Skattebo, Ø., Hallén, J., Rønnestad, B., Losnegard, T. (2015). Upper body heavy strength training does not affect performance in junior female cross-country skiers. *Scandinavian Journal of Medicine & Science in Sports, 26*, 1007-1016.

Title: Upper-body heavy strength training does not affect performance in junior female cross-country skiers

Submission type:

Original investigation

Authors:

Øyvind Skattebo[1], Jostein Hallén[1], Bent Ronny Rønnestad[2] & Thomas Losnegard[1]

Institutions:

[1]Department of Physical Performance, The Norwegian School of Sport Sciences, Oslo, Norway, [2]Section of Sport Sciences, Lillehammer University College, Lillehammer, Norway

Address for correspondence:

Thomas Losnegard

The Norwegian School of Sport Sciences

Post box 4014 Ullevål Stadion, 0806 Oslo, Norway

Email: thomas.losnegard@nih.no

Fax: +47 22 23 42 20

Tel: +47 22 26 22 73

Running head: Heavy strength training in female XC skiers

Abstract Word Count: 197

Number of Figures and Tables: 5 figures and 3 tables

Abstract

We investigated the effects of adding heavy strength training to a high volume of endurance training on performance and related physiological determinants in junior female cross-country skiers. Sixteen well-trained athletes (17 ± 1 yrs, 60 ± 6 kg, 169 ± 6 cm, VO_{2max} running: 60 ± 5 mL · kg^{-1} · min^{-1}) were assigned either to an intervention group (INT; n=9) or a control group (CON; n=7). INT completed two weekly sessions of upper-body heavy strength training in a linear periodized fashion for 10 weeks. Both groups continued their normal aerobic endurance- and muscular endurance training. One repetition maximum in seated pull-down increased significantly more in INT than in CON, with a group difference of 15 ± 8% ($p<0.01$). Performance, expressed as average power output on a double poling ergometer over 20 sec and as 3 min with maximal effort in both rested (sprint-test) and fatigued states (finishing-test), showed similar changes in both groups. Sub-maximal O_2-cost and VO_{2peak} in double poling showed similar changes or were unchanged in both groups. In conclusion, ten weeks of heavy strength training increased upper-body strength but had trivial effects on performance in a double poling ergometer in junior female cross-country skiers.

Key words: Cross-country skiing, elite athletes, junior athletes, resistance training

Introduction

Over the last two decades the race velocity in distance cross country (XC) skiing has increased ~5-8% in World Cup races, and in sprint-skiing (≤1.8 km), an approximately 20% higher average velocity is reached compared to distance skiing (≥10/15 km, women/men) (Losnegard, 2013). In World Championships and Olympics, 5 of 6 race events are sprints (~3 min) or mass starts (~40-150 min) and the final ranks are often determined in a final sprint. Therefore, the ability to perform supra-maximal workloads seems important. As a consequence, heavy strength training has attracted interest, both in research and practical situations, as a training model for improving such abilities (Stöggl et al., 2007; Losnegard & Hallén, 2014). Losnegard et al. (2011) and Stöggl et al. (2007) documented high to almost perfect correlations between upper-body strength and power and double poling performance on both roller skis and double poling ergometers. Since the correlation between upper-body 1 repetition maximum (RM) strength and double poling performance was greater for women than men, it was argued that women could potentially improve performance by increasing muscle strength (Losnegard et al., 2011). As differences in performance between gender are larger with exercises that involve a significant upper-body contribution (Sandbakk et al., 2014), improved upper-body strength can potentially improve double poling performance in female XC skiers.

Studies have demonstrated that systematic heavy strength training (and subsequent strength gains) can reduce O_2-cost during double poling on roller skis and double poling ergometers in highly trained skiers (Østerås et al., 2002; Mikkola et al., 2007), albeit that not all studies have confirmed this during roller ski skating (Losnegard et al., 2011; Rønnestad et al., 2012). Increased maximal strength has resulted in increased time to exhaustion (Hoff et al.,1999, 2002; Østerås et al., 2002) and 5 min power output on double poling ergometers (Losnegard et al., 2011). However, improvements in skating or double poling time trials (1.1 - 7.5 km) have not been found after concurrent strength and endurance training in XC skiers (Mikkola et al., 2007; Losnegard et al., 2011; Rønnestad et al., 2012).

Given the inconsistent findings above, we examined the effect of supplementing endurance training with heavy strength training in junior female XC skiing from a performance perspective. In addition, the main determinants of performance in a sport-specific exercise were addressed. Performance in endurance sports such as XC skiing is mainly determined by energy turnover (maximal oxygen uptake; VO_{2max}, utilization of VO_{2max} and anaerobic capacity) and work economy (e.g. O_2-cost) (di Prampero, 2003). Since no earlier study has included a performance test after a prolonged fatiguing sub-maximal load to simulate intensive skiing at the end of a mass start, a performance test was also designed to capture this. We hypothesized that heavy strength training would lead to increased muscle strength and reduced sub-maximal O_2-cost with subsequent increased double poling performance.

Method

Subjects

A total of 19 female junior XC skiers started the study, but 3 participants were subsequently excluded due to injuries, illness or inability to complete the required number of strength training sessions (minimum 85% adherence). There were no inclusion or exclusion criteria related to previous strength training experience. The participants were all among the top 30 in their respective competition classes in the Norwegian national cup (mean: 12; range 1-30) and 3 of the athletes participated in the Junior World Championship the following winter. Detailed written information was given to all participants before inclusion, and the participants gave their written informed consent before study participation. Parental approval was obtained for participants younger than 18 yrs. The project was evaluated by the Norwegian Regional Committee for Medical Research Ethics and performed according to the Declaration of Helsinki.

After pre-testing, subjects were self-selected either to an intervention group (INT: 18 ± 1 yrs, 171 ± 5 cm, 61 ± 4 kg, n=9) or a control group (CON: 17 ± 1 yrs, 166 ± 6 cm, 60 ± 9 kg, n=7). The self-selection was based on the athlete's preferences in collaboration with their respective coaches. Such selection has previously been frequently used when studying high level athletes (Losnegard et al., 2011; Rønnestad et al., 2012). There were no significant differences in anthropometrics between groups, but INT was slightly older than CON ($p<0.05$).

Study design

Table 1 shows the experimental design. A week before pre-testing, the skiers underwent an extensive familiarization with protocols and testing equipment. The main tests were performed over two weeks, with four test days and in the following order: Double poling protocol 1 (double poling ergometer); Double poling protocol 2 (double poling ergometer); 1RM strength test; and VO_{2max} running test. All tests were separated by a minimum of 48 hours, except the VO_{2max} running test, which was performed 24 hours after the 1RM strength test for half of the subjects. On test days, subjects did not perform any training before testing. The day before test days, subjects exercised for a maximum of two hours at low intensities (<75% of maximum heart rate (HR_{max})). Subjects were instructed to perform the same training procedures during the test period at both pre- and post-testing.

Between test periods, INT supplemented their endurance training with two weekly sessions of upper-body heavy strength training. CON, on the other hand, continued their normal endurance training and did not include heavy strength training. The intervention lasted 10 weeks, from mid-September to the end of November (late pre-competition period for XC skiers).

