Lilli Marie Myhre Ofstad

Comparison of test-retest reliability of peak oxygen uptake during different upper-body poling protocols

Master's thesis in Human Movement Science

Trondheim, June 2016

Norwegian University of Science and Technology

Faculty of Medicine

Department of Neuroscience (INM)

Abstract

Background and purpose: Upper body poling is a sport-specific exercise mode for several Paralympic sports such as Nordic sit ski and ice sledge hockey. However, there is a lack of reliable and standardized tests to evaluate changes in physical fitness, such as peak oxygen uptake (VO_{2peak}), in this mode. The primary purpose of this study was therefore to examine the test-retest reliability of VO_{2peak} and the corresponding physiological variables in three different test-protocols in upper body poling. The secondary purpose was to investigate which protocol resulted in the highest VO_{2peak}. **Methods:** 23 healthy upper-body trained men participated in the study, which involved two days of testing (3.9±3.3 days apart). Body composition was assessed with an InBody scale. On both test days, subjects performed the three different test protocols to exhaustion; a 1-min and a 3-min self-paced all-out test as well as an incremental test to failure. The test order was randomized across subjects, but the same order was applied both days. The subjects were placed in front of a poling ergometer in a seated position and with feet locked. Variables included in the study were $VO_{2peak,}$ peak heart rate (HR_{peak}), peak blood lactate concentration (BLa_{peak}), muscular-, respiratory- and over-all rating of perceived exertion (RPE) and power output. Statistical significance was set as $p \leq 0.05$. Reliability was determined by intraclass correlation coefficient (ICC) and coefficient of variation (CV). Limits of agreement (LOA) were also investigated. **Results:** A high ICC (0.938-0.950) and low CV (0.4-1.2 %) between test and retest show a high reliability in VO_{2peak} withtin all three test protocols. There were no significant differences in VO_{2peak}, RER_{peak}, HR_{peak}, respiratory RPE, over-all RPE or BLa_{peak} between test and retest in the 1-min test, 3-min test or the incremental test (all $p<0.005$). A significantly higher power output in the retest for all three protocols were found ($p<0.05$), as well as a significantly higher V_{Epeak} in the retest for the incremental protocol ($p<0.05$) and a lower muscular RPE in the retest for the 1-min protocol ($p<0.05$). Subjects reached 5.5 % ($p <0.05$) and 7.9 %($p< 0.05$) higher VO_{2peak} in the 3-min protocol and the incremental test compared to the 1-min test, respectively. The incremental protocol resulted in a 4.6 % higher VO_{2peak} than the 3-min test ($p< 0.05$) **Conclusion:** The 1-min protocol, 3-min protocol and the incremental test do all have a high reliability when it comes to VO_{2peak}, RER_{peak}, HR_{peak}, respiratory RPE, over-all RPE and $BLa_{peak.}$ A significant difference in power output might be due to a systematic error in terms of a learning effect. However, all three test protocols were found reliable for testing of VO_{2peak} in upper-body poling. The highest VO_{2peak}

were obtained with the incremental protocol, and should be the recommended test for detection of VO_{2peak} in upper-body poling.

Sammendrag

Bakgrunn og mål med studien: Overkroppsstaking er en sportsspesifikk øvelse for flere paralympiske idretter, blant annet kjelkehockey og skipigging. Det finnes derimot ingen reliable og standardiserte tester for å evaluere fysisk kapasitet og peak oksygen opptak (VO_{2peak}) i denne bevegelsesformen. Hovedmålet med studien var derfor å undersøke test-retest reliabiliteten av VO_{2peak} og de korresponderende fysiologiske variablene i tre ulike testprotokoller i overkroppsstaking. Sekundært ønsket vi å undersøke hvilken av de tre protokollene som ga de høyeste VO_{2peak}-verdier. **Metode:** 23 friske, overkroppstrente menn deltok i studien, som involverte to dager med testing (3.9±3.3 dagers mellomrom). Kroppssammensetning ble målt med en InBody vekt. Begge dagene ble subjektene testet i de tre ulike protokollene, utført til utmattelse. De respektive protokollene var en 1-min test, 3-min test og en inkrementell test. Rekkefølgen på gjennomføring av protokollene ble randomisert mellom subjektene, hvor den samme rekkefølgen gjaldt for både test og retest. Subjektene ble plassert foran et stakeergometer i en sittende posisjon med beina låst. Variabler som ble inkludert i studien var VO_{2peak}, peak heart rate (HR_{peak}), peak blood lactate (BLa_{peak}), muskulær-, respiratorisk- og over-all rating of perceived exertion (RPE) og power output. Reliabilitet ble fastslått med intraclass correlation coefficient (ICC) og coefficient of variation (CV). Limits of agreement (LOA) ble også undersøkt. Statistisk signifikans var satt til $p \leq 0.05$. **Resultat:** En høy ICC (0.938-0.950) og lav CV (0.4-1.2 %) mellom test og retest viser en høy reliabilitet i VO_{2peak} i alle tre protokollene. Det var ingen signifikante forskjeller i VO_{2peak}, RER_{peak}, HR_{peak}, respiratorisk RPE, over-all RPE eller BLa_{peak} i 1-min testen, 3-min testen eller den inkrementelle testen. En siginifikant høyere power output i retest ble funnet i alle tre protokollene($p<0.05$), samt høyere V_{Epeak} i den inkrementelle testen ($p<0.05$) og lavere muskulær RPE i 1-min testen ($p<0.05$). Subjektene oppnådde henholdsvis 5.5%($p<0.05$) og 7.9 %($p<0.05$) høyere VO_{2peak} i 3-min protokollen og den inkrementelle protokollen sammenlignet med 1-min protokollen ($p<0.05$), og henholdsvis 4.6% høyere VO_{2peak} i den inkrementelle testen sammenlignet med 3-min testen ($p< 0.05$). **Konklusjon:** 1-min testen, 3-min testen og den inkrementelle testen har alle en høy reliabilitet når det kommer til VO_{2peak}, RER_{peak}, HR_{peak}, respiratorisk RPE, over-all RPE og BLa_{peak}. En signifikant høyere power output i retest kan skyldes en systematisk error i form av en læringseffekt. Alle tre protokollene er derfor funnet reliable for

testing av VO_{2peak} i overkroppsstaking. Den høyeste VO_{2peak}-verdien ble oppnådd med den inkrementelle protokollen, og bør være den anbefalte metoden for å teste VO_{2peak} i overkroppsstaking.

