

# Arbeidskrav 1 - reliabilitet og reproduserbar dataanalyse

## Introduksjon

Hensikten med denne studien er å finne ut hvor reproduserbar en  $\text{VO}_{2\text{maks}}$ -test på sykkel er. Dette er interessant fordi bedre reliabilitet betyr at vi kan stole enda mer på de resultatene vi får fra enkelttester. Da kan vi stole enda mer på de resultatene vi får når enkelttester i en fysiologisk testlabb brukes for å måle endringer over tid (Hopkins 2000).

Faktorer som påvirker reproduserbarheten av testen gjelder dagsvariasjoner hos forsøkspersonene, men også variasjoner i måleinstrumentene, samt hvordan vi instruerer underveis. Derfor var gjennomføringen av testene et sentralt fokus i gjennomføringen for å sikre et best mulig mål på reliabiliteten til testen. Basert på størrelsen på utvalget av testpersoner og vår erfaring med gjennomføring av fysiologisk tester, så visste vi på forhånd at vi måtte være forsiktige med å trekke konklusjoner om målevariasjonen til oksygenanalysatoren. Uansett får vi et svar på hvordan reproduserbarheten til testen er med de gitte premissene.

## Metode

### Forsøkspersonene

Syv mannlige deltaker ble rekruttert til prosjektet (alder =  $25.7 \pm 7$  år, vekt =  $75.7 \pm 10.8$  kg, høyde =  $181.3 \pm 6.7$  cm) (Table 1). Alle deltakerne trente regelmessig, men erfaring med trening på sykkel varierte innad i gruppa. Ingen hadde noe særlig erfaring med sykkeltestene vi gjennomførte.

Table 1: Dataene er presentert som gjennomsnitt  $\pm$  standardavvik.

Karakteristikker av forsøkspersonene basert på første test

Alder (år)	$25.7 \pm 3.5$
Høyde (cm)	$181.3 \pm 4.6$
Vekt (kg)	$75.7 \pm 7.4$
VO2maks	$66.2 \pm 7.2$
Wmaks	$410.0 \pm 58.9$

### Studiedesign

Prosjektets testdager bestod av fire dager, der halvparten av gruppa ble testet hver dag. Testdag 1 og 2 ble gjennomført som test 1 (t1), mens testdag 3 og 4 ble gjennomført som test 2 (t2). «Hviledagen» til forsøkspersonene bestod av rolig trening eller hvile. Dette var for å sikre at de var tilnærmet likt restituert før hver test. Testdagene prøvde vi å ha så identiske som mulig for alle deltakerne, i form av bruk av samme testleder på hver test (ett unntak pga. logistiske utfordringer for testleder), likt tidspunkt på døgnet  $\pm$  2 timer, ga dem beskjed om at siste måltid skulle være det samme og like lenge før test (Hopkins 2000).

### Kalibrering - klargjøring til test

Før hver test til forsøkspersonene kalibrerte vi Oxycon Pro. Vi sjekket at luftfuktigheten og temperaturen i rommet stemte overens med Oxycon sin estimering. Godkjenning av volum- og gasskalibrering var  $\pm$  1% for å minimere eventuelle feilmålinger. I tillegg ble Lode Excalibur-sykkelen innstilt likt ved t1 og t2.

### Testprosedyre

Deltakerne startet med en syv minutters lang oppvarming på ergometer sykkel, med en gradvis økning i intensitet. Intensiteten ble styrt etter Borgs 6-20 skala (Heath 1998). De syklet 3 min

tilsvarende opplevd anstrengelse på 11, 2 min på 13 og 2 min på 15/16 på Borgs 6-20 skala.

Del 1 styrketest: Etter oppvarming gjennomførte deltakerne en kort styrketest (knebøy power test). Den bestod av tre løft med 20 kg (oppvarming), og tre løft med henholdsvis 30-, 60- og 75 % av egen kroppsvekt. Målet var å gjennomføre løftet så hurtig som mulig, og kraftutviklingen ble målt med en muscle lab hastighetsmåler. Beste forsøk på hver belastning ble tellende. Vi går ikke noe mer inn på styrketesten, fordi det er på sykkeltesten vi har gjort analyser.

Del 2 sykkeltester: Deltakerne gikk direkte fra styrketesten til sykkeltestene. Her gjennomførte de en tredelt test, som først bestod av to submaksimale drag, deretter en  $VO_{2maks}$ -test og til slutt en Maximal Accumulated Oxygen Deficit (MAOD)-test. Grunnen til at vi kjørte de submaksimale dragene var for å estimere oksygenkrav på effekten (W) som ble syklet under MAOD-testen.