<<Table 1 near here>>

Testing Procedures

Familiarization

Two familiarization sessions were conducted, using both the strength training apparatus and the double poling ergometer (ThoraxTrainer Elite, ThoraxTrainerApS, Kokkedal, Denmark). Strength familiarization started with a technical introduction to the exercise used for 1RM-testing, followed by three sub-maximal sets with increasing load (10 repetitions at 40%, 6 repetitions at 75% and 3 repetitions at 85% of estimated 1RM). Then two heavier, almost maximal lifts were performed to estimate 1RM. Double poling familiarization started with a 10 min unspecified general testing of the ergometer, followed by either Double poling protocol 1 or 2 (described later). Familiarization was also performed two weeks before post-testing, and consisted of one strength session (same procedures as before pre-testing) and one double poling session (Double poling protocol 2).

Double poling protocol 1 (test day 1)

The same warm-up procedure was used before Double poling protocol 1 and 2, and consisted of 5 min running and 3 min double poling (60-75% of HR_{max}). Double poling protocol 1 started with two continuous sub-maximal workloads (30W and 40W), each lasting 5 min and followed by 3 min rest (Figure 1). Thereafter, subjects performed 2 bouts with maximal effort, each lasting 20 sec and separated by rests of 2 min 40 sec. The highest mean power output of the two attempts was considered as the performance on the 20 sec test. After 4 min and 40 sec rest, subjects performed a 3 min bout with maximal effort (sprint-test). Effect production (W) was fixed for the first 30 sec to avoid "overpacing" and was individually set, based on preliminary tests (familiarization). O_2-uptake was measured continuously during both double poling protocols with an automated system (Oxycon Pro; Jaeger Instrument, Hoechberg, Germany), using breath-by-breath technique validated by Rietjens et al. (2001). The gas analyzers and the flow turbine (Triple V; Erick Jaeger GmbH, Hoechberg, Germany) were calibrated according to the instruction manual as described in detail previously (Losnegard et al., 2011). In addition, the system was calibrated for air humidity before each test using a hygrometer. HR was measured continuously (Polar RS800; Polar Electro Oy, Kempele, Finland) and rate of perceived exertion (RPE) (Borg, 1982) was recorded after sub-maximal workloads. Sub-maximal O_2-cost was calculated as average O_2-uptake from the 3rd to 5th min at each exercise stage. Due to the short duration of the test-protocol with maximal effort (3 min), the highest 30 sec O_2-uptake was considered as VO_{2peak}, a method found not to differ from averaging over 60 sec in longer incremental protocols in our laboratory (Losnegard et al., 2012).

<<Figure 1 near here>>

Double poling protocol 2 (test day 2)

Double poling protocol 2 (Figure 1) consisted of three continuous sub-maximal workloads (60%, 70% and 80% of VO_{2peak}), each lasting 5 min and followed by 2 min rest. Thereafter, a 10 min sub-maximal workload at 70% of VO_{2peak} was conducted, directly followed by a 3 min bout with maximal effort (finishing-test). Relative sub-maximal workloads were calculated individually by linear extrapolation from sub-maximal O_2-cost and VO_{2peak} measured in Double poling protocol 1. Double poling cycle rate was recorded using ThoraxTrainer software (ThoraxTrainerAnalyzer 1.52, ThoraxTrainer Aps, Kokkedal, Denmark).

1RM strength test and upper-arm circumference (test day 3)

Upper-arm circumference was measured with a measuring tape (Seca model 201; Seca GmbH & Co., Hamburg, Germany). Segment-length of the upper-arm was considered as the length between the angulus acromii (scapula) and olecranon (ulna). Circumference was measured at 50%, 60% and 70% of the segment-length from the angulus acromii at both arms, and given as the average of all six measurements. 1RM strength was measured in seated pull-down (Losnegard et al., 2011). First, subjects warmed up with 10 min running (60-70% of HR_{max}) followed by three sub-maximal sets in seated pull-down with gradually increasing load (10 repetitions at 40%, 6 repetitions at 75% and 3 repetitions at 85% of estimated 1RM). The starting load was 95% of estimated 1RM. After each successful attempt, the load was increased by 2-5% until the subject failed to lift the load. The rest periods between attempts were 2-3 min. All tests were supervised by the same investigator at both pre- and post-testing. Criteria for accepted lifts are described elsewhere (Losnegard et al., 2011).

VO_{2max} running (test day 4)

After 25 min standardized warm-up (60-80% of HR_{max}), subjects ran at a constant incline (10.5%) on a treadmill (Woodway Desmo-Evo, Woodway GmbH, Weil am Rhein, Germany) while the speed increased progressively by 1 $km \cdot h^{-1}$ every 60 sec until exhaustion. Starting speed was 7 $km \cdot h^{-1}$ and the subjects were exhausted within 6.0-8.5 min. HR and O_2-uptake were measured continuously. VO_{2max} was taken as the highest average O_2-uptake over 60 sec and the ergospirometry system was identical to the double poling tests. Before all ergospirometry testing, body weight was measured (Seca model nr: 877; Seca GmbH & Co., Hamburg, Germany).

Intervention

Before each training session, subjects performed a 10 min general aerobic warm-up, running or cycling. Thereafter, subjects performed two sub-maximal warm-up sets in seated pull-down (10 repetitions at 40% and 6 repetitions at 75% of pre-test 1RM). The strength training program consisted of three exercises with an identical order for all training sessions: seated pull-down, standing double poling and triceps press. All exercises utilized a custom-made handlebar designed to imitate the pole-grip in XC skiing (Losnegard et al., 2011). All exercises were conducted with 3 sets per exercise and the training load and number of repetitions ranged from 10RM to 4RM through the intervention (Table 2). Each repetition was conducted with maximal mobilization in the concentric phase (lasting approximately 1sec) followed by a slower eccentric phase (2-3sec). Rests between sets were 2-3 min. Each session, including warm-up, lasted ~40 min.

<<Table 2 near here>>

Subjects in INT were encouraged to attend strength training sessions with trained coaches, which were scheduled 2-3 times per week. This was to ensure that the subjects lifted with proper technique and optimal load. Both groups could undertake strength training on the upper-body muscles with a moderate training load (>25 RM) and is further referred to as muscular endurance training. The muscular endurance training consisted mainly of two exercises; dips on bench and knee push-ups/ incline push-ups. Both groups also performed core stability training which consisted of sling exercises, Swiss Ball exercises and various mat exercises. The aerobic endurance training in INT and CON was managed by the athletes themselves after consulting with their respective coaches. All except one athlete in INT and CON were part of the same training group and had ~6 joint training sessions per week, mostly consisting of aerobic endurance training. Training through the intervention period was logged according to training mode and intensity, and given to the project coordinator. Aerobic endurance training was categorized into three intensity zones: low-intensity training (LIT: 60-81% of HR_{max}), moderate-intensity training (MIT: 82-87% of HR_{max}) and high-intensity training (HIT: ≥88% of HR_{max}).

Reliability and validity of the double poling test

Sub-maximal O_2-cost on the double poling ergometer had a typical error (coefficient of variation) of 2.2 ± 1.4%, including the biological variation in the test subjects and the error of the ergospirometry system, based on 35 steady state measurements (intraclass correlation= 1.00). Average power output on the sprint-test did not change from the first familiarization session to pre-test (difference: 1.7 ± 3.5%; $p>0.5$; ES: 0.09) (n=19). One week after post-testing, the participants in the present study performed a sprint-prologue skating and a 5 km distance race classic style on snow (national qualification races for the Junior World Championship). Average power outputs on the 3 min sprint-test were very highly correlated with the sprint-prologue time (r ± 90% confidence interval: r= -0.73 ± 0.24; $p<0.01$; n=14) and highly correlated with the 5 km time (r= -0.54 ± 0.32; $p<0.05$; n=16).