Acknowledgements

To PhD student Julia Baumgart for all help and cooperation on the project, hours spent together in the lab, patience and good advice

To my supervisor Dr. Øyvind Sandbakk for letting me have this opportunity, all moral- and scientific support, 24/7 service and positive appearance

To my co-supervisor Prof. Gertjan Ettema

To PhD student and handyman Knut Skovereng for all help and support all year through

To the guy next door, Post Doc. David McGhie for always keeping your door open

To all the great people at Olympiatoppen Midt-Norge, Centre for Elite Sports Reaserch (SenTIF) and the rest in Granåsen Toppidrettssenter for making this year a joy, and every day at the office spent in gratitude.

To all subjects volunteering for the study, your time and effort, blood, sweat, good spirit and sportsmanship

Thank you!

Table of contents

1. Introduction

Sport specific testing is an important tool for evaluation of the effect of training and physical fitness in sports, and standardized test-protocols are developed in several exercise modes, as cycling and running, to assure reliable and optimal test methods (ACSM, 2006). Upper body poling is a sport specific exercise mode for several Paralympic sports such as Nordic sit ski and ice sledge hockey, with requirement of high aerobic capacity in upper-body. However, this mode lacks reliable and standardized tests to evaluate changes in physical fitness, such as peak oxygen uptake (VO_{2peak}). Reliability studies of VO_{2peak}- protocols are frequent in whole- and lower body exercises as running (Kirkeberg, Dalleck, Kamphoff, & Pettitt, 2011; Peyer, Pivarnik, & Coe, 2011), cycling (Andersen, 1995; Burnley, Doust, & Vanhatalo, 2006)and roller skiing (Verges, Flore, Laplaud, Guinot, & Favre-Juvin, 2006). The reliability of VO_{2peak}-protocols is also investigated in upper body modes as arm cycling, arm cranking, where the reliability is found similar to what observed in whole-and lower body exercises (Flueck, Lienert, Schaufelberger, & Perret, 2015; Leicht, Sealey, & Sinclair, 2009).When testing athletes one wish to test in a close to competition-like mode as possible, in order to assess the muscles and physiological factors involved in the given sport. During testing both physiological variables, such as peak VO_{2peak}, peak minute ventilation (V_{Epeak}) and peak heart rate (HR_{peak}), as well as subjective parameters as muscular, cardio-respiratory and overall rate of perceived exertion (RPE) (Borg, 1998) can be assessed.

Studies that have investigated the reliability of VO_{2peak} protocols in upper-body modes have applied various protocols, but there is no consensus of which are the most valid and reliable one. Bar -Or and Zwiren tested the reliability in an incremental test of 25 W increases every 2 min in arm cranking, and found high test- retest reliability (ICC 0.94) (Bar-Or & Zwiren, 1975). Holmberg and Nilsson performed a reliability test on a 6-min all-out protocol in a standing poling ergometry, where they found a high reliability (CV 2.4 %) (Holmberg & Nilsson, 2008), with test-retest reliability values that are similar to what observed in other activities as treadmill running(Fielding, Frontera, Hughes, Fisher, & Evans, 1997; Figueroa-Colon et al., 2000) (Pivarnik, Dwyer, & Lauderdale, 1996) and cycling (Andersen, 1995).

Only a few studies a have performed VO_{2peak}-tests in upper-body poling (Forbes, Chilibeck, Craven, & Bhambhani, 2010; Hegge et al., 2015). A study of the Canadian sit ski team, compared a double poling ergometer with field testing for evaluating VO_2, using a 3-min self-paced protocol.

The 3-min protocol were repeated three times with 1.5 min recovery between repetitions (Forbes et al., 2010), and the study concluded that the double poling ergometer were just as reliable as field testing to evaluate VO_{2peak} in elite cross-country sit-skiers. Hegge et al. utilized a 3-min self-paced protocol to compare the aerobic capacity in upper-body poling in women and men, and found values of 56.1±4.8 mL min^{-1} kg^{-1} in men and 41.7±3.5 mL min^{-1} kg^{-1} in women, respectively (Hegge et al., 2015). However, reliability was not investigated and in this study upper-body poling was defines as a standing position with feet locked, something that allowed a different muscle activation than what seen in the seated position in Paralympic sports.

The reliability, or reproducibility, is important when determining VO_{2peak}. High reliability allows a smaller sample size and gives a stronger confident that a real change has occurred when testing an individual repeated times (Midgley, McNaughton, & Carroll, 2007). There are no consensus for what level of reliability that should be clinical acceptable, yet there is a common agreement that no single reliability measures are valid alone, and that several measures should be applied together (Bruton, Conway, & Holgate, 2000). Baumgarter defined two types of reliability; absolute and relative reliability. The absolute reliability indicate how repeated measures differ for individuals, where less variability means a higher reliability. The relative reliability gives an indicator of to which level individuals remain their position in a group over repeated measurements (Baumgarter, 1989).

To ensure a reliable and valid test, reaching VO_{2max} requires that the subject meet several criteria. Traditionally a plateau in VO_2, despite an increase in workload traditionally has been the main criteria for VO_{2max} (Howley, Bassett, & Welch, 1995). However, later research have stated that this plateau does not occur in all individuals (Bassett & Howley, 2000; Day, Rossiter, Coats, Skasick, & Whipp, 2003), and that the absence of a plateau is more frequently observed during upper-body exercise (Smith, Amaral, Doherty, Price, & Jones, 2006). Therefore VO_{2peak} is a more correct term to use when referring to the highest achieved oxygen uptake instead of VO_{2max}, as it does not require any plateau to occur (Rossiter, Kowalchuk, & Whipp, 2006). In absence of the plateau, several other criteria are applied to state a peak value in whole- and lower-body exercise. As responses in upper body exercise differ from whole- and lower body exercise, other criteria is needed to determine the VO_{2peak} in upper-body mode. However, there is still a gap in literature in this field.

Due to the lack of reliable and standardized test-protocols to assess VO_{2peak} in upper-body poling, the primary purpose of this study was to examine the test-retest reliability of VO_{2peak} and the corresponding physiological variables using three different test-protocols in upper body poling. The secondary purpose was to investigate which protocol resulted in the highest VO_{2peak} values.

