFit athletes, although adjustment for lean or muscle mass reduced some of these differences. Such findings reinforce the distinction between absolute performance capacity and performance expressed relative to body or muscle mass.

Training status does not eliminate sex differences, but adaptation to training can be similar in relative terms. Roberts, Nuckols and Krieger (2020) found broadly similar relative hypertrophic and lower-body strength adaptations between men and women following resistance training, while women showed a larger relative effect for upper-body strength improvement. These findings are important when interpreting multi-year competition results: the persistence of sex differences in absolute capacity does not imply that women have a lower capacity for training adaptation.

2.2 Sex differences in fatigability and repeated-effort performance

Fitness Challenge is not a single maximal-effort test. Athletes repeatedly perform submaximal or near-maximal contractions for two-minute intervals and must then reproduce performance at subsequent stations. Sex differences in fatigability are therefore potentially relevant to competition outcomes.

Women are often less fatigable than men during sustained contractions performed at the same relative intensity, although the magnitude and even direction of the difference depend on contraction mode, muscle group, velocity, perfusion, and absolute force requirements (Hunter, 2014, 2016). Earlier experimental work demonstrated that lower absolute force demands can contribute substantially to longer endurance times in women during submaximal contractions (Hunter and Enoka, 2001). Consequently, a competition format that reduces women's external loads may do more than simply account for maximal-strength differences: it can alter the balance between force requirement and fatigue resistance throughout a repeated-repetition task.

This interaction is likely movement specific. Bodyweight pulling and pressing tasks may preserve larger between-sex differences because the external challenge cannot be scaled independently of the athlete's body mass and relative upper-body strength. In contrast, externally loaded lower-body or whole-body exercises can be adjusted directly through sex-specific resistance prescriptions. A central expectation of the present study was therefore that sex ratios would differ by station type rather than remain constant across the circuit.

2.3 Sex-specific scaling in functional fitness competition

Functional fitness competitions frequently use sex-specific scaling of external loads. Schlegel and Křehký (2022) noted that CrossFit competition typically prescribes women's weightlifting loads at a fraction of men's loads while often retaining identical or near-identical gymnastics and metabolic tasks. Their analysis of the 2021 CrossFit Games found that performance differences remained in most events, demonstrating that scaling can reduce but does not universally remove sex-based differences.

Mangine, Grundlingh and Feito (2023a) extended this work by creating normative scores from more than a decade of CrossFit Open competition data. Significant sex differences were present in most workouts, but the magnitude varied with scoring type, workout duration, movement selection, and resistance characteristics. In repeated workouts, men and women also demonstrated different improvement trajectories (Mangine, Grundlingh and Feito, 2023b). Together, these findings suggest that competition scaling should be evaluated empirically rather than assumed to produce equivalent challenge.

The predictive importance of pacing and relative resistance further complicates interpretation. Mangine et al. (2024) found that repetition pace, break strategy, and consistency were central predictors of CrossFit Open performance, with their influence modified by movement complexity and relative load. These factors are directly relevant to Fitness Challenge because valid repetition accumulation within fixed work windows is the primary scoring mechanism.

The broader functional-fitness literature supports a multidimensional interpretation of performance. Systematic reviews characterize high-intensity functional training as a multimodal model in which strength, power, aerobic fitness, and muscular endurance interact rather than operate as isolated qualities (Feito et al., 2018; Claudino et al., 2018; Wang et al., 2023). This is important for sex comparisons because the magnitude of a performance gap will depend on which of these qualities is most strongly constrained by a specific station.

Competition-focused studies further show that aerobic capacity, anaerobic power, maximal and relative strength, body composition, technical skill, and sport-specific experience can each contribute to functional-fitness performance, with the importance of individual predictors varying by task structure (Bellar et al., 2015; Butcher et al., 2015; Serafini et al., 2018; Martínez-Gómez et al., 2019, 2020; Mangine et al., 2020; Meier, 2021; Meier et al., 2023). The importance of muscular strength as a general athletic foundation is also well established (Suchomel et al., 2016), while elite athlete profiling confirms substantial sex-specific differences in several absolute physiological capacities even among highly trained functional-fitness competitors (Sauvé et al., 2024).

Research on newer standardized hybrid competition provides an additional methodological analogue. Brandt et al. (2025) demonstrated that physiological profiling can identify meaningful performance correlates in an emerging competition format, although the running-dominant structure of HYROX differs substantially from Fitness Challenge. Taken together, these studies justify evaluating Fitness Challenge at both aggregate and station levels rather than assuming that a single sex-scaling ratio or a single physiological characteristic can explain total performance.

2.4 Fitness Challenge as a standardized comparative model

The current International Federation of Bodybuilding and Fitness rules define Fitness Challenge as a six-station circuit in which athletes perform the maximum possible number of valid repetitions during two-minute work periods. Judges control technical execution and record valid repetitions, and external loads are prescribed by sex and competitive standard (International Federation of Bodybuilding and Fitness, 2024a; International Federation of Bodybuilding and Fitness, 2025b).

