

Mappeeksamen

Hermann Moen

2021-12-01

Reliabilitet

Introduksjon

Maksimalt oksygenopptak VO₂max ble først beskrevet av Hill og Lupton i 1923, og kan defineres som kroppens evne til å ta opp og forbruke oksygen per tidsenhet (Bassett and Howley 2000; Hill and Lupton 1923). Innen toppidrett måles ofte det maksimale oksygenopptaket for å måle utøverens kapasitet opp mot arbeidskravet i den spesifikke idretten, og VO₂max kan i så måte også sees på som et mål på den aerobe effekten til utøveren (Bassett and Howley 2000). I Olympiatoppens testprotokoller benytter de flere definerte hjelperkriterier for å sikre at man faktisk har funnet deltakerens maksimale oksygenopptak (Tønnessen et al. 2017). Følgende kriterier er beskrevet; platå i O₂ er oppnådd, økning i ventilasjon med utflating av O₂ verdi, RER-verdi over 1.10 (1.05 om gjennomført laktatprofiltest i forkant) og blodlaktat over 8 (Tønnessen et al. 2017).

Metode

I forkant av testen målte alle deltakerne kroppsvekten i samme klær som ble brukt under testen, men ble bedt om å ta av seg skoene. Kroppsvekten som senere brukt i beregningen av maksimalt oksygenopptak ($\text{ml kg}^{-1} \text{ min}^{-1}$) er kroppsvekten målt i forkant av test, etter at 300g har blitt trukket av for å ta høyde for vekten av klærne. For å sikre intern validitet ble deltakerne bedt om å avstå fra anstrengende fysisk aktivitet dagen før test, standardisere måltidet i forkant av test samt avstå fra inntak av koffein under de siste 12 timene før testen (Halperin, Pyne, and Martin 2015). Pre- og post-tester ble gjennomført på samme tid på døgnet under standardiserte forhold.. Post-test ble gjennomført 6 dager etter gjennomført pretest. Det ble ikke kontrollert for fysisk aktivitet mellom testdagene.

Alle deltakerne gjennomførte en 10 minutter lang oppvarmingsprotokoll på tredemøllen (Woodway 4Front, Waukesha, USA), beskrevet for deltakerne i forkant av testen. Denne oppvarmingsprotokollen bestod av fem minutter på 11-13 i

Borg 6-20 RPE skala (Borg 1982), etterfulgt av 2x1min på starthastighet og stigning med 30 sekund mellom. Siste tre min var også 11-13 i borg. Etter oppvarming var det to min pause før testen begynte. Starthastighet for begge kjønn var satt til 8km/t, med stigning på 10.5% og 5.5% for henholdsvis menn og kvinner.

VO2max ble målt ved hjelp av en metabolsk analysator med miksekammer (vyntus CPX, mixingchamber (Vyntus CPX, Jaeger-CareFusion, UK)). Forut for alle tester ble analysatoren gass og volumkalibrert med en feilmargin på henholdsvis 2% og 0.2%. Analysatoren ble stilt inn til å gjøre målinger hvert 30sek, og VO2max ble kalkulert gjennom å bruke snittet av de to høyeste påfølgende målingene av O2. Underveis i testen mottok alle deltakerne en høylytt verbal oppmuntring fra testleder. Alle deltakerne gjennomførte også begge testene med samme testleder og med samme personer til stede i rommet for å redusere konfundering (Halperin, Pyne, and Martin 2015).

For hvert medgåtte minutt av testen ble hastigheten på møllen økt med 1km/t, helt til utmattelse, hvor testen ble avsluttet. Deltakernes hjerterefrekvens ble også registrert under hele testen. Når testen ble avsluttet ble deltakerne bedt om å rapportere opplevd anstrengelse ved hjelp av Borg-skala (Borg 1982). Maksimal hjerterefrekvens under testen ble også registrert. Ett minutt etter avsluttet test ble hjerterefrekvens registrert, og det ble målt og analysert blodlaktat (Biosen C-line, EKF Diagnostics, Barleben, Germany).

Resultater

	Kvinner	Menn
N	4	7
Alder (år)	24.5 (1.29)	23.9 (1.77)
Vekt (kg)	58.9 (6.28)	74.8 (5.55)
Høyde (cm)	166 (2.99)	180 (3.1)
Verdier er gitt som gjennomsnitt og (Standardavvik)		

Det var 11 deltakere i studien, samtlige deltakere er studenter ved Høgskolen i Innlandet. Deskriptive data for disse deltakerne er vist i Tabell 1, i Figur 1 kan man se utviklingen fra pre-test til post-test fordelt på kjønn. Det typiske målefeilet (typical error, (Hopkins 2000)) fra pre til post-test er utregnet til å være 4.04%.