2. Methods

Participants

23 healthy, upper body trained men (28 ± 10 years, 77.3 ± 3.0 kg, 182± 7 cm), familiar to cross country skiing, or other endurance sports with requirement of high aerobic capacity in upper body volunteered to participate in the study. All participants answered a pre-screening questionnaire, to make sure they fitted the inclusion criteria (appendix B). The study included athletes at regional, national, recreational level in cross country skiing (n=22) and rowing (n=1), with training level of 37.7 ± 7.9 h/month (n=19). A written consent form was obtained from each participant before the data collection started, and the study was approved by the Regional Ethical Committee (REK). In advance of the testing, participants were informed about the testing procedure, and requested not to perform any extraneous training or consume alcohol 24 h before testing. In addition, they were told to avoid snus, caffeine and other substances on the day of the tests, and not to consume any food 2 h before testing. During testing the participants had an ad libitum intake of water and sports drink.

Experimental design

The study involved two days of testing, with 3.9 ± 3.3 days apart. Both days the participants performed a standardized warm-up and the same three randomly ordered test protocols performed to exhaustion. The three protocols performed in the study were a 1-min self-paced all-out test, a 3-min self-paced all-out test and an incremental test. On test day 1, a written consent was obtained from the subjects, and they were asked to fill out a questionnaire regarding sleep, nutrition, training, illness and injuries (Appendix A). Before the physical tests started, body composition and resting lactate were measured. Resistance of the ergometer was set at maximum (i.e.10) for all participants in all three protocols. The physical testing started with 2 min of familiarization to the set up at a very easy pace, followed by a standardized warm-up. The warm-up consisted of 4*5 min bouts of intensity corresponding to RPE 9, 11, 13 and 15 on the Borgs scale (Borg, 1998).

After the warm-up, subjects performed the three test protocols in a randomized order, where they had 15- 20 min of active break and a 5-min standardized warm-up in between each protocol. The 5-min warm-up were divided into 3 min at the same watt as produced in warm-up stage 3 (RPE

13), and the two last minutes at the same watt as produced in warm-up stage 4 (RPE 15). During the third minute the subjects were told to perform two submaximal sprints of 5 s, to get the muscles ready for maximal effort in the coming max test.

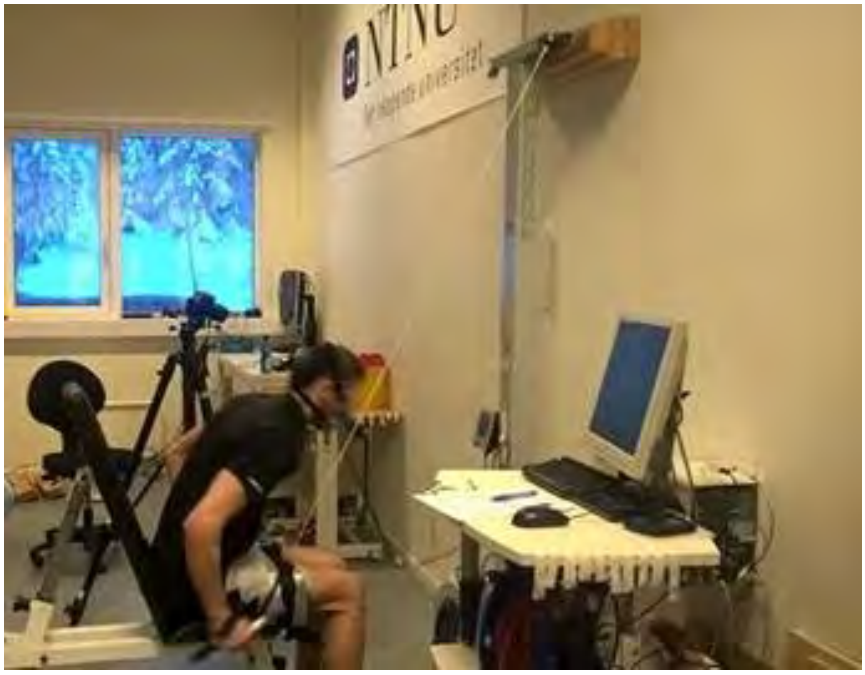
On test day 2, participants performed the same three protocols (retest) at the same time of the day, and in the same order as test day 1. The only difference from test day 1 were that the participants were given the watts from test day 1, and told to reproduce the same watt in the warm-up. They were asked to fill out the same questionnaire regarding sleep, nutrition, training, illness and injuries as test day 1.

Subjects were seated in front of a SkiErg with feet locked and strapped to the seating. Variables that were collected during testing were respiratory variables, BLa, HR, RPE and power output. VO_{2peak} and V_{Epeak} were determined by the highest value, after attenuating the variability from the raw data with 21-breath averages (Robergs, Dwyer, & Astorino, 2010). Lactate was determined as the highest of measurements 1- and 3 min after the max test.

Equipment

For the analysis of body composition an InBody scale (InBody 720, InBody, Seoul, Korea) were applied. Upper-body poling was performed on a modified Concept2 SkiErg apparatus (Morrisville, VT, USA), and subject were seated and strapped on to a modified strength bench in order to isolate lower body from the poling movement. The seating was set in a horizontal position, and the back of the seat were set in ~90 degrees. The two straps were placed over the pelvis and one on upper part of thigh, allowing free motion of upper body. Feet were placed in front of the bench, with a ~90-degree angle in knees. Distance between the seat and SkiErg were standardized, and were taken in consideration for the subjects not to hit their head in the SkiErg when going forward in the poling movement. Power output was monitored by the SkiErgs integrated software. Respiratory variables were measured with a breath-by-breath system, with the expired gas passing through the mixing chamber (Oxycon Pro apparatus, Jaeger GmbH, Hoechberg, Germany). An adult mask, medium size with dead space 96 was used. The system was calibrated against an ambient- and a commercial mixture of O_2 and CO_2. The O_2 and CO_2 contents were calibrated with a 3-L high precision calibration syringe (Calibration syringe D, SensorMedics, Yorba Linda, CA,

USA). Heart rate was monitored with a Polar M400 (Polar Electronics, Kempele, Finland) and blood lactate was measured with blood samples from the fingertip, analyzed with a Biosen C-Line Sport lactate assay system (EKF-diagnostic GmbH, Magdeburg, Germany).



Protocols

1-min all-out exercise test

For the 1-min test, the subjects were instructed to go out in a high a pace as possible they could maintain for one minute, while one researcher verbally encouraged the participants to use maximal effort. Definition of VO_{2peak} was the highest VO_2 measured during the 1-min period after analyzing the raw data. The 1-min protocol was an experimental design to investigate if a shorter and more time- and cost efficient test could be just as reliable as the two other protocols.