Statistics

Raw data are expressed as mean ± standard deviation if not otherwise stated. Relative changes from pre- to post-test are expressed as mean change ± 90% confidence interval (CI). Group-comparison at baseline was examined using an unpaired Student's t-test. Within-group and between-group changes were detected with a paired Student's t-test and two-way repeated-measures ANOVA respectively. If confounding factors were suspected, ANCOVA was conducted. Within- and between-group changes (% and magnitude of difference as Cohen's *d* effect size (ES)) were calculated via log-transformation of raw data (Hopkins, 2006). The magnitude of the difference was classified as trivial (ES<0.2), small (0.2≤ES<0.6), moderate

(0.6≤ES<1.2), large (1.2≤ES<2.0) or very large (ES≥2.0) (Hopkins, 2000). All ESs are expressed in favor of the INT group (i.e. a larger reduction in a variable in INT compared to CON gives a negative ES when groups are compared). Classifications of magnitude of difference in performance changes between INT and CON were also described with probabilities. The probabilistic terms to describe beneficial, harmful or trivial effects were used with the following scale: <0.5%, most unlikely; 0.5-5%, very unlikely; 5-25%, unlikely; 25-75%, possibly; 75-95%, likely; 95-99.5%, very likely; >99.5%, most likely (Hopkins et al., 2009). Correlation analyses were conducted with Pearson's product-moment correlation and correlation coefficients (r) were classified as small (0.1≤r<0.3), moderate (0.3≤r<0.5), high (0.5≤r<0.7), very high (0.7≤r<0.9) and almost perfect (r≥0.9) (Hopkins, 2000). Microsoft Office Excel 2007 (Microsoft, Redmond, USA) and IBM SPSS Statistics 20.0 (International Business Machines, New York, USA) were used for statistical analysis. The level of confidence was set to 90% and a p-value ≤0.1 was, therefore, considered statistically significant.

Results

Training during the intervention period

The groups displayed trivial to small group differences in the volumes of aerobic endurance training and core stability training during the 10-week intervention period (Table 3). Only INT underwent heavy strength training, and the subjects attended 98 ± 3% of the planned strength training sessions (range: 95-100%). CON performed ~11 min more muscular endurance training than INT per week (Table 3; p<0.01). In terms of exercise modes of aerobic endurance training, both groups performed similar amounts of ski-skating (~27%), ski-classic (~28%), running (~40 %) and other exercise (~5%; cycling, rowing etc.).

<<Table 3 near here>>

Muscle strength and anthropometrics

1RM increased significantly more in INT (24 ± 5%; p<0.01) than in CON (8 ± 7%; p<0.05), with a group difference of 15 ± 8% (p<0.01; ES: 0.90) (Figure 2A). At post-test, INT had significantly higher 1RM than CON (p<0.05). Upper-arm circumference increased significantly more in INT (3.3 ± 0.7%; p<0.01) than in CON (2.0 ± 1.2; p<0.05), with a group difference of 1.3 ± 1.3% (p=0.05; ES: 0.18). Body weight increased significantly in both INT (2.5 ± 1.2%; p<0.01) and in CON (2.6 ± 1.9%; p<0.05), with no significant group difference (-0.1 ± 2.1%; ES: -0.01).

Double poling performance

No significant group differences were seen before or after the intervention period in any of the performance tests (Figure 2B-D). Group comparisons showed possibly to likely trivial intervention effects in all three tests (ES of the difference: 0.04-0.07). The probabilities for beneficial or harmful effects were unlikely to very unlikely. Average power output (W) increased 17.1 ± 6.9% in INT and 16.2 ± 3.0% in CON in the sprint-test (both p<0.01; Figure 2C), 14.9 ± 4.9% in INT and 13.1 ± 3.5% in CON in the finishing-test (both p<0.01; Figure 2D) and 17.1 ± 5.7% in INT and 15.7 ± 4.4% in CON in the 20 sec test (both p<0.01; Figure 2B). For all performance tests, the poling frequency was not significantly changed from pre- to post-test.

<<Figure 2 near here>>

Sub-maximal O_2-cost during double poling

No significant changes were seen in sub-maximal O_2-cost from pre- to post-test in either INT or CON (Figure 3). There were no significant group differences, but INT reduced O_2-cost by small ESs on all sub-maximal workloads compared with CON (ES: -0.27 to -0.31). HR was significantly reduced from pre- to post-test in CON ($p<0.05$) (Figure 3), and reduced by small to moderate ESs compared to INT (ES: 0.47-0.68). RPE was significantly reduced in both INT and CON at 60% of VO_{2peak} and only in INT at 70% of VO_{2peak} (Figure 3).

<<Figure 3 near here>>

Maximal aerobic power

VO_{2max} running was statistically unchanged in both INT (-2.0 ± 2.1%) and CON (1.7 ± 3.5%) in absolute terms ($L \cdot min^{-1}$) (Figure 4A). Standardized to body-weight ($mL \cdot kg^{-1} \cdot min^{-1}$), VO_{2max} reduced significantly in INT (-3.7 ± 2.2; $p<0.05$) and was unchanged in CON (0.0 ± 2.8%) (Figure 4B). VO_{2peak} double poling increased significantly in both INT (2.9 ± 2.8 %; $p<0.1$) and CON (7.7 ± 6.4%; $p<0.1$) in absolute terms (Figure 4C). Standardized to body-weight, VO_{2peak} was unchanged in both INT (0.6 ± 1.9%) and CON (4.7 ± 5.0%) (Figure 4D). Group comparisons showed small negative intervention effects in INT compared to CON on all aerobic measurement terms (ES: -0.32 to -0.48). Pre- to post-test changes in VO_{2max} and VO_{2peak} (absolute and relative) and age were included as covariates in ANCOVA models to test whether these variables confounded the results on the sprint- and finishing-test. There were still no group differences between INT and CON ($p \geq 0.2$ for all models). On average for INT and CON, VO_{2peak} double poling changed significantly from 88.4 ± 1.7% to 92.4 ± 2.2% of VO_{2max} running between pre- and post-test ($p<0.01$).

<<Figure 4 near here>>

Correlations

At pre-test, 1RM strength showed high to very high significant correlations with power output on the sprint-test (r ± 90% CI: r=0.70 ± 0.24; $p<0.01$), finishing-test (r=0.65 ± 0.27; $p<0.01$) and the 20 sec test (r=0.74 ± 0.21; $p<0.01$) (Figure 5A-C). However, there were only small and non-significant correlations between pre- to post-test changes in 1RM strength and changes in power output in the three double poling tests (r=0.19 ± 0.41, r=0.17 ± 0.42 and r=0.19 ± 0.41 respectively). Upper-arm circumference had a very high correlation with 1RM strength at pre-test (r=0.88 ± 0.11; $p<0.01$) (Figure 5D), and the relative changes of the two variables were also highly correlated (r=0.54 ± 0.32; $p<0.05$).

<<Figure 5 near here>>

Discussion

The main findings of this study were that adding heavy strength training to a high volume of endurance training for 10 weeks in junior female XC skiers increased muscle strength and upper-arm circumference, but had only trivial to small effects on performance, sub-maximal O_2-cost and VO_{2peak} in double poling.