Recent Gold competition includes strict pull-ups, a kettlebell squat-and-pull, dips, loaded walking lunges, toes-to-bar, and devil press. Silver includes lying pull-ups, a loaded lifting station, bench dips, dumbbell lunges, weighted sit-ups, and dumbbell snatch. The sex-specific load prescription is particularly relevant because some stations are externally scalable whereas others are performed with body mass. This mixed structure allows the competition to be viewed as a natural experiment in task-specific sex performance under a standardized scoring framework.

At the same time, the sport has evolved since its international introduction in 2021. Changes in station configuration and technical regulation mean that raw scores from different years cannot be interpreted as if they came from an invariant laboratory test. Longitudinal sex comparisons must therefore emphasize relative male-female output within each season and regulatory configuration rather than simply comparing absolute scores across all years.

2.5 Research gap, objectives, and hypotheses

The existing literature establishes that sex differences in functional-fitness performance are task dependent and influenced by scaling, relative load, fatigue, and pacing. However, no peer-reviewed study has yet characterized sex-specific competition output in IFBB Fitness Challenge across multiple international seasons or examined how aggregate parity relates to station-specific differences.

The study addressed four research questions: (RQ1) How has the female-to-male competitive output ratio changed across international Fitness Challenge seasons? (RQ2) Does the magnitude of the sex difference differ between Gold and Silver competition? (RQ3) Which station types show persistent male advantages, approximate parity, or higher female repetition output under sex-specific loading? (RQ4) Are recent station-level patterns sufficiently consistent across seasons to suggest systematic effects of competition design rather than isolated event-specific variation?

The analysis was guided by three exploratory hypotheses. First, total-score sex differences would be smaller in recent seasons than in the inaugural competition period. Second, externally loaded stations with sex-specific resistance prescriptions would show smaller male-female gaps than upper-body bodyweight stations. Third, Silver competition would show greater aggregate male-female convergence than Gold because its movement complexity and absolute force requirements are lower.

3. Methodology

3.1 Study design

A retrospective, repeated cross-sectional competition-analysis design was used. The analytical unit was an officially recorded athlete performance. No primary physiological testing or intervention was conducted. The study combined a multi-year benchmark analysis (2021-2026) with a detailed station-level analysis of recent European Championship datasets for which complete individual results were available.

3.2 Data sources and study period

Competition data were derived from official International Federation of Bodybuilding and Fitness results and from the authors' harmonized competition database compiled from those records. The multi-year benchmark series included senior individual Gold and Silver winning performances from the 2021-2026 international competition period, drawing on the inaugural 2021 World Championships, the separate 2022 women's and men's world-level Fitness Challenge events, the 2023 World Championships, and subsequent European reference competitions (International Federation of Bodybuilding and Fitness, 2021; 2022a; 2022b; 2023; 2024b; 2025a; 2026). Detailed station-level analyses used the 2025 and 2026 European Fitness, Bodybuilding & Fitness Challenge Championships, because complete male and female station scores were available and the competitions represented consecutive high-level European events. The 2026 season was still ongoing at the analytical cut-off; consequently, 2026 values represent an early-season benchmark rather than a complete annual summary.

3.3 Eligibility and data-quality procedures

The principal analysis included adult/senior individual Gold and Silver categories in which both male and female results were available. Junior, Masters, Paired, Team, Bronze, Children's Fitness Challenge, and time-to-completion SpeedFit formats were not pooled into the primary sex-comparison dataset because their age structure, task configuration, or scoring systems differed. Performances recorded as complete zero rows, disqualifications, or non-scoring entries were excluded from descriptive calculations. Station scores were cross-checked against published totals whenever possible. Missing station values were not imputed; the affected observation was retained for total-score analyses only when an official total was available.

3.4 Exposure: standardized sex-specific loading

The phrase standardized loading conditions refers to standardized competition prescriptions within each sex division, not identical absolute loading between men and women. External loads are deliberately sex differentiated. In the current Gold standard, for example, walking lunges are prescribed at 50 kg for men and 30 kg for women, while other external-load stations likewise use sex-specific resistance. Bodyweight stations, by contrast, rely on the athlete's own mass and cannot be scaled through an externally prescribed load in the same way. This distinction formed the basis for station-level interpretation.

3.5 Outcome variables

Primary outcomes were total valid repetitions and individual-station valid repetitions. The central comparative metric was the female-to-male performance ratio, calculated as female output divided by male output and multiplied by 100. A ratio of 100% indicates equal repetition output under the prescribed sex-specific competition conditions; values below 100% indicate lower female repetition output, and values above 100% indicate higher female repetition output. The ratio is a competition-output metric and must not be interpreted as evidence of physiological equivalence.

For the multi-year benchmark analysis, ratios were calculated from winning senior Gold and Silver scores. For the detailed recent analysis, mean male and female totals and station scores were calculated. Sample standard deviations were reported for total scores. Station-level ratios were used to identify whether differences clustered systematically by movement type.