3-min all-out self-paced exercise test

For the 3-min test, subjects were told to go out in a high a pace as possible they could maintain for three minutes. One researcher verbally encouraged the participants to use maximal effort. Definition of VO_{2peak} was the highest VO_2 measured during the 3-min period after analyzing the raw data.

Incremental exercise test

The incremental test started at the same watt as produced in the third 5-min bout in the warm up (RPE = 13), rounded to the closest 5 or 10. Onward the subjects were told to increase intensity by 10 W every 30 seconds until exhaustion. When peak wattage was reached, the subjects were encouraged to maintain a high power as possible until either a plateau in the VO_2 occured, or until VO_2 significantly declined. One researcher instructed the subject during the increments and verbally encouraged the subject to use maximal effort.

Statistics

Statistical significance was set to $p \leq 0.05$. Data were checked for normality, and significant differences in physiological variables between test and retest were detected using a paired sampled t-test. Reliability for peak physiological variables were determined by intraclass correlation coefficient (ICC) (two way mixed) and coefficient of variation (CV) (Hopkins, 2000). Reliability of the ICC was interpreted with Munro's classification (Munro BH, 1986): 0.26-0.49 are associated with low correlation. 0.50-0.69 indicate a moderate correlation, 0.70-0.89 is associated with a high correlation, while 0.90-1.0 reflects a very high correlation.

Limits of agreement (LOA) were also investigated (Bland & Altman, 1986; Bruton et al., 2000). All statistical analysis was conducted with the Statistical Package for Social Science (SPSS, v14, Chicago, Illinois, USA).

3. Results

There was no significant difference in VO_{2peak} between test and retest in the 1-min test, 3-min test or the incremental test. Table 1 shows the group averages from test-, retest- and delta values in the respective protocols. All three tests had a significant difference in power output ($p<0.05$). In addition, there was a significant difference in muscular RPE in the 1-min test ($p<0.05$), and a significant difference in V_{Epeak} ($p<0.05$) for the incremental test. Calculated from an average of the VO_{2peak} of test and retest, the subjects reached 5.5 % and 7.9 % higher values in the 3 min (45.3 $ml \cdot kg^{-1} \cdot min^{-1}$) and incremental (46.4 $ml \cdot kg^{-1} \cdot min^{-1}$) test compared to the 1-min test (43.0 $ml \cdot kg^{-1} \cdot min^{-1}$), respectively ($p<0.05$ for all). The incremental test resulted in 4.7 % higher VO_{2peak}-values than the 3-min test ($p< 0.05$).

Table 1 Group averages for all variables. Test-, retest- and delta values of VO_{2peak}, V_{Epeak}, RER_{peak}, HR_{peak}, power output, rate of perceived exertion (RPE) and BLa_{peak} in three VO_{2peak}-protocols in upper-body poling (n=23). A 1-min self-paced all-out protocol, a 3-min self-paced all-out protocol and an incremental protocol.

	1-min all-out			**3-min all-out**			**Incremental**		
	Day 1	**Day 2**	**Delta**	**Day 1**	**Day 2**	**Delta**	**Day1**	**Day2**	**Delta**
VO_{2peak} ($ml \cdot kg^{-1} \cdot min^{-1}$)	42.5	43.4	0.9	45.3	43.3	-2.0	45.8	46.9	1.1
V_{Epeak} ($L \cdot min^{-1}$)	156	157	1	168	170	2	163	170	7*
RER_{peak} ($VCO_2 \cdot VO_2^{-1}$)	1.20	1.26	0.06	1.23	1.25	0.02	1.21	1.22	0.01
HR_{peak} (beats$\cdot$ min^{-1})	167	165	-2	171	170	-1	171	170	-1
W_{Avg}[1] (W)	232	245	13*	180	188	8*	194	200	6*
RPE_{oa}	18.1	17.9	-0.3	18.2	18.2	0.0	18.2	18.1	-0.1
RPE_{mu}	18.4	18.2	-0.2*	18.5	18.5	0.0	18.6	18.4	-0.2
RPE_{resp}	17.5	17.4	-0.1	17.70	17.80	0.10	17.63	17.76	0.13
BLa_{peak} ($mmol \cdot L^{-1}$)	10.9	10.9	0.0	11.5	11.8	0.3	11.3	11.9	0.6

[1]**Peak watt for Incremental test *Significant differences.** Peak oxygen uptake ($VO2_{peak}$). Peak ventilator equivalent (VE_{peak}). Peak respiratory exchange ratio (RER_{peak}). Peak heart rate (HR_{peak}). Average watt (W_{Avg}). Overall rate of perceived exertion (RPE_{oa}). Muscular rate of perceived exertion (RPE_{mu}). Respiratory rate of perceived exertion (RPE_{resp}). Peak blood lactate (BLa_{peak}).

There was a high reliability in VO_{2peak}, power output, HR_{peak} and BLa_{peak} in all three protocols, assessed with ICC and CV. Table 2 show high an ICC and low CV in all four variables within the 1-min test, 3-min test and the incremental tests.

Table 2 Intraclass correlation coefficient (ICC) and coefficient of variation (CV) of test-retest reliability of VO_{2peak}, power output, HR_{peak} and BLa in three VO_{2peak}-protocols upper-body poling (n=23). A 1-min self-paced all-out protocol, a 3-min self-paced all-out protocol and an incremental protocol.

	1-min		**3-min**		**Incremental**	
	ICC	**CV (%)**	**ICC**	**CV(%)**	**ICC**	**CV(%)**
VO_{2peak} ($ml \cdot kg^{-1} \cdot min^{-1}$)	0.952	0.74	0.959	0.41	0.938	1.2
WAvg[A] (W)	0.969	3.15	0.965	1.3	0.958	1.28
HR_{peak} ($beats \cdot min^{-1}$)	0.892	0.12	0.949	0.47	0.948	0.37
BLa ($mmol \cdot L^{-1}$)	0.851	2.38	0.933	1.57	0.755	1.95

[A]**PeakW for incremental test.** Peak oxygen uptake (VO_{2peak}). Average watt (W_{Avg}). Peak heart rate (HR_{peak}). Peak blood lactate (BLa).