The increase in 1RM for INT (24%) is in the upper range compared with previous studies involving concurrent upper-body strength and endurance training in XC skiers (10-23%) (Hoff et al., 1999, 2002; Østerås et al., 2002; Losnegard et al., 2011; Rønnestad et al., 2012). Notably, 1RM was also significantly increased in CON (8%), which could be related to the subjects maturation state (16-19 yrs) and possibly influenced by CON being one year younger than INT. The increased 1RM in CON could also be an effect of ski-specific training, since upper-body muscles are used to a large extent in all ski techniques (Calbet et al., 2005; Smith et al., 2009; Bojsen-Møller et al., 2010). In addition, both groups performed some muscular endurance training and "strength training" with a resistance even as low as 15% of 1RM can increase muscle strength (Moss et al., 1997). Nevertheless, INT increased 1RM and upper-arm circumference significantly more than CON during the intervention period and was significantly stronger than CON at post-test.

The very high correlation found between upper-arm circumference and 1RM seated pull-down is in accordance with Losnegard et al. (2011), who showed an almost perfect (r=0.91) correlation between 1RM seated pull-down and the cross-sectional area of m. Triceps Brachii. Losnegard et al. (2011) did not, however, find any significant correlation between pre- to post-test changes in the two variables (Losnegard, T., unpublished observation). A mismatch between strength gains and muscle growth is repeatedly found when subjects inexperienced in heavy strength training start using this training mode regularly. This is often explained by neural adaptations, especially in coordinatively demanding exercises (Moritani & deVries, 1979; Folland & Williams, 2007) and has likely also contributed to the strength gains in this study.

Surprisingly, INT showed small negative changes in VO_{2max} and VO_{2peak} compared with CON during the intervention period. This is in contrast with previous studies, where concurrent heavy strength and endurance training compared with only endurance training has led to similar or more favorable changes in VO_{2max} and VO_{2peak} (Hoff et al., 1999, 2002; Østerås et al., 2002; Mikkola et al., 2007; Losnegard et al., 2011; Rønnestad et al., 2012). Heavy strength training has a long recovery period, where muscle function can be lowered for more than 48 hours after exercise (Raastad et al., 2003). Combined with the fact that INT exercised ~1 hour per week more than CON, the negative developments in VO_{2max} and VO_{2peak} could, therefore, be a result of a to high training strain. Interestingly, when group data were pooled, the difference between VO_{2max} running and VO_{2peak} double poling significantly reduced between pre- and post-test. We speculate that ski-specific training during the intervention period led to this change and that this emphasizes the need for specificity in training.

Sub-maximal O_2-cost did not significantly change in the present study; this is both in conflict (Hoff et al., 1999, 2002; Østerås et al., 2002; Mikkola et al., 2007) and in agreement (Losnegard et al., 2011; Rønnestad et al., 2012) with previous findings conducted in XC skiing. Although the intervention groups in Mikkola et al. (2007) and Østerås et al. (2002) showed significant reductions in O_2-cost between pre- and post-test, the changes did not significantly differ from the control groups. In addition, Hoff et al. (1999, 2002) used unconventional methods to measure O_2-cost. Hoff et al. (1999) measured O_2-cost as O_2-uptake divided by the speed of double poling during a time to exhaustion test at maximal aerobic velocity. Furthermore in Hoff et al. (2002), O_2-cost was taken as the O_2-uptake one minute after an increase in speed in an incremental protocol with the duration of stages being only two minutes (load corresponding to ~75% of $VO_{2\text{-peak}}$ double poling at pre-test). Since a steady state O_2-uptake is obtained after ~3 min at sub-maximal workloads (Whipp & Wasserman, 1972) and will continue to rise and not reach a steady state at maximal aerobic velocity (Jones et al., 2011), it can be argued that Hoff et al. (1999, 2002) never measured O_2-cost in the steady state. Therefore, no study has, to our knowledge, found a convincing beneficial effect of heavy strength training on exercise economy in XC skiers. However, INT showed a reduced O_2-cost with small ESs compared with CON from pre- to post-test in the present study. Because of this and the fact that only the intervention groups in Mikkola et al. (2007) and Østerås et al. (2002) reduced their O_2-cost, we cannot rule out the possibility that heavy strength training could produce a significant beneficial effect on O_2-cost during longer experiments.

Even though 1RM and power output in the double poling tests were highly correlated at baseline, no significant effect of adding heavy strength training (with subsequent strength gains) was found on performance in our group of XC skiers. In fact, the probabilities for beneficial or harmful effects were unlikely to very unlikely. Nor was there found any significant correlation between the delta changes of 1RM and performance. The inconsistencies could be explained by a spurious relationship between 1RM strength and performance at baseline. Correlation plots revealed two clear outliers (the strongest and the weakest athletes in terms of 1RM strength performed best and worst on the double poling tests respectively). When these individuals were removed from the plots there was no apparent relationship between 1RM strength and power output on the double poling tests (r= 0.12-0.35).

The finishing-test was performed after sub-maximal exercise in order to mimic performance in a fatigued state. In cyclists, it has been observed that adding heavy strength training to the normal endurance training improved 5 min all-out power after 3 hours sub-maximal cycling (Rønnestad et al., 2011). The latter indicates that strength training might reduce or delay the development of fatigue during prolonged sub-maximal exercise. However, this was not observed in the present study. This different finding may be related to the fact that there were no differences in mean power output during the 3 min sprint-test in a "fresh" state and the 3 min finishing-test in a "fatigued" state. This indicates that the load during the 27 min sub-maximal exercise in the present study was too small to induce a performance decline and may potentially mask the effects of strength training in a more fatigued state.

The small negative pre- to post-test changes in VO_{2max} and VO_{2peak} and the higher age in INT compared with CON could possibly have affected the pre- to post-test changes in performance on the sprint- and finishing-test. However, when age and changes in VO_{2max} or VO_{2peak} were adjusted for in ANCOVA models, there was still no significant group difference. This strengthens the conclusion that heavy strength training had a trivial effect on the present measurements of performance in junior female XC skiers and falsifies our hypothesis. The results contradicts the findings of Hoff et al. (1999, 2002) and Østerås et al. (2002), whom found large improvements in a time to exhaustion test in a double poling ergometer after concurrent strength and endurance training (31-79% larger than the control groups). The reason for the discrepancy is unknown, but different test types (i.e. time to exhaustion vs. constant duration test), training programs and equipment (i.e. different types of poling ergometers) could partly explain the differing results between these studies. Time to exhaustion tests can induce large changes over time, simply because only a small change in an athlete's ability to produce power will result in a large change in time to exhaustion (Hopkins et al., 2001). The same custom-built ergometer was used in Hoff et al. (1999, 2002) and Østerås et al. (2002), and significant developments in double poling technique and ergometers have occurred over the last decades.

In the present study a linear periodized heavy strength training program was used. Training load was determined on individual RM-loads at each training session. The athletes, therefore, constantly challenged the neuromuscular system, which can explain the great strength gains, but a progressively increasing neuromuscular fatigue is also plausible. In addition, the experimental design did not include any tapering period. We cannot, therefore, rule out that an accumulated fatigue influenced INT's post-test performance. Future studies, should therefore, include an end-intervention tapering period and/or include lower intensity periods or days to minimize the risk of an accumulated neuromuscular fatigue.

Perspectives

The present study showed increased maximal strength, but not altered performance, after adding heavy strength training to endurance training in female XC skiers. Further, heavy strength training had a small beneficial, but not statistically significant effect on sub-maximal O_2-cost, and a small harmful effect on VO_{2max} running and VO_{2peak} double poling. A high VO_{2max} and fractional utilization of VO_{2max} together with optimal ski technique are important factors affecting performance in XC sprint and distance races. Hence, we recommend XC skiers and trainers to prioritize aerobic endurance training and carefully consider heavy strength training as a training mode for optimizing performance in XC skiing. However, the present and earlier studies have been of short duration (8-12 weeks) and there is still strength training protocols that needs to be examined. We cannot therefore rule out the possibility that heavy strength training or other strength training regimens could be performance-enhancing in the long run.