3.6 Statistical approach

Given the small and unequal category samples typical of the early competitive development of Fitness Challenge, and because station configuration changed across the observation period, the analysis was intentionally descriptive. Means, sample standard deviations, absolute score differences, and female-to-male ratios were calculated. No claim of statistical equivalence was made from non-significant or visually similar values. Inferential testing across the entire 2021-2026 period was considered inappropriate because it would conflate athlete effects, field-composition effects, and regulatory changes. The analytical emphasis was therefore on magnitude, direction, consistency, and task specificity of sex-related output patterns.

4. Analysis and Results Interpretation

4.1 Multi-year pattern of male and female winning output

The six-season benchmark series shows substantial narrowing of the aggregate sex difference compared with the inaugural competitive period, although the pattern is nonlinear and varies by standard. Table 1 summarizes senior Gold and Silver winning scores and the corresponding female-to-male performance ratio.

Table 1. Multi-year senior winning scores and female-to-male output ratios

Year	Reference event	Men Gold	Women Gold	F:M Gold %	Men Silver	Women Silver	F:M Silver %
2021	World Championships	326	203	62.3	433	373	86.1
2022	World championship season	306	231	75.5	498	417	83.7
2023	World Championships	297	292	98.3	467	450	96.4
2024	European Championships	319	273	85.6	500	486	97.2
2025	European Championships	327	297	90.8	483	502	103.9

Year	Reference event	Men Gold	Women Gold	F:M Gold %	Men Silver	Women Silver	F:M Silver %
2026*	European Championships	356	295	82.9	476	482	101.3

*2026 represents a partial competitive season. Ratios describe repetition output under sex-specific competition prescriptions and are not measures of physiological equivalence.

Gold competition displayed the greatest longitudinal variability. The female-to-male winning ratio increased from 62.3% in 2021 to 98.3% in 2023, then remained between 82.9% and 90.8% in the 2024-2026 reference competitions. Thus, the early sex gap narrowed markedly but did not disappear consistently at Gold level.

Silver displayed a different pattern. The female-to-male winning ratio was 86.1% in 2021 and 83.7% in 2022, increased to 96.4% in 2023 and 97.2% in 2024, and exceeded 100% in the 2025 and 2026 reference competitions. The result suggests that recent Silver prescriptions generated highly comparable aggregate repetition output between male and female winners.

4.2 Recent group-level comparisons

Complete recent European Championship results allow the sex comparison to be extended beyond winners. Table 2 presents mean total scores for 2025 and 2026 senior Gold and Silver fields.

Table 2. Recent European Championship total-score profiles

Year	Standard	Sex	n	Total repetitions, mean $\pm$ SD	Female:male mean ratio
2025	Gold	Men	10	293.1 $\pm$ 29.9	—
2025	Gold	Women	5	270.4 $\pm$ 27.0	92.3% of men
2025	Silver	Men	4	421.2 $\pm$ 77.6	—
2025	Silver	Women	6	420.3 $\pm$ 79.4	99.8% of men
2026	Gold	Men	8	281.1 $\pm$ 66.3	—
2026	Gold	Women	3	283.0 $\pm$ 17.4	100.7% of men
2026	Silver	Men	6	360.7 $\pm$ 76.0	—
2026	Silver	Women	6	374.8 $\pm$ 72.7	103.9% of men

At the 2025 European Championships, Silver mean performance was almost identical between sexes: 421.2 repetitions in men and 420.3 in women. Gold retained a larger aggregate difference, with women averaging 92.3% of the male mean. In 2026, the pattern shifted further toward aggregate parity: women averaged 100.7% of the male Gold mean and 103.9% of the male Silver mean.

These values should not be interpreted as evidence that men and women possess equal absolute strength or power. Instead, they show that the competition prescriptions can produce similar aggregate repetition totals from athlete groups whose physiological characteristics differ. The more informative question is therefore whether this apparent parity is distributed uniformly across stations.

4.3 Station-specific sex patterns in Gold competition

Station-level analysis demonstrates that aggregate parity conceals substantial task-specific differences. Table 3 shows female-to-male ratios calculated from mean station output in the 2025 and 2026 European Gold competitions.

Table 3. Female-to-male mean station-output ratios in Gold competition

Station / movement domain	2025 F:M %	2026 F:M %
Pull-up	69.9	89.7
KB squat-and-pull	102.4	107.9
Dips	75.5	74.2
Walking lunges	114.9	125.1
Core station*	78.9	97.5
Final whole-body station	112.6	113.7

*The core station changed across regulatory periods; ratios are interpreted as movement-domain indicators rather than direct test-retest comparisons.

The upper-body bodyweight stations showed the clearest persistent male advantage. Mean female pull-up output was 69.9% of male output in 2025 and 89.7% in 2026. Dips showed an even more stable pattern, with ratios of 75.5% and 74.2%, respectively. These results are directionally consistent with known sex differences in absolute and relative upper-body strength and with functional-fitness evidence showing larger sex gaps in gymnastics-type tasks when the movement standard itself is not externally scaled (Schlegel and Křehký, 2022).

In contrast, the externally loaded squat-and-pull station produced female outputs slightly above male means in both seasons (102.4% and 107.9%). Walking lunges showed an even larger reversal, with women averaging 114.9% of male output in 2025 and 125.1% in 2026. The final whole-body station also produced ratios above 110% in both competitions. These repeated patterns strongly suggest that sex-specific external-load calibration interacts with movement characteristics to reshape the observable competition gap.