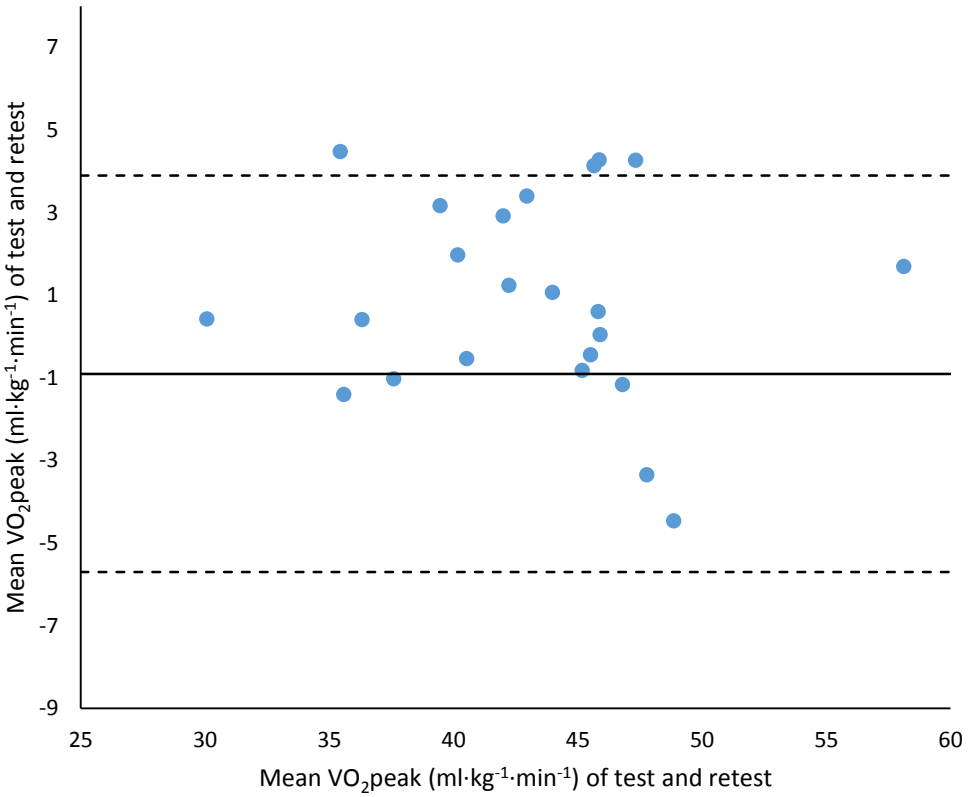


Fig. 1 Bland-Altman plot mean VO_{2peak} 1-min test. Y-axis show difference in VO_{2peak} between test and retest. Solid line show mean difference in VO_{2peak} between test and retest. Dotted lines show bias and limits of agreement (LOA).

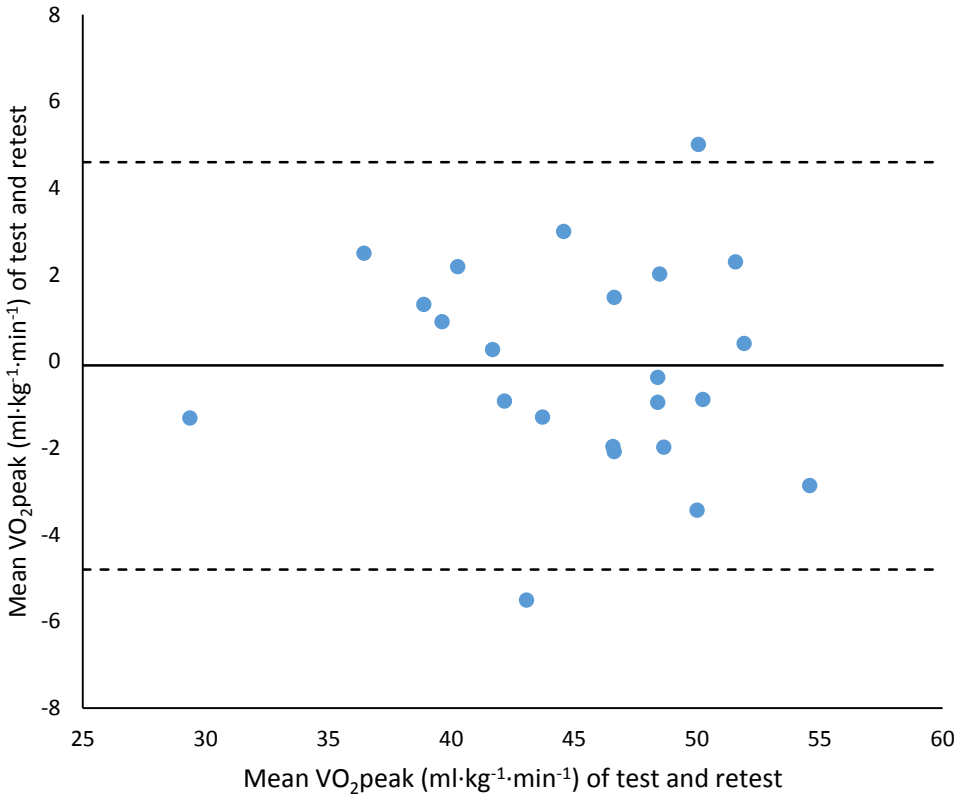


Fig. Bland-Altman plot mean VO_{2peak} 3-min test. Y-axis show difference in VO_{2peak} between test and retest. Solid line show mean difference in VO_{2peak} between test and retest. Dotted lines show bias and limits of agreement (LOA).