Under hele sykkeltesten prøvde vi å kjøre mest mulig lik tilbakemelding og engasjement hver gang. Det var lite tilbakemeldinger under de submaksimale dragene, og verbal oppmuntring under  $VO_{2maks}$ -testen, spesielt mot slutten. Vi ga ingen opplysninger om oksygenopptak ( $VO_2$ ) underveis, men de fikk vite effekten de syklet på. I tillegg til at de hele tiden kunne se tråkkfrekvensen og tid under alle testene. Oxycon målte gjennomsnittlig  $O_2$ -opptak hvert 30. sek som vi noterte ned (både på  $VO_{2maks}$ -testen og MAOD-testen), i tillegg til at vi noterte ned effekt, tråkkfrekvens (rpm), puls og varigheten på de to siste testene.

### **Submaksimale drag**

Den submaksimale testen bestod av to drag på fire minutter. For seks av deltakerne ble første belastning på testen gjennomført med en effekt på 100 W og andre belastning på 150 W, mens én deltaker syklet på 75 W og 125 W. Tilpasningene ble gjort for å få en mer optimal test. Forsøkspersonene syklet med neseclupe og munnstykke de siste to minuttene av hvert drag (begynte å ta i da det hadde gått 1,5 min), for bestemmelse av oksygenopptak på submaksimale belastninger. Deltakerne ble oppfordret til å holde en tråkkfrekvens på  $> 80$  rpm. Hver belastning ble gjennomført på samme måte, og gikk direkte over i hverandre. Vi spurte om Borgs 6-20 skala etter hvert drag. Etter dragene var det to minutter pause der deltakerne satt helt i ro. Tråkkfrekvensen til en bestemt deltaker ble reproduert på alle andre submaksimale drag og under MAOD-testen så lenge de klarte.

## **VO<sub>2maks</sub>-test**

VO<sub>2maks</sub>-testen startet for de fleste på 200 W, og økte med 25 W hvert minutt helt til utmattelse. For deltakeren som hadde litt lavere effekt på de submaksimale dragene startet VO<sub>2maks</sub>-testen på 150 W. Testen var ferdig når tråkkfrekvensen var < 60 rpm. Det var fri tråkkfrekvens og vi målte oksygenopptaket under hele testen. Vi spurte om Borgs-skala rett etter VO<sub>2maks</sub>-testen. Etter avsluttet test fikk forsøkspersonen fem minutter pause. Det første minuttet etter avsluttet test satt personen helt i ro, mens de neste fire minuttene ble gjennomført som rolig sykling på 50 W. Valgfri tråkkfrekvens, men den skulle være lik under pausen på t2.

## **MAOD-test**

MAOD-testens starteffekt baserte seg på VO<sub>2maks</sub>-testen. Effekten de startet på var den siste belastningen deltakeren syklet 30.sek eller mer på under VO<sub>2maks</sub>-testen. Den belastningen som ble brukt under t1 ble også brukt på t2 uavhengig av hvordan de presterte på VO<sub>2maks</sub>-test ved t2. Deltakerne syklet med neseklype og munnstykke under hele testen, og startet med “flying start” fra 50 W. Belastningen ble satt klart på maskinen, slik at den var klar når testleder ga beskjed om at testen skulle starte. Deltakerne syklet så lenge som mulig, og testen var over når tråkkfrekvensen var < 60 rpm. Vi spurte om Borgs 6-20 skala rett etter avsluttet test.

## **Datainnsamling og -behandling**

Etter å ha gjennomført testene samlet vi inn dataene vi skulle bruke for å gjøre statistiske analyser. Vi noterte ned VO<sub>2</sub> på de submaksimale dragene, og regnet ut VO<sub>2maks</sub> fra VO<sub>2maks</sub>-testen i Excel. Samtidig noterte vi ned andre variabler etter testen, slik som maksimal hjertefrekvens, gjennomsnittlig effekt fra siste minutt av VO<sub>2maks</sub>-testen, maksimal respiratory exchange rate (RER), maksimal pustefrekvens (BF), maksimal ventilasjon (VE), hvor lenge personen syklet, og hvilken effekt personen avsluttet på, samt opplevd anstrengelse (Heath 1998). Alle maksimale variabler fra Oxycon Pro regnet vi ut som gjennomsnittet av de to verdiene som var samtidig som de to høyeste etterfølgende VO<sub>2</sub>-målingene.