The central implication is that no single male-female scaling factor can characterize Fitness Challenge performance. The effect of sex-specific prescription depends on whether the task is limited primarily by upper-body relative strength, lower-body muscular endurance, movement economy, metabolic tolerance, or an externally scalable resistance.

4.4 Station-specific sex patterns in Silver competition

Silver competition showed a similar but more pronounced pattern of aggregate convergence. Table 4 summarizes recent female-to-male station ratios.

Table 4. Female-to-male mean station-output ratios in Silver competition

Station	2025 F:M %	2026 F:M %
Lying pull-up	85.4	91.4
Loaded lift (Jefferson / Hexa Lift)	111.1	112.9
Bench dips	80.9	91.5
Dumbbell lunges	114.9	130.3
Sit-ups	102.1	105.5
Dumbbell snatch	108.3	117.9

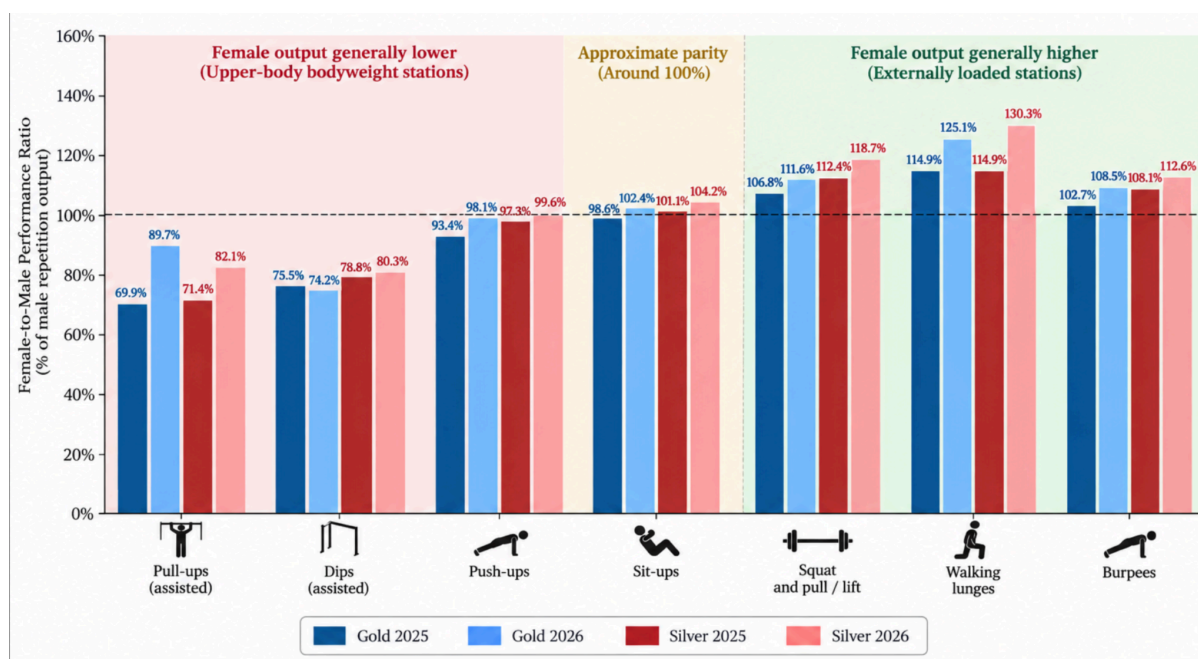
For the 2026 women's Silver data, one published station value was missing; that station mean was calculated from available observations only, with no imputation.

Silver reproduced the same directional distinction observed at Gold. Women remained below men on upper-body pulling and pressing stations, although the gaps were smaller than in Gold. Conversely, women exceeded male mean output on the externally loaded lift, dumbbell lunges, abdominal station, and dumbbell snatch in both recent seasons.

The most pronounced difference occurred in dumbbell lunges, where the female-to-male ratio increased from 114.9% in 2025 to 130.3% in 2026. This should not be interpreted as a general female physiological advantage. Rather, it indicates that the prescribed female load, combined with sex-specific anthropometry, strength-endurance characteristics, technique, and pacing, allowed greater repetition accumulation under the competition rules.

The repeated direction of these station patterns across consecutive seasons provides stronger evidence than a single-event comparison. Despite changes to the Silver lifting station between 2025 and 2026, the loaded-lift ratio remained above 110% in both seasons, whereas upper-body bodyweight stations remained below 100%.

Figure 1 provides an integrated visualization of these station-specific female-to-male output ratios across Gold and Silver competition. The 100% reference line highlights the systematic contrast between upper-body bodyweight stations, where female output generally remained below male output, and several externally loaded stations, where female output frequently exceeded male output.