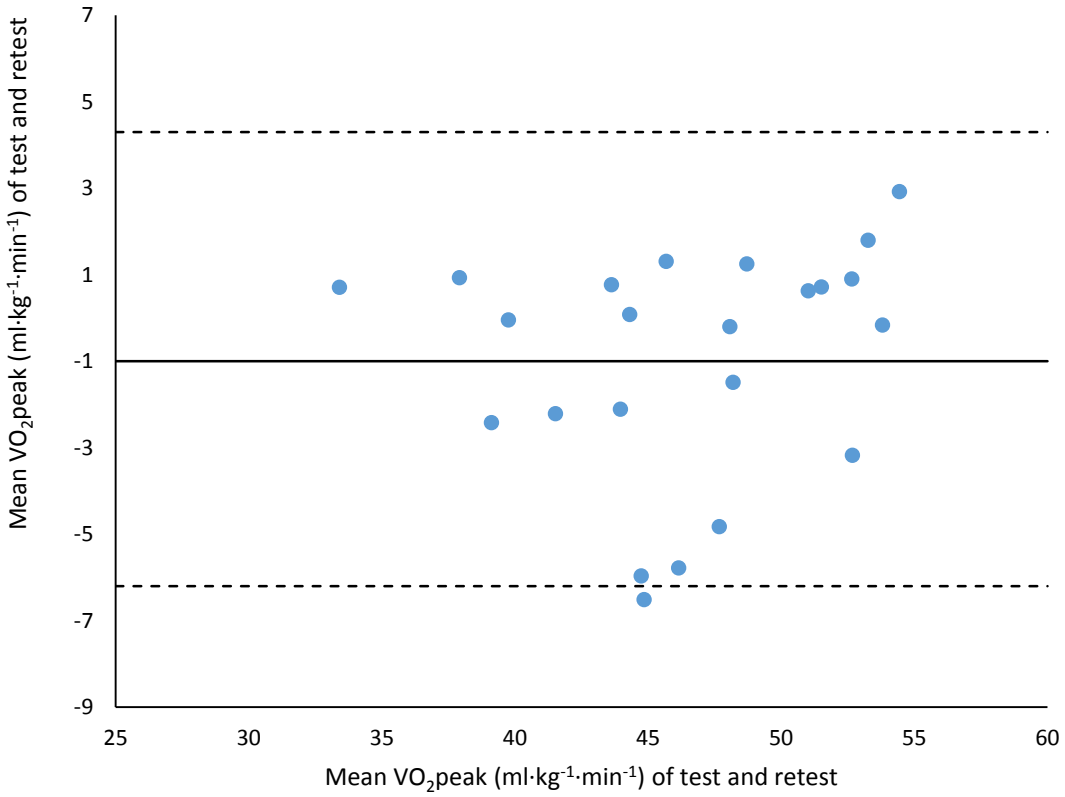


Fig. 3 Bland-Altman plot mean VO_{2peak} 1-min test. Y-axis show difference in VO_{2peak} between test and retest. Solid line show mean difference in VO_{2peak} between test and retest. Dotted lines show bias and limits of agreement (LOA).

4. Discussion

Until now there has been a lack of reliable and standardized test protocols to assess physical fitness in upper-body poling. Therefore, the aim of the present study was to investigate the test-retest reliability of VO_{2peak}, power output, RPE and physiological variables as BLa_{peak} and HR_{peak} between a 1-min all-out test, a 3-min all-out self-paced test and an incremental test, in addition to determine which of the three test protocols resulted in the highest VO_{2peak} values. The main findings were as following; 1) the results show no differences from pre- to post-test and a high test-retest reliability for VO_{2peak} (ICC 0.94-0.96, CV 0.4-1.2 %), HR_{peak} (ICC 0.90-0.95, CV 0.1-0.5 %) and BLa_{peak} (ICC 0.75-0.93, CV 1.6-2.4 %) within all three protocols, except for power output which systematically increased from pre- to post-test, and 2) the highest VO_{2peak} was obtained in the incremental test, followed by the 3-min test and the 1-min test.

The results show a high test-retest reliability, based on ICC and CV, for VO_{2peak} in this mode for all three protocols. This applied within all three protocols. In the 1-min test there was in addition high reliability in power output, though a slightly lower reliability seen in HR_{peak} and BLa_{peak}. The 3-min showed a high reliability in all examined variables; the power output, HR_{peak} and BLa_{peak}, while we found a high reliability in power output and HR_{peak}, though a lower reliability for the BLa_{peak} in the incremental test.

This finding shows that this test mode have a high reliability in all three protocols. This is in line with previous research on comparable protocols using other modes such a running, cycling, poling and arm crank (Bar-Or & Zwiren, 1975; Davis, Vodak, Wilmore, Vodak, & Kurtz, 1976; Fielding et al., 1997; Figueroa-Colon et al., 2000).

Despite the fact that there was a high reliability in power output in all three test, there was a significant difference in power output on test and retest. Even though subjects were pre-screened and familiar with upper-body training, none of them had earlier performed any isolated upper-body poling. A systematic error in terms of a learning effect, and a more efficient and economic movement might explain the higher watt values seen in all three retests. This can also be supported by the lower HR values seen in the retest for all three protocols, even though it was not a significant difference.

This indicate that even though it was a learning effect in power output, there were no difference in other physiological variables, something which indicate that the work-loads in all three protocols were appropriate to take out the physiological responses.

We found no differences in reliability between test protocols, by applying commonly used and recommended reliability measures (Atkinson & Nevill, 1998). The 1-min had an ICC of 0.95, the 3 min of 0.96 and the incremental an ICC of 0.94. Looking at the CV, difference between test were also low and relatively similar, ranging from 0.4 to 1.2 %. Though significant differences in power output from pre- to post-tests for all protocols, this does not seem to affect the VO_{2peak} values and reliability of the protocols.

It is agreed that one reliability measure alone is not sufficient to give the whole aspect of reliability, and that different reliability measures should apply together(Bruton et al., 2000). This study has applied both ICC and CV as estimates for reliability, as well as investigated the LOA. The CV is a measurement of absolute reliability, and the variation of repeated measurements within individuals, Limits of agreement were also investigated, and show the test-retest within participant variability (determined by 95% LOA). For VO_{2peak} tests, a CV and LOA within 3.8-8.5% is within the biological variation according to Katch et. al (Katch, Sady, & Freedson, 1982). It is debated which reliability estimates that are appropriate under given circumstances (Bruton et al., 2000), and some instances have suggested a need for a consensus in order to make a universal and standardized for reliability estimates (Rankin & Stokes, 1998).

The current study finds a high reproducibility in all three VO_{2peak} protocols, based up on several reliability measures (ICC, CV and LOA). As these three reliability measures conform, and based up on this strong results, we conclude that the 1-min protocol, the 3-min protocol and the incremental protocol are reliable.

The level of reliability found in this study also agrees with what found in previous studies on arm ergometry (Bar-Or & Zwiren, 1975; Smith et al., 2006) and treadmill running(Fielding et al., 1997; Figueroa-Colon et al., 2000), despite the fact that different protocols were applied.

The highest VO_{2peak} values were obtained with the incremental test, followed by the 3-min protocol and the 1 min-protocol, in which the incremental and 3-min test resulted in 7.9% and 5.5% higher

values than the 1-min respectively. When testing for reliability there will always be a measurement error, as the true peak value is impossible to detect (Bruton et al., 2000). By assessing VO_{2peak} with the same reliable protocol consequently, one will be able to monitor changes in physical fitness, but will not necessary detect the real VO_{2peak} and physical fitness. Despite the fact that all three test protocols were reliable, one wish to assess the VO_2 closest to the real $VO_{2peak.}$ Therefor it is important to know which protocol to apply, in order to obtain the highest possible value.