Acknowledgements

The authors would like to thank the athletes and coaches for their participation, interest and cooperation in this study. In addition, we would like to thank Jørgen Eikås, Mikael B. Gransjøen and Emilie Kristoffersen for their helpful assistance in the training intervention and laboratory testing. We also acknowledge NTG Lillehammer and Head Coach Per Ola Gasmann for help during the study.

References

Bojsen-Møller, J., Losnegard, T., Kemppainen, J., Viljanen, T., Kalliokoski, K. K., & Hallén, J. Muscle use during double poling evaluated by positron emission tomography. J Appl Physiol 2010: 109: 1895-1903.

Borg, G. A. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982: 14: 377-381.

Calbet, J. A., Holmberg, H. C., Rosdahl, H., van Hall, G., Jensen-Urstad, M., & Saltin, B. Why do arms extract less oxygen than legs during exercise? Am J Physiol Regul Integr Comp Physiol 2005: 289: R1448-1458.

di Prampero, P. E. Factors limiting maximal performance in humans. Eur J Appl Physiol 2003: 90: 420-429.

Folland, J. P., & Williams, A. G. The adaptations to strength training: morphological and neurological contributions to increased strength. Sports Med 2007: 37: 145-168.

Hoff, J., Gran, A., & Helgerud, J. Maximal strength training improves aerobic endurance performance. Scand J Med Sci Sports 2002: 12: 288-295.

Hoff, J., Helgerud, J., & Wisloff, U. Maximal strength training improves work economy in trained female cross-country skiers. Med Sci Sports Exerc 1999: 31: 870-877.

Hopkins, W. G. A scale of magnitudes for effect statistics. I: A New View of Statistics. Internet Society for Sport Science 2000: http://www.sportsci.org/resource/stats/.

Hopkins, W. G. Spreadsheets for Analysis of Controlled Trials, with Adjustments for Subjects Characteristic. Sportscience 2006: 10: 46-50. http://www.sportsci.org/2006/wghcontrial.htm.

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 2009: 41: 3-13.

Hopkins, W. G., Schabort, E. J., & Hawley, J. A. Reliability of power in physical performance tests. Sports Med 2001: 31: 211-234.

Jones, A. M., Grassi, B., Christensen, P. M., Krustrup, P., Bangsbo, J., & Poole, D. C. Slow component of VO2 kinetics: mechanistic bases and practical applications. Med Sci Sports Exerc 2011: 43: 2046-2062.

Losnegard, T. Physiological determinants of performance in modern elite cross-country skiing (PhD thesis). Norwegian School of Sport Sciences, Oslo 2013.

Losnegard, T., & Hallén, J. Physiological differences between sprint- and distance-specialized cross-country skiers. Int J Sports Physiol Perform 2014: 9: 25-31.

Losnegard, T., Mikkelsen, K., Rønnestad, B. R., Hallén, J., Rud, B., & Raastad, T. The effect of heavy strength training on muscle mass and physical performance in elite cross country skiers. Scand J Med Sci Sports 2011: 21: 389-401.

Losnegard, T., Myklebust, H., & Hallén, J. Anaerobic capacity as a determinant of performance in sprint skiing. Med Sci Sports Exerc 2012: 44: 673-681.

Mikkola, J., Rusko, H. K., Nummela, A. T., Paavolainen, L. M., & Häkkinen, K. Concurrent endurance and explosive type strength training increases activation and fast force production of leg extensor muscles in endurance athletes. J Strength Cond Res 2007: 21: 613-620.

Moritani, T., & deVries, H. A. Neural factors versus hypertrophy in the time course of muscle strength gain. Am J Phys Med 1979: 58: 115-130.

Moss, B. M., Refsnes, P. E., Abildgaard, A., Nicolaysen, K., & Jensen, J. Effects of maximal effort strength training with different loads on dynamic strength, cross-sectional area, load-power and load-velocity relationships. Eur J Appl Physiol Occup Physiol 1997: 75: 193-199.

Raastad, T., Risøy, B. A., Benestad, H. B., Fjeld, J. G., & Hallén, J. Temporal relation between leukocyte accumulation in muscles and halted recovery 10-20 h after strength exercise. J Appl Physiol 2003: 95: 2503-2509.

Rietjens, G. J., Kuipers, H., Kester, A. D., & Keizer, H. A. Validation of a computerized metabolic measurement system (Oxycon-Pro) during low and high intensity exercise. Int J Sports Med 2001: 22: 291-294.

Rønnestad, B. R., Hansen, E. A., & Raastad, T. Strength training improves 5-min all-out performance following 185 min of cycling. Scand J Med Sci Sports 2011: 21: 250-259.

Rønnestad, B. R., Kojedal, O., Losnegard, T., Kvamme, B., & Raastad, T. Effect of heavy strength training on muscle thickness, strength, jump performance, and endurance performance in well-trained Nordic Combined athletes. Eur J Appl Physiol 2012: 112: 2341-2352.

Sandbakk, Ø., Ettema, G., & Holmberg, H. C. Gender differences in endurance performance by elite cross-country skiers are influenced by the contribution from poling. Scand J Med Sci Sports 2014: 24: 28-33.

Smith, G., Kvamme, B., & Jakobsen, V. Effectiveness of ski and pole forces in ski skating. In: Müller, E., Lindinger, S. & Stöggl, T. (Eds.), Science and skiing IV. Maidenhead, UK: Meyer & Meyer Sport, 2009: 647-656.

Stöggl, T., Lindinger, S., & Müller, E. Evaluation of an upper-body strength test for the cross-country skiing sprint. Med Sci Sports Exerc 2007: 39: 1160-1169.

Whipp, B. J., & Wasserman, K. Oxygen uptake kinetics for various intensities of constant-load work. J Appl Physiol 1972: 33: 351-356.
Østerås, H., Helgerud, J., & Hoff, J. Maximal strength-training effects on force-velocity and force-power relationships explain increases in aerobic performance in humans. Eur J Appl Physiol 2002: 88: 255-263.

Table 1. Experimental design.

Familiarization	x											x			
Double poling protocol 1		x												x	
Double poling protocol 2		x												x	
1RM strength test			x												x
VO_{2max} running test			x												x
		Pre-test					Intervention							Post-test	
		←→			←								→	←→	
Week	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12

RM, repetition maximum; VO_{2max}, maximal oxygen uptake.

Table 2. Strength training program for the intervention group.

Week	**1-4**	**5-7**	**8-10**
Day 1 (sets x repetitions)	3 x 10 RM	3 x 8 RM	3 x 6 RM
Day 2 (sets x repetitions)	3 x 6 RM	3 x 5 RM	3 x 4 RM

RM, repetition maximum.

Table 3. Training through the 10-week intervention period for both the intervention and control group expressed in hours per week and as percentage distribution.