Source: Authors' own conceptual contribution

Figure 1. Female-to-male mean station-output ratios in Gold and Silver IFBB Fitness Challenge competition, 2025–2026. Values represent female mean repetition output divided by male mean repetition output $\times 100$. The horizontal reference at 100% indicates equal repetition output under the prescribed sex-specific competition conditions. Values below 100% indicate higher male output, whereas values above 100% indicate higher female output. Ratios represent competition output and should not be interpreted as physiological equivalence.

4.5 Aggregate parity versus task-specific inequality

Taken together, the results reveal a central paradox of the current Fitness Challenge scoring model: men and women can produce almost identical total scores while reaching those totals through different station-level performance distributions. Aggregate parity is therefore not synonymous with station-level parity.

For competition design, this distinction is important. If the objective of sex-specific loading is to produce comparable aggregate competitive opportunity and similar total repetition ranges, the recent Silver and Gold results suggest that the current approach can achieve substantial convergence. If the objective is to equalize performance at every station, however, the data indicate that a single sex-specific loading philosophy is insufficient because bodyweight upper-body tasks and externally loaded lower-body tasks respond differently.

5. Discussion

5.1 Principal findings

This study provides the first multi-year analysis focused specifically on sex-related performance patterns in IFBB Fitness Challenge. Four findings are most important. First, the aggregate male-female competition gap narrowed markedly after the inaugural period, especially at Silver level. Second, recent group means demonstrate near-equivalent total repetition output under sex-specific competition prescriptions. Third, this aggregate convergence masks stable station-specific differences. Fourth, the direction of those station differences is strongly associated with task type: upper-body bodyweight stations generally favor men, whereas several externally loaded lower-body and whole-body stations produce female output at or above male levels.

These findings support the central hypothesis that sex-specific load prescriptions can attenuate aggregate performance differences without eliminating task-specific biological and mechanical effects.

5.2 Why upper-body bodyweight stations retain a male advantage

The most consistent male advantages occurred in pull-up and dip variants. This pattern is physiologically plausible. Men possess greater absolute upper-body muscle mass and strength after puberty, and the proportional

sex difference is generally larger in upper-body than lower-body force production (Hunter et al., 2023; Hunter and Senefeld, 2024). When an exercise uses body mass as the external resistance, organizers cannot independently reduce the prescribed load for women in the same manner as for a barbell, kettlebell, or dumbbell task.

Competition performance also depends on anthropometry. D'Hulst et al. (2024) reported that anthropometric characteristics were associated with CrossFit Open performance and emphasized the potential influence of limb dimensions on functional-fitness tasks. In pull-ups and dips, body mass, segment lengths, muscle cross-sectional area, and moment arms directly affect mechanical demand. Consequently, a standardized technical movement can remain sex differentiated even when overall competition scoring is well calibrated.

The persistence of a male advantage in these stations is consistent with CrossFit evidence. Schlegel and Křehký (2022) observed that scaling external weight does not automatically eliminate sex differences when gymnastics tasks remain essentially unscaled. The present Fitness Challenge data extend this concept to a more standardized six-station circuit.

5.3 Why externally loaded stations can show parity or higher female output

Externally loaded stations generated a different profile. Women frequently matched or exceeded male repetition output in squat-and-pull, loaded lifting, lunges, and dumbbell snatch. The most likely explanation is an interaction between prescribed load and the physiological characteristics relevant to repeated submaximal work.

Sex-specific external loading directly reduces absolute force requirements for women. Once the prescribed load falls within a manageable relative-intensity range, women's generally lower absolute force production may become less limiting, while local muscular endurance and lower fatigability can become more influential. Experimental literature indicates that women may sustain selected submaximal contractions longer than men, although this advantage is strongly task specific (Hunter, 2014, 2016). Fitness Challenge therefore provides an applied context in which the balance between strength reserve and fatigue resistance may shift according to station design.

This interpretation should remain cautious. The present study did not measure one-repetition maximum strength, relative load, muscle mass, lactate, or oxygen uptake. It cannot determine whether women experienced a lower percentage of maximal strength at a given station. Future laboratory and field testing should quantify the actual relative intensity represented by competition loads in elite male and female Fitness Challenge athletes.

5.4 Silver versus Gold: evidence for differential calibration

Silver produced the strongest aggregate convergence. Female winning output approached or exceeded male output in the later competitive seasons, and recent group means were nearly identical or modestly higher in women. Gold retained more variable sex differences across years despite near parity in the 2026 group mean.

This divergence is likely related to task complexity and the role of upper-body bodyweight performance. Gold includes strict pull-ups, dips, toes-to-bar, and higher-demand loaded movements. Silver uses less technically demanding variants and lower external loading. Because the larger biological sex differences are often expressed in absolute strength and power, especially in upper-body tasks, Gold may retain greater sensitivity to sex-specific physical characteristics even after external-load adjustment.

The finding has practical implications for competition governance. Load calibration should not necessarily be identical in conceptual form across competitive levels. The relative contribution of each station to total score, and the extent to which a station systematically advantages one sex division in repetition opportunity, should be monitored independently at Bronze, Silver, and Gold.