There are no existing recommendations towards duration in VO_{2peak} tests in upper-body, but from the results in the present study and in literature, it is likely that a certain duration should be recommended, as we the highest VO_{2peak} were obtained with the incremental test. The lower values seen in the 1-min test can be related to the shorter time at high intensity but also to a higher power output, and hence a higher anaerobic energy contribution and muscle tension, reducing the peripheral oxygen transport and limit the VO_{2peak} (Petrofsky & Hendershot). Comparing the 3-min and incremental test, a high power output and muscle tension that limits the VO_{2peak} might also be the reason why higher values were seen in the incremental test. This is supported by Leicht et al. , which recommended a longer lasting incremental protocol to avoid premature muscular fatigue before reaching the VO_{2peak} (Leicht et al., 2009).

Reliable tests in upper-body poling might also be relevant for sport disciplines where a upper-body poling movement plays a central role in the performance, such as cross country skiing, where upper-body capacity have a high correlation with performance (Mygind, Larsson, & Klausen, 1991; Watts et al., 1993).

Methodological considerations

For subjects that are novice to this exercise mode, it can be hard to choose an appropriate power output in the self-paced test, especially on the first day of testing. Ideally, subjects should have a familiarization session to all three protocols, in order to avoid a learning effect.

Despite the fact that most subjects in this study were cross country skiers, they were not necessary a homogenous group in terms of training and training distribution. It is likely to think that subjects who have had more focus on upper body training have a better pacing strategy as they are known with the responses and earlier onset of fatigue that occur with upper body exercise.

We regarded 15-20 minutes of rest between the three VO_{2peak}-tests to be sufficient to recover and the order of tests were evenly distributed and randomized within the participants in order to prevent any biases due to this. Although Hall-Lopez et al. tested the reliability of maximal treadmill test with only 10 min recovery between, which did not affect the reliability (Hall-Lopez et al., 2015),this study was performed in a lower-body mode, and recovery time might not necessary be the same in upper-body mode.

The method for attenuating the variability in the raw data from the breath-by-breath measurements could be a factor that affects the VO_{2peak} values. Time averaging for set periods as 15, 30 or 60 seconds or averages of a set number for breaths is the most common methods to approach this (Robergs et al., 2010). But though choice of time-averaging method for the breath-by-breath measurements might affect values, Midgley et al. concluded that reproducibility is not affected by the length of the VO_2 time-average interval (Midgley et al., 2007).

Conclusion

This study finds a high reproducibility in VO_{2peak}, HR_{peak}, BLa_{peak} and power output for all the three protocols; the 1-min all-out self-paced, 3-min self-paced and the incremental test. There were no significant differences from pre- to post-test in these, except a significantly higher average watt and peak watt in the retest for all tests, most likely due to a learning effect. Overall, this indicates that all these tests can be used in various experiments and training evaluations with high reproducibility. The highest VO_{2peak} measurements were obtained with the incremental test protocol, suggesting that, together with the high reliability, the incremental test could be recommended as a standardized protocol for upper-body poling testing. The findings in this study might contribute to improve training- and testing methods in several Paralympic sports, as well as in cross country skiing. The literature in this field is still scarce, and more research is needed to strengthen these findings. For further research, the reliability of these protocols in Paralympic athletes with various physical impairments and autonomic dysfunction would be of high relevance and interest for the field.

References

ACSM. (2006). ACSM's Guidelines for Exercise Testing and Prescription *American College of Sports Medicine. ACSM's Guidelines for Exercise Testing and Prescription* (pp. 99-102). Philadelphia: Lippincott, Williams & Wilkins.

Andersen, L. B. (1995). A maximal cycle exercise protocol to predict maximal oxygen uptake. *Scand J Med Sci Sports, 5*(3), 143-146.

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med, 26*(4), 217-238.

Bar-Or, O., & Zwiren, L. D. (1975). Maximal oxygen consumption test during arm exercise--reliability and validity. *J Appl Physiol, 38*(3), 424-426.

Bassett, D. R., Jr., & Howley, E. T. (2000). Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med Sci Sports Exerc, 32*(1), 70-84.

Baumgarter, E. R. (1989). Norm-referenced measurement: Reliability. In M. W. Safrit, TM. (Ed.), *Measurements Concepts in Physical Education and Exercise Science* (pp. 45-72): Champaign, IL.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*(8476), 307-310.

Borg, G. (1998). *Borg`s Perceived Exertion and Pain Scales*. Leeds, UK.

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy, 86*(2), 94-99.

Burnley, M., Doust, J. H., & Vanhatalo, A. (2006). A 3-min all-out test to determine peak oxygen uptake and the maximal steady state. *Med Sci Sports Exerc, 38*(11), 1995-2003. doi:10.1249/01.mss.0000232024.06114.a6

Davis, J. A., Vodak, P., Wilmore, J. H., Vodak, J., & Kurtz, P. (1976). Anaerobic threshold and maximal aerobic power for three modes of exercise. *J Appl Physiol, 41*(4), 544-550.

Day, J. R., Rossiter, H. B., Coats, E. M., Skasick, A., & Whipp, B. J. (2003). The maximally attainable VO2 during exercise in humans: the peak vs. maximum issue. *J Appl Physiol (1985), 95*(5), 1901-1907. doi:10.1152/japplphysiol.00024.2003

Fielding, R. A., Frontera, W. R., Hughes, V. A., Fisher, E. C., & Evans, W. J. (1997). The reproducibility of the Bruce protocol exercise test for the determination of aerobic capacity in older women. *Med Sci Sports Exerc, 29*(8), 1109-1113.

Figueroa-Colon, R., Hunter, G. R., Mayo, M. S., Aldridge, R. A., Goran, M. I., & Weinsier, R. L. (2000). Reliability of treadmill measures and criteria to determine VO2max in prepubertal girls. *Med Sci Sports Exerc, 32*(4), 865-869.