	Intervention group (n = 9)			**Control group (n= 7)**		**Group comparison**
	Hours		**%**	**Hours**	**%**	**Cohen's *d* ES**
HIT (≥88% of HR_{max})	0.5 ± 0.2		4 ± 1	0.5 ± 0.2	4 ± 1	0.00-trivial
MIT (82-87% of HR_{max})	0.3 ± 0.2		2 ± 1	0.3 ± 0.1	2 ± 1	0.11-trivial
LIT (60-81% of HR_{max})	10.3 ± 1.6		75 ± 3	9.7 ± 1.3	76 ± 5	0.28-small
Total endurance training	11.2 ± 1.8		81 ± 3	10.6 ± 1.3	83 ± 4	0.27-small
Heavy strength training	1.1 ± 0.1		8 ± 1	-	-	-
Core stability training	0.9 ± 0.5		7 ± 3	1.2 ± 0.3	9 ± 2	-0.45-small
Muscular endurance training	0.1 ± 0.1	*	1 ± 0	0.3 ± 0.1	2 ± 1	-1.41-large
Other training	0.5 ± 0.2		4 ± 2	0.7 ± 0.4	5 ± 3	-0.48-small
Total training	13.8 ± 2.3		100	12.7 ± 1.4	100	0.34-small

Data are mean ± standard deviation; LIT, low intensity training; MIT, moderate intensity training; HIT, high intensity training; HR_{max}, maximal heart rate; Core stability training: sling- Swiss Ball- and mat exercises; Muscular endurance training: dips on bench and knee push-ups with a resistance that could be performed 25 times or more per set; ES, effect size; *Significantly different from the control group ($p<0.01$).

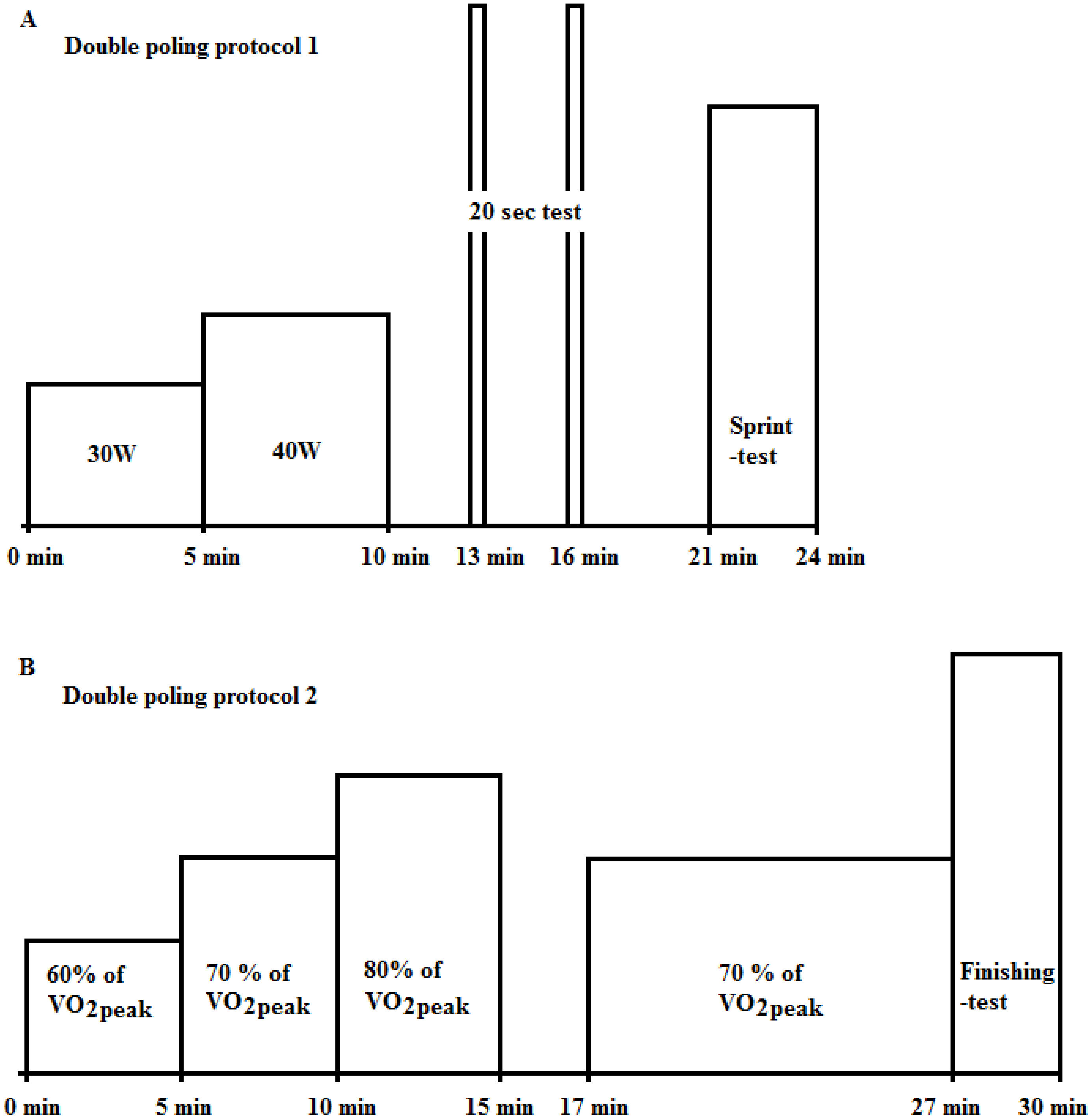

Fig. 1. Double poling protocol 1(A) and 2 (B).