Etter MAOD-testen regnet vi ut VO<sub>2maks</sub>, oksygenkravet ved belastning under MAOD-test (L/min), det totale oksygenkravet som måtte dekkes (L), akkumulert oksygenopptak på testene (L), akkumulert oksyngjeld og prosent av arbeidet som ble dekket anaerobt (%). Samtidig

noterte vi ned hvor lenge personen syklet (i sekunder), maksimal hjerterefrekvens og opplevd anstrengelse (Heath 1998).

Alle data i resultatkapittelet er presentert som gjennomsnitt  $\pm$  standardavvik.

## Resultater

For relativ  $\text{VO}_{2\text{maks}}$  var differansen mellom testene  $0.62 \pm 2.41 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . Med en typisk målefeil på 1.7, og en variasjonskoeffisient (CV) på 2.6. I prosent var det  $0.94 \pm 3.65 \%$  forskjell.

For  $W_{\text{maks}}$  og ventilasjonen er differansen mellom testene henholdsvis  $5.71 \pm 13.80 \text{ W}$  og  $2.21 \pm 11.42 \text{ L/min}$ . Den typiske målefeilen var hhv. 9.8 og 8.1, mens CV var 2.4 og 4.

Gjennomsnittet ved t1 var ( $66.2 \pm 7.2$ ), og ved t2 var det ( $65.6 \pm 7.1$ ), fremstilt i Figure 1.

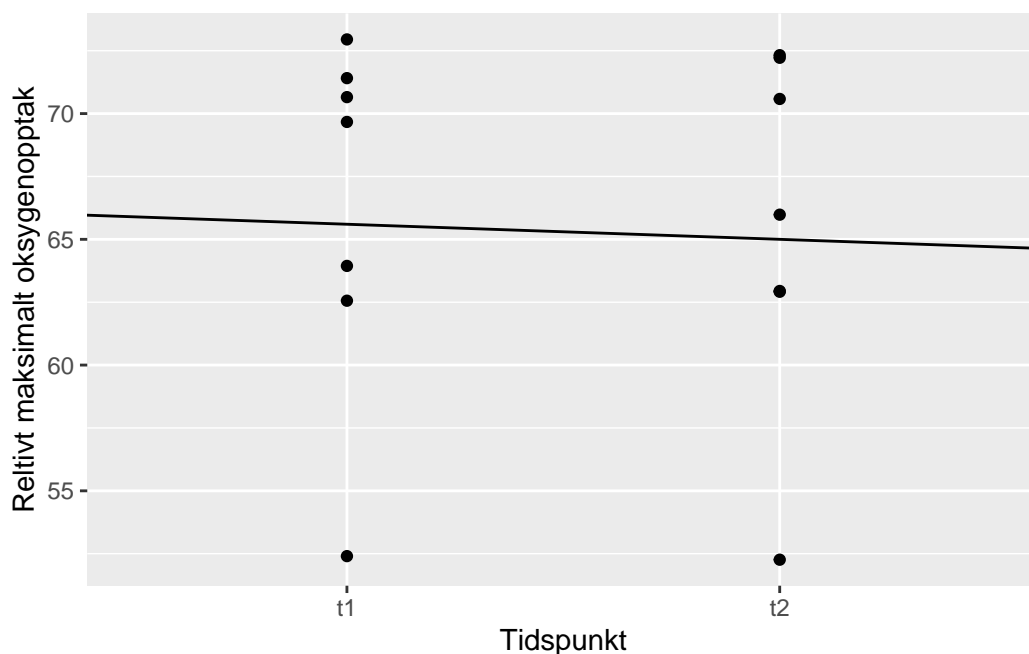


Figure 1: Relativt maksimalt oksygenopptak ved de to testene. Linjen fra fra gjennomsnittet fra første test (t1) til gjennomsnittet fra den andre (t2).

## Diskusjon

### Diskusjon av resultater

8 forsøkspersoner er et lite utvalg når vi skal måle reliabiliteten til en test (Hopkins 2000). Dette gjør at det er mer tilfeldigheter som kan påvirke resultatene våre. Ved første test var  $VO_{2maks}$  på  $66.2 \pm 7.2$  ml/kg/min, noe som gjør at de kan defineres som godt trente. Godt trente utøvere vil også produsere høyere effekt. Når både oksygenopptak og wattverdier er høyere vil det være desto mer beskrivende å se på CV som sier noe om prosentvis i forhold til gjennomsnittet. Den var på 2.6 og 2.4 for henholdsvis  $VO_{2maks}$  og  $W_{maks}$ . Med et lite utvalg blir det vanskelig å konkludere noe om reliabiliteten til testen, men det ser ut som vi har gjennomført testene godt med de feilkildene som nevnes i neste delkapittel.