Flueck, J. L., Lienert, M., Schaufelberger, F., & Perret, C. (2015). Reliability of a 3-min all-out Arm Crank Ergometer Exercise Test. *Int J Sports Med, 36*(10), 809-813. doi:10.1055/s-0035-1548811

Forbes, S. C., Chilibeck, P. D., Craven, B., & Bhambhani, Y. (2010). Comparison of a double poling ergometer and field test for elite cross country sit skiers. *N Am J Sports Phys Ther, 5*(2), 40-46.

Hall-Lopez, J. A., Ochoa-Martinez, P. Y., Moncada-Jimenez, J., Ocampo Mendez, M. A., Martinez Garcia, I., & Martinez Garcia, M. A. (2015). Reliability of the maximal oxygen uptake following two consecutive trials by indirect calorimetry. *Nutr Hosp, 31*(4), 1726-1732. doi:10.3305/nh.2015.31.4.8487

Hegge, A. M., Bucher, E., Ettema, G., Faude, O., Holmberg, H. C., & Sandbakk, O. (2015). Gender differences in power production, energetic capacity and efficiency of elite cross-country skiers during whole-body, upper-body, and arm poling. *Eur J Appl Physiol*. doi:10.1007/s00421-015-3281-y

Holmberg, H. C., & Nilsson, J. (2008). Reliability and validity of a new double poling ergometer for cross-country skiers. *J Sports Sci, 26*(2), 171-179. doi:10.1080/02640410701372685

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Med, 30*(1), 1-15.

Howley, E. T., Bassett, D. R., Jr., & Welch, H. G. (1995). Criteria for maximal oxygen uptake: review and commentary. *Med Sci Sports Exerc, 27*(9), 1292-1301.

Katch, V. L., Sady, S. S., & Freedson, P. (1982). Biological variability in maximum aerobic power. *Med Sci Sports Exerc, 14*(1), 21-25.

Kirkeberg, J. M., Dalleck, L. C., Kamphoff, C. S., & Pettitt, R. W. (2011). Validity of 3 protocols for verifying VO2 max. *Int J Sports Med, 32*(4), 266-270. doi:10.1055/s-0030-1269914

Leicht, A. S., Sealey, R. M., & Sinclair, W. H. (2009). The reliability of VO2(peak) determination in healthy females during an incremental arm ergometry test. *Int J Sports Med, 30*(7), 509-515. doi:10.1055/s-0029-1202351

Midgley, A. W., McNaughton, L. R., & Carroll, S. (2007). Effect of the VO2 time-averaging interval on the reproducibility of VO2max in healthy athletic subjects. *Clin Physiol Funct Imaging, 27*(2), 122-125. doi:10.1111/j.1475-097X.2007.00725.x

Munro BH, V. M., Page EB (Ed.) (1986). *Statistical methods for health care research*. Philadelphia: JB Lippincott.

Mygind, E., Larsson, B., & Klausen, T. (1991). Evaluation of a specific test in cross-country skiing. *J Sports Sci, 9*(3), 249-257. doi:10.1080/02640419108729887

Petrofsky, J. S., & Hendershot, D. M. (1984). The interrelationship between blood pressure, intramuscular pressure, and isometric endurance in fast and slow twitch skeletal muscle in the cat. *Eur J Appl Physiol Occup Physiol, 53*(2), 106-111.

Peyer, K., Pivarnik, J. M., & Coe, D. P. (2011). The relationship among HRpeak, RERpeak, and VO2peak during treadmill testing in girls. *Res Q Exerc Sport, 82*(4), 685-692. doi:10.1080/02701367.2011.10599805

Pivarnik, J. M., Dwyer, M. C., & Lauderdale, M. A. (1996). The reliability of aerobic capacity (VO2max) testing in adolescent girls. *Res Q Exerc Sport, 67*(3), 345-348. doi:10.1080/02701367.1996.10607962

Rankin, G., & Stokes, M. (1998). Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil, 12*(3), 187-199.

Robergs, R. A., Dwyer, D., & Astorino, T. (2010). Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Med, 40*(2), 95-111. doi:10.2165/11319670-000000000-00000

Rossiter, H. B., Kowalchuk, J. M., & Whipp, B. J. (2006). A test to establish maximum O2 uptake despite no plateau in the O2 uptake response to ramp incremental exercise. *J Appl Physiol (1985), 100*(3), 764-770. doi:10.1152/japplphysiol.00932.2005

Smith, P. M., Amaral, I., Doherty, M., Price, M. J., & Jones, A. M. (2006). The influence of ramp rate on VO2peak and "excess" VO2 during arm crank ergometry. *Int J Sports Med, 27*(8), 610-616. doi:10.1055/s-2005-865857

Verges, S., Flore, P., Laplaud, D., Guinot, M., & Favre-Juvin, A. (2006). Laboratory running test vs. field roller skiing test in cross-country skiers: a longitudinal study. *Int J Sports Med, 27*(4), 307-313. doi:10.1055/s-2005-865664

Watts, P. B., Hoffman, M. D., Sulentic, J. E., Drobish, K. M., Gibbons, T. P., Newbury, V. S., . . . Clifford, P. S. (1993). Physiological responses to specific maximal exercise tests for cross-country skiing. *Can J Appl Physiol, 18*(4), 359-365.

Appendix A
- Spørreskjema deltakere prosjekt overkroppsstaking

Navn:

E-mail adresse:

Mobil nr:

Søvn

Hvor mange timer søvn hadde du I natt?

Hvor mange timer har du sovet i snitt de siste 5 nettene?

Kosthold

Hvor lenge siden er det du spiste?

Hva/ hvor mye spiste du?

Trening

Siste treningsøkt?

Siste hardøkt?

Hvor mange timer har du trent de siste uka?

Gjennomsnittlig treningstimer pr måned det siste året?

Rusmidler

Sist alkoholinntak?

Siste inntak av koffein

Har du drukket koffein i dag?

Sykdom

Har du vært syk eller hatt noen skader siste 4 uker?

Appendix B- Inklusjonskriterier

Mann ☐ Kvinne ☐

Alder: _____

Vekt: _____

Høyde: _____

Under svarer du kun dersom sykdom kan ha påvirkning på resultatet, eller om du er usikker på det.

Kroniske sykdommer/ tilstander: _____

Sykdommer siste 3 mnd.: _____

Virus/ Infeksjoner siste året: _____

Idrettsgren (er): _____

Totalt antall treningstimer pr uke: _____

Stakeøkter/ overkroppsøkter pr uke: _____

Hvilket nivå er du på som idrettsutøver?

Regionalt ☐

Nasjonalt ☐

Internasjonalt ☐

Turrenn ☐

Mosjonist ☐