5.5 Comparison with functional-fitness research

The present results differ in an important way from the broad pattern reported in CrossFit. Schlegel and Křehký (2022) found that men achieved better results in 14 of 15 CrossFit Games events despite scaling, with sex differences ranging from trivial to large. Maigne, Grundlingh and Feito (2023a) likewise found significant sex differences in most CrossFit Open workouts. In Fitness Challenge, by contrast, recent aggregate repetition output is frequently very similar between divisions.

This may reflect structural differences between sports. CrossFit competition includes wider variation in workout duration, load, gymnastics complexity, monostructural conditioning, and scoring format. Fitness Challenge uses a fixed two-minute repetition window at each of six stations and a more constrained movement

architecture. Standardization may allow sex-specific load prescriptions to be calibrated more directly to desired repetition ranges.

Nevertheless, the station-level results show that the same problem identified in CrossFit remains: scaling one variable does not equalize all performance determinants. Pacing, technical proficiency, anthropometry, movement complexity, and the relative intensity of bodyweight tasks continue to shape sex-specific outcomes (Mangine et al., 2021; Mangine et al., 2024).

5.6 Implications for coaching and athlete profiling

The practical value of the study extends beyond competition regulation. Coaches should not interpret similar total scores in male and female athletes as evidence that identical training emphasis is appropriate. The station profile suggests that upper-body relative-strength development may be particularly important for female Gold athletes, whereas men may gain relatively more from improving fatigue resistance and repetition economy in loaded lower-body stations where female output is already highly competitive.

For both sexes, athlete profiling should distinguish total score from station-specific strengths and weaknesses. A male athlete can achieve a strong total despite losing relative ground on high-repetition lunges, while a female athlete can compensate for lower pull-up or dip output through superior performance on loaded and endurance-oriented stations. This creates different pathways to the same aggregate score.

Future performance models should therefore include station z-scores or within-sex percentiles rather than relying only on total repetitions. Such profiles could support individualized training prescription, national-team selection, and identification of station-specific performance ceilings.

5.7 Implications for competition design and load calibration

The findings also offer a data-driven framework for future load calibration. Rather than applying a universal percentage reduction from men's loads, organizers can evaluate observed female-to-male output ratios for each externally loaded station over multiple international events. Persistent ratios far above or below the intended competitive range could then trigger a technical review.

However, calibration should not aim automatically for identical repetition counts at every station. A multidimensional sport may legitimately contain sex-specific strengths and weaknesses provided that the aggregate competition remains balanced, safe, and technically meaningful. The key is transparency regarding the objective: equal physiological stress, equal relative load, equal repetition opportunity, or comparable total scoring are different design goals and cannot all be assumed to occur simultaneously.

A stronger future approach would combine competition output with direct physiological measurements. If the same sex-specific loads were expressed as percentages of individual one-repetition maximum or force capacity, researchers could determine whether apparently comparable repetition output reflects similar relative intensity. Heart rate, blood lactate, perceived exertion, and repetition-velocity loss could further establish whether internal load is comparable even when external output is not.

5.8 Strengths and limitations

The principal strength of this study is its use of official real-world competition data from the formative years of a standardized emerging sport. The analysis also distinguishes aggregate output from station-level performance, allowing a more nuanced interpretation of sex-specific scaling than winner-score comparisons alone.

Several limitations must be acknowledged. Competitive field sizes were small in some divisions, particularly among women at Gold level. Athlete samples were not random and represent self-selected international competitors. The same athletes did not participate in every season, so multi-year trends reflect both athlete development and changes in field composition. Exercise configurations and technical regulations evolved during the observation period, limiting direct absolute-score comparisons across all years. The detailed station analysis focused on recent European Championships and should not automatically be generalized to every region or competitive level.

Finally, competition data provide external performance but not the physiological mechanisms underlying it. The study cannot establish causal explanations for sex-specific station patterns. The proposed interpretations regarding strength reserve, fatigability, anthropometry, and relative intensity are supported by established exercise physiology but require direct testing in Fitness Challenge athletes.

6. Conclusions

Sex-specific performance in IFBB Fitness Challenge is best understood as a pattern of aggregate convergence combined with task-specific divergence. Across the 2021-2026 competitive period, female-to-male winning-output ratios increased substantially from the inaugural season, particularly at Silver level. In recent European competitions, male and female mean total scores were frequently almost identical under standardized sex-specific competition prescriptions.

The apparent parity of total scores did not extend uniformly to individual stations. Upper-body bodyweight pulling and pressing movements generally retained a male performance advantage, whereas externally loaded lower-body and whole-body stations frequently produced female repetition output comparable with or greater than male output. This pattern was reproduced across consecutive recent seasons and across Gold and Silver competition.

The findings indicate that sex-specific external-load scaling can successfully attenuate differences in aggregate repetition-based performance, but it does not remove the influence of sex-specific physiology, anthropometry, movement mechanics, and fatigue characteristics. Accordingly, total-score equality should not be interpreted as physiological equivalence.