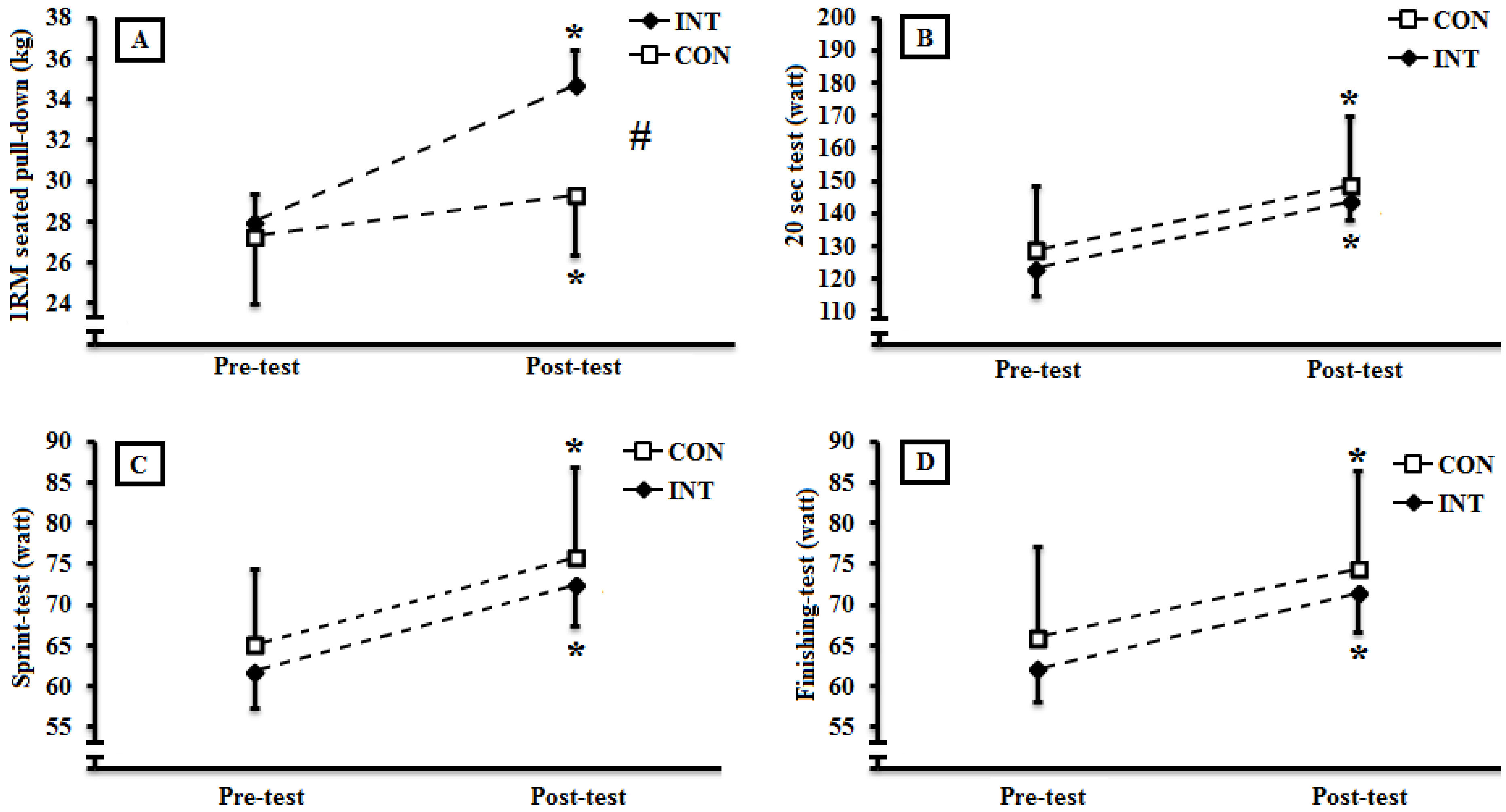

Fig. 2. The figure shows pre- and post-test results for 1RM strength in seated pull-down (A) and power output in the 20 sec test (B), the 3 min sprint-test (C) and the 3 min finishing-test (D). Data are expressed as group mean ± 90% confidence interval for both the intervention group (INT) and the control group (CON). * Significant change from pre- to post-test ($p<0.05$). # Significant difference in pre- to post-test change between INT and CON ($p<0.01$).

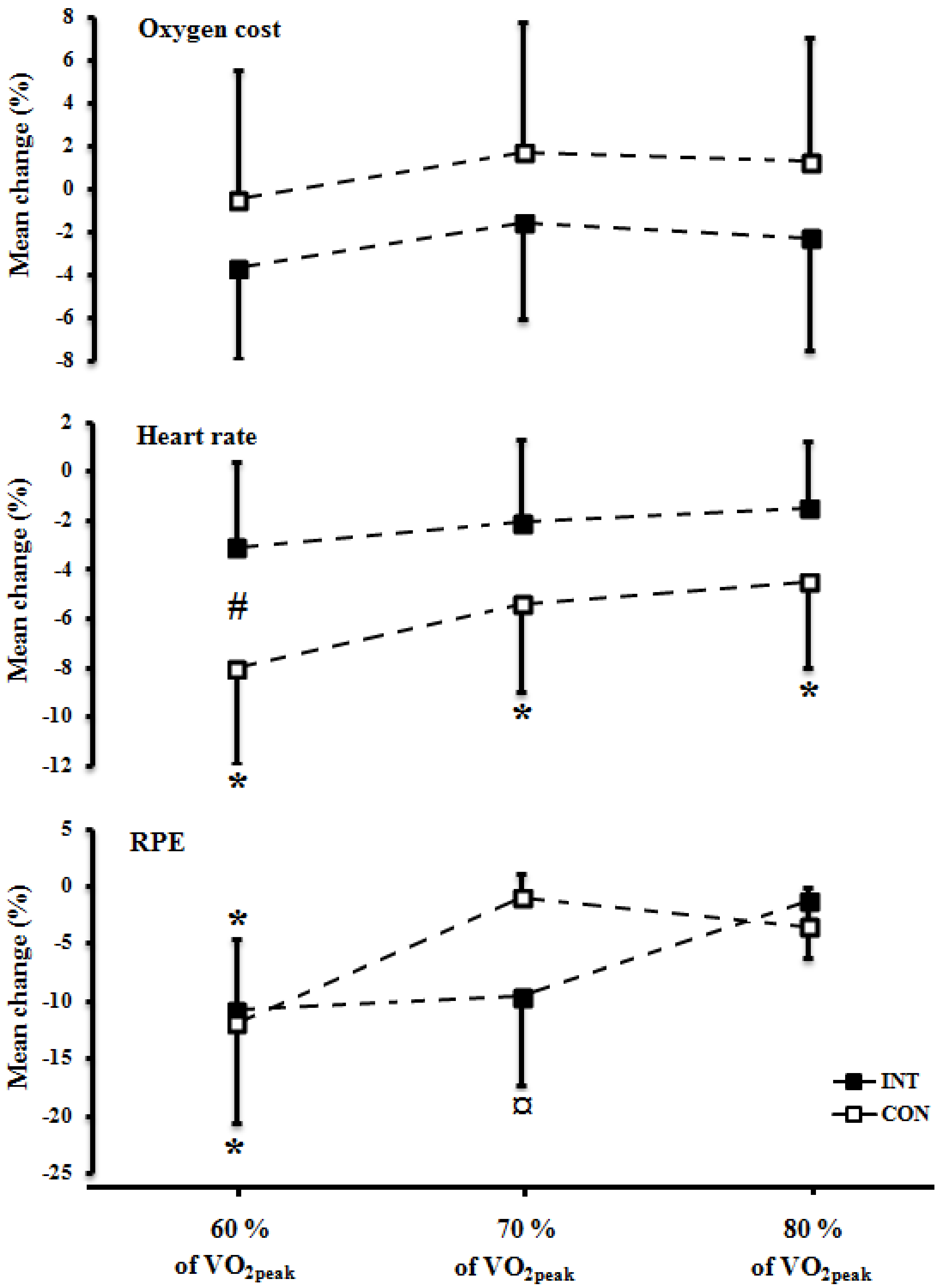

Fig. 3. Pre- to post-test changes in sub-maximal oxygen cost (L · min^{-1}), heart rate (beats · min^{-1}) and rating of perceived exertion (RPE) at 60, 70 and 80% of VO_{2peak}. Data are expressed as group mean ± 90% confidence interval. */¤ Significant change from pre- to post-test ($p<0.05$/ $p<0.1$). # Significant difference in change from pre- to post-test between groups ($p<0.1$).