### Tiltak for å sikre god reliabilitet

For å oppnå en størst mulig grad av validitet og reliabilitet er det nødvendig å ta stilling til ytre variabler som kan påvirke resultatet. Ved å ta hensyn til potensielle forstyrrende variabler reduseres risikoen for feilmålinger, og gjør funnene mer pålitelige (Israel Halperin and Martin 2015).

For å sikre god reliabilitet på de fysiologiske testene hadde vi flere tiltak for å redusere risikoen for forstyrrende variabler som kan påvirke resultatene. For det første ble begge testene gjennomført så likt som mulig, med en standardisert protokoll. Testene for hver forsøksperson ble avholdt på omtrent samme tidspunkt ( $\pm 2$  timer). Vi ga også beskjed om at de bare kunne trene rolig dagen før t1, og dagen i mellom testene. På den måten var det et tiltak for å sikre at de var restituert før begge testene. Deltakerne fikk også beskjed om at siste måltid før begge testene skulle være like, og til omtrent samme tidspunkt.

Deltakerne fikk beskjed om å ha samme tråkkfrekvensen ved første submaksimale belastningstrinn og ved MAOD-testen, og dette ble kopiert ved t2. Belastning og lengde på pausen før hver MAOD-test var lik ved begge testene. Det nevnte var svært viktig å ta høyde for, fordi oksygenopptaket varierer med ulik tråkkfrekvens (Gottshall, Bauer, and Fahrner 1996). Det var lik belastning og lengde på pause før hver MAOD-test begge dager uansett utfall på  $VO_{2maks}$ -testen. I tillegg ble hele sykkeltesten gjennomført sittende.

Vi valgte også å ha samme testleder for hver enkelt forsøksperson ved både t1 og t2. Den muntlige oppmuntringen og tilbakemeldingene underveis på testene var lik, og vi sørget for å gi like instruksjoner om utførelsen av testene og målet med hver test. Det ble også kjørt kalibrering av Oxycon pro før hver test, og vi satte godkjent kalibrering på volum til  $\pm 1 \%$ , og godkjent kalibrering av gass med en differanse på maksimalt  $\pm 1.0 \%$ . I vårt utvalg av forsøkspersoner, hadde alle gjennomført testing på et fysiologisk testlaboratorium før. Så de var kjent med å måle oksygenopptak, men det var liten eller ingen erfaring ved slik testing på sykkel. Dermed ble t1 mye læring for mange av personene, så for å sikre enda bedre reliabilitet hadde det vært viktig å la forsøkspersonene bli kjent med det å gjennomføre en sykkeltest på forhånd. Et punkt på dette som var usikkert for flere var hvordan de ulike belastningene føltes noe som kunne vært et enkelt tiltak å gjennomføre.

Studiens testledere hadde også en noe ulik erfaring ved å være testledere. Noe som førte til noen feil i gjennomføringen av protokollen. Dette var feil som forsøkspersonene ikke oppfattet underveis i deres tester, men som kan ha hatt innvirkning på testresultatet. Vi ser at ved senere studier vil det være hensiktsmessig for reliabiliteten til testene, at vi gjennomfører pilottesting, for å bli bedre kjent med gjennomføring. Dette vil i stor grad være det som i litteraturen blir referert til som tilfeldig forandring.

### Referanseliste:

- Gottshall, R., T. Bauer, and S. Fahrner. 1996. "Cycling Cadence Alters Exercise Hemodynamics." *International Journal of Sports Medicine* 17 (01): 17–21. <https://doi.org/10.1055/s-2007-972802>.
- Heath, Edward M. 1998. "Borg's Perceived Exertion and Pain Scales." *Medicine& Science in Sports & Exercise* 30 (9): 1461. <https://doi.org/10.1097/00005768-199809000-00018>.
- Hopkins, Will G. 2000. "Measures of Reliability in Sports Medicine and Science." *Sports Medicine* 30 (1): 1–15. <https://doi.org/10.2165/00007256-200030010-00001>.
- Israel Halperin, David B. Pyne, and David T. Martin. 2015. "Threats to Internal Validity in Exercise Science: A Review of Overlooked Confounding Variables." *International Journal of Sports Physiology and Performance* 10 (7): 823–29.