For coaches, the results support sex- and station-specific athlete profiling. For competition organizers, they provide an empirical basis for monitoring load calibration by station rather than relying on a single global scaling assumption. Future research should combine international competition databases with direct measurements of relative strength, aerobic and anaerobic capacity, muscle fatigability, anthropometry, repetition velocity, and internal load to determine whether current prescriptions produce comparable relative physiological demands in male and female Fitness Challenge athletes.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with literature identification, manuscript structuring, language refinement, and descriptive calculation support. All competition data, calculations, references, interpretations, and AI-assisted content were subsequently reviewed and verified by the authors. The authors take full responsibility for the accuracy, integrity, and final content of the publication.

Data Availability Statement

The primary competition results analyzed in this study were derived from officially published International Federation of Bodybuilding and Fitness competition result documents. The harmonized analytical dataset and derived descriptive statistics may be made available by the corresponding author upon reasonable request, subject to the conditions governing redistribution of the original federation records.

Conflict of Interest

The authors declare no conflict of interest.

Funding

The authors received no specific funding for this study.

References

- Bellar, D., Hatchett, A., Judge, L.W., Breaux, M.E. and Marcus, L. (2015) 'The relationship of aerobic capacity, anaerobic peak power and experience to performance in CrossFit exercise', *Biology of Sport*, 32(4), pp. 315-320. doi:10.5604/20831862.1174771.
- Brandt, T., Ebel, C., Lebahn, C. and Schmidt, A. (2025) 'Acute physiological responses and performance determinants in Hyrox® - a new running-focused high intensity functional fitness trend', *Frontiers in Physiology*, 16, 1519240. doi:10.3389/fphys.2025.1519240.
- Butcher, S.J., Neyedly, T.J., Horvey, K.J. and Benko, C.R. (2015) 'Do physiological measures predict selected CrossFit® benchmark performance?', *Open Access Journal of Sports Medicine*, 6, pp. 241-247. doi:10.2147/OAJSM.S88265.
- Claudino, J.G., Gabbett, T.J., Bourgeois, F., Souza, H.S., Miranda, R.C., Mezêncio, B., Soncin, R., Cardoso Filho, C.A., Bottaro, M., Hernandez, A.J., Amadio, A.C. and Serrão, J.C. (2018) 'CrossFit overview: Systematic review and meta-analysis', *Sports Medicine - Open*, 4, 11. doi:10.1186/s40798-018-0124-5.

- D'Hulst, G., Hodžić, D., Leuenberger, R., Arnet, J., Westerhuis, E., Roth, R., Schmidt-Trucksäss, A., Knaier, R. and Wagner, J. (2024) 'Physiological profiles of male and female CrossFit athletes', *International Journal of Sports Physiology and Performance*, 19(8), pp. 780-791. doi:10.1123/ijsp.2023-0386.
- Feito, Y., Heinrich, K.M., Butcher, S.J. and Poston, W.S.C. (2018) 'High-intensity functional training: Definition and research implications for improved fitness', *Sports*, 6(3), 76. doi:10.3390/sports6030076.
- Handelsman, D.J., Hirschberg, A.L. and Berman, S. (2018) 'Circulating testosterone as the hormonal basis of sex differences in athletic performance', *Endocrine Reviews*, 39(5), pp. 803-829. doi:10.1210/er.2018-00020.
- Hunter, S.K. (2014) 'Sex differences in human fatigability', *Acta Physiologica*, 210(4), pp. 768-789. doi:10.1111/apha.12234.
- Hunter, S.K. (2016) 'Sex differences in fatigability of dynamic contractions', *Experimental Physiology*, 101(2), pp. 250-255. doi:10.1113/EP085370.
- Hunter, S.K. and Enoka, R.M. (2001) 'Sex differences in the fatigability of arm muscles depends on absolute force during isometric contractions', *Journal of Applied Physiology*, 91(6), pp. 2686-2694. doi:10.1152/jappl.2001.91.6.2686.
- Hunter, S.K., Angadi, S.S., Bhargava, A., Harper, J., Hirschberg, A.L., Levine, B.D., Moreau, K.L., Nokoff, N.J., Stachenfeld, N.S. and Berman, S. (2023) 'The biological basis of sex differences in athletic performance: Consensus statement for the American College of Sports Medicine', *Medicine & Science in Sports & Exercise*, 55(12), pp. 2328-2360. doi:10.1249/MSS.0000000000003300.
- Hunter, S.K. and Senefeld, J.W. (2024) 'Sex differences in human performance', *The Journal of Physiology*, 602(17), pp. 4129-4156. doi:10.1113/JP284198.
- International Federation of Bodybuilding and Fitness (2021) IFBB Fitness Challenge World Championships: Official Results. Santa Susanna, Spain, 3-8 November 2021. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2022a) IFBB Fitness Challenge Female World Championship: Official Results. Santa Susanna, Spain, 2-7 November 2022. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2022b) IFBB Fitness Challenge Men's World Championships: Official Results. Tehran, Iran, 7-11 December 2022. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2023) 2023 IFBB World Fitness Challenge Championships: Official Results. Kyrgyzstan. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2024a) Rules - IFBB Fitness Challenge. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2024b) European Fitness Challenge Competition 2024: Official Results. Santa Susanna, Spain. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2025a) IFBB European Fitness, Bodybuilding & Fitness Challenge Championships 2025: Official Results. Santa Susanna, Spain, 30 April-5 May 2025. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2025b) IFBB SpeedFit Fitness Challenge: Rules, July 2025. Madrid: IFBB.
- International Federation of Bodybuilding and Fitness (2026) IFBB European Fitness, Bodybuilding & Fitness Challenge Championships: Fitness Challenge Official Results. Santa Susanna, Spain, 30 April-4 May 2026. Madrid: IFBB.
- Joyner, M.J., Hunter, S.K. and Senefeld, J.W. (2025) 'Evidence on sex differences in sports performance', *Journal of Applied Physiology*, 138(1), pp. 274-281. doi:10.1152/jappphysiol.00615.2024.
- Mangine, G.T., Tankersley, J.E., McDougale, J.M., Velazquez, N., Roberts, M.D., Esmat, T.A., VanDusseldorp, T.A. and Feito, Y. (2020) 'Predictors of CrossFit Open performance', *Sports*, 8(7), 102. doi:10.3390/sports8070102.
- Mangine, G.T., Feito, Y., Tankersley, J.E., McDougale, J.M. and Kliszczewicz, B.M. (2021) 'Workout pacing predictors of CrossFit® Open performance: A pilot study', *Journal of Human Kinetics*, 78(1), pp. 89-100. doi:10.2478/hukin-2021-0043.
- Mangine, G.T., Grundlingh, N. and Feito, Y. (2023a) 'Normative scores for CrossFit® Open workouts: 2011-2022', *Sports*, 11(2), 24. doi:10.3390/sports11020024.
- Mangine, G.T., Grundlingh, N. and Feito, Y. (2023b) 'Differential improvements between men and women in repeated CrossFit Open workouts', *PLOS ONE*, 18(11), e0283910. doi:10.1371/journal.pone.0283910.
- Mangine, G.T., Zeitz, E.K., Dexheimer, J.D., Hines, A., Lively, B. and Kliszczewicz, B.M. (2024) 'The influence of sex-division, experience, and pacing strategy on performance in the 2020 CrossFit® Open', *Frontiers in Sports and Active Living*, 6, 1344036. doi:10.3389/fspor.2024.1344036.
- Martínez-Gómez, R., Valenzuela, P.L., Barranco-Gil, D., Moral-González, S., García-González, A. and Lucia, A. (2019) 'Full-squat as a determinant of performance in CrossFit', *International Journal of Sports Medicine*, 40(9), pp. 592-596. doi:10.1055/a-0960-9717.
- Martínez-Gómez, R., Valenzuela, P.L., Alejo, L.B., Gil-Cabrera, J., Montalvo-Pérez, A., Talavera, E., Lucia, A. and Barranco-Gil, D. (2020) 'Physiological predictors of competition performance in CrossFit athletes', *International Journal of Environmental Research and Public Health*, 17(10), 3699. doi:10.3390/ijerph17103699.