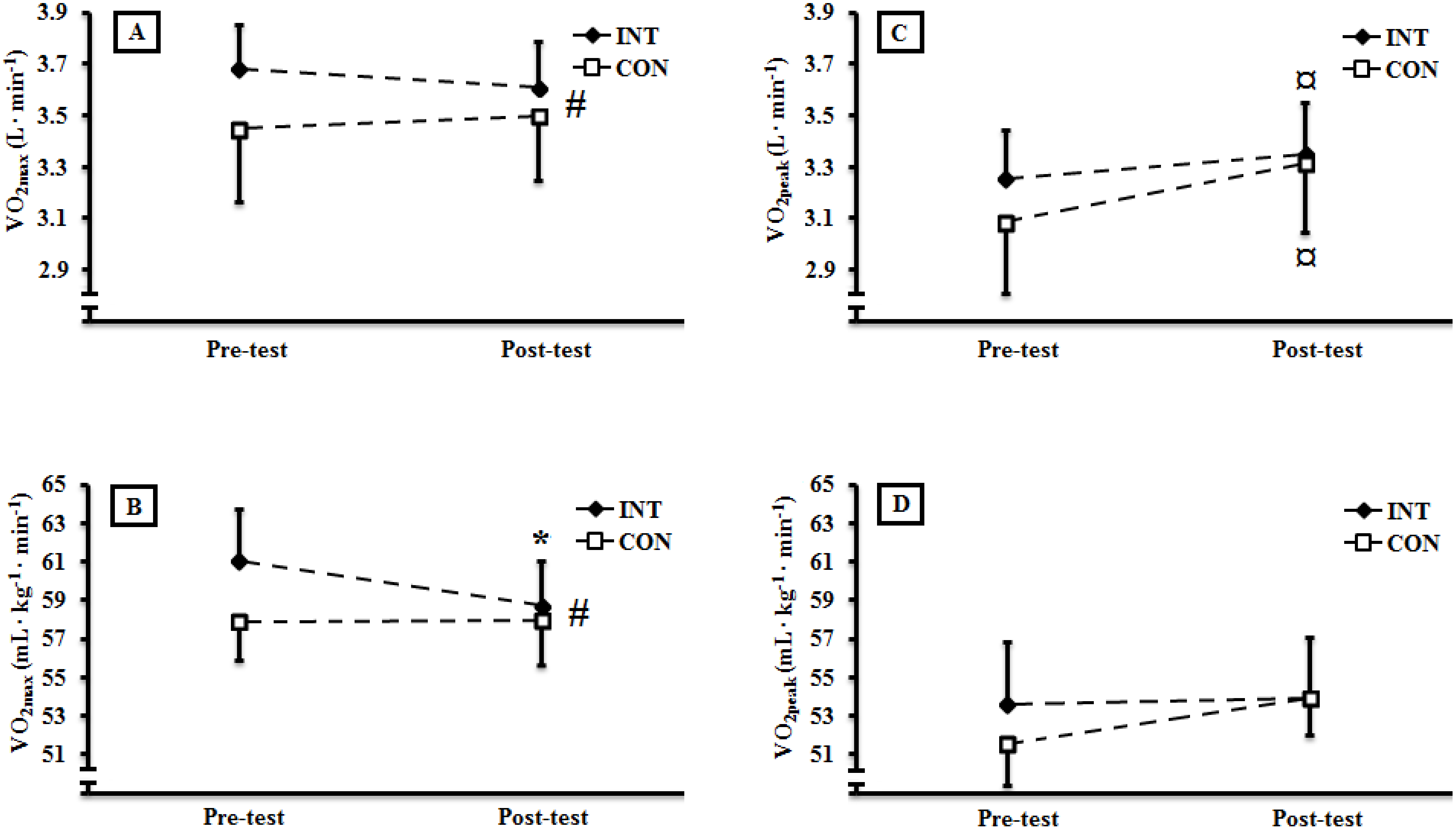

Fig. 4. VO_{2max} running is shown in A ($L \cdot min^{-1}$) and B ($mL \cdot kg^{-1} \cdot min^{-1}$) and the highest VO_{2peak} double poling measured in either the sprint-test or the finishing-test is shown in C ($L \cdot min^{-1}$) and D ($mL \cdot kg^{-1} \cdot min^{-1}$). The data are expressed as group mean ± 90% confidence interval for both the intervention group (INT) and the control group (CON) at pre- and post-test. */¤ Significant change from pre- to post-test ($p<0.05$/ $p<0.1$). # Significant difference in pre- to post-test change between INT and CON ($p \leq 0.1$).

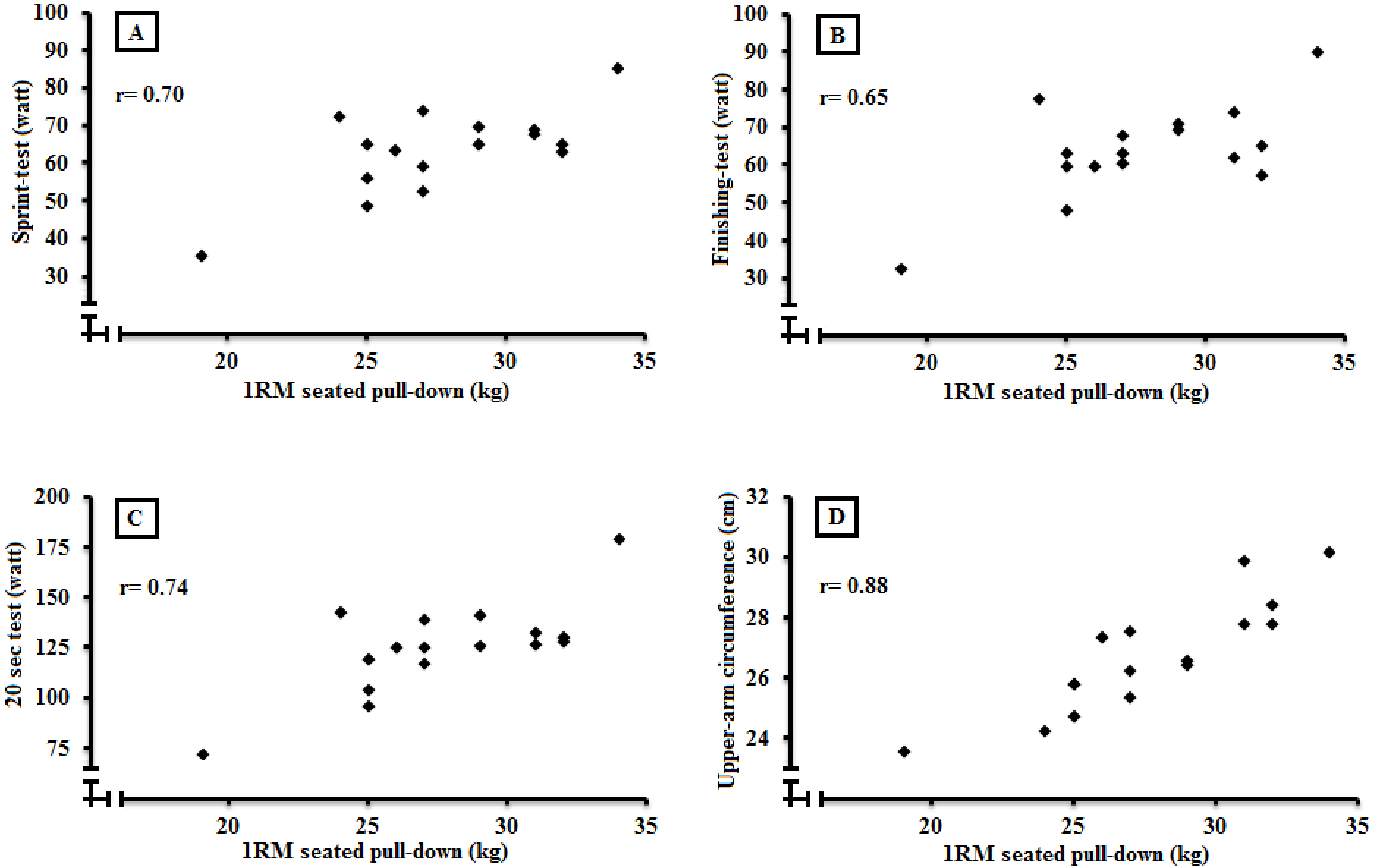

Fig. 5. Plots A-C show the bivariate relationship between 1RM strength seated pull-down and power output in the sprint-test (A), the finishing-test (B) and the 20 sec test (C). Plot D shows the relationship between 1RM strength and upper-arm circumference. All data were collected at pre-test.