- Meier, N. (2021) 'Determination of a CrossFit® benchmark performance profile', *Sports*, 9(6), 80. doi:10.3390/sports9060080.
- Meier, N., Schlie, J. and Schmidt, A. (2023) 'CrossFit®: Unknowable or predictable? A systematic review on predictors of CrossFit® performance', *Sports*, 11(6), 112. doi:10.3390/sports11060112.
- Ponce-García, T., García-Romero, J., Carrasco-Fernández, L., Castillo-Domínguez, A. and Benítez-Porres, J. (2025) 'Sex differences in anaerobic performance in CrossFit® athletes: A comparison of three different all-out tests', *PeerJ*, 13, e18930. doi:10.7717/peerj.18930.
- Roberts, B.M., Nuckols, G. and Krieger, J.W. (2020) 'Sex differences in resistance training: A systematic review and meta-analysis', *Journal of Strength and Conditioning Research*, 34(5), pp. 1448-1460. doi:10.1519/JSC.0000000000003521.
- Sauvé, B., Haugan, M. and Paulsen, G. (2024) 'Physical and physiological characteristics of elite CrossFit athletes', *Sports*, 12(6), 162. doi:10.3390/sports12060162.
- Schlegel, P. and Křehký, A. (2022) 'Performance sex differences in CrossFit®', *Sports*, 10(11), 165. doi:10.3390/sports10110165.
- Serafini, P.R., Feito, Y. and Mangine, G.T. (2018) 'Self-reported measures of strength and sport-specific skills distinguish ranking in an international online fitness competition', *Journal of Strength and Conditioning Research*, 32(12), pp. 3474-3484. doi:10.1519/JSC.0000000000001843.
- Suchomel, T.J., Nimphius, S. and Stone, M.H. (2016) 'The importance of muscular strength in athletic performance', *Sports Medicine*, 46(10), pp. 1419-1449. doi:10.1007/s40279-016-0486-0.
- Wang, X., Soh, K.G., Samsudin, S., Deng, N., Liu, X., Zhao, Y. and Akbar, S. (2023) 'Effects of high-intensity functional training on physical fitness and sport-specific performance among athletes: A systematic review with meta-analysis', *PLOS ONE*, 18(12), e0295531. doi:10.1371/journal.pone.0295531.