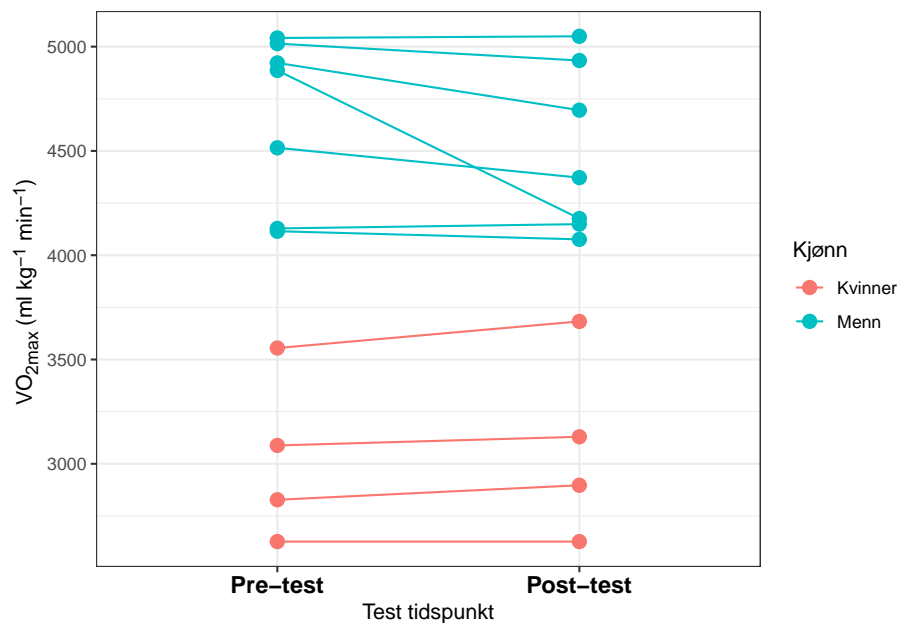


Figure 1: Figurtekst legg til...

Diskusjon

Ettersom testing av maksimalt oksygenopptak er en test som gjennomføres til utmattelse, vil man kunne forvente en viss variasjon i testresultatene ettersom opplevd anstrengelse kan påvirkes av flere ulike variabler (Halperin, Pyne, and Martin 2015). For å redusere konfundering vil flere faktorer være nyttig å ta hensyn til under slik testing. Som nevnt i metoden vil standardisering av matinntak, koffeininntak, utstyr og tidspunkt for gjennomføring av test være med på å kunne sikre intern validitet i resultatene. Eksempler er deltakernes kjennskap til testen, verbal oppmuntring og personer tilstede under testen er andre faktorer som potensielt kan bidra til konfundering. Felles for alle faktorer er at graden av påvirkning på resultatene muligens reduseres ved hjelp av en standardisert testprotokoll. Deltakerne - og testlederne, sin kjennskap til testen er en annen faktor som trolig påvirker resultatene i vårt prosjekt. I dette tilfellet fantes det enkelte deltakere som hadde gjennomført en liknende test flere ganger, og en kan da forvente en mindre grad av variasjon mellom resultatene på pre og post test, sammenlignet med de deltakerne som gjennomførte testen for første gang på pretest. Dette fordi kjennskapen og kunnskapen de tilegnet seg på pre-test, trolig spiller inn på testresultatene. Standardfeilen på 4.04% kan også tyde på at enkelte av disse resultatene kan være utsatt for konfundering av ulik sort (Hopkins 2000).

Grunnen til at vi snakker om standardfeil er at når vi ønsker å måle påvirkningen av trening på en gruppe individer er det viktig å kunne si noe om hva som er endring og hva som er støy (målefeil). Desto mindre støy en test innebærer jo bedre er målingen. Målet som brukes er standardfeil. Hva som danner denne variasjonen som representeres ved typical error er multifaktorelt, men hoveddelen er som oftest biologisk (Hopkins 2000).

For å måle standardfeil har vi brukt within subject deviation metoden. Denne metoden påvirkes ikke av at gjennomsnittet endrer seg fra test til test (Hopkins 2000). Data for målinger i VO₂max fra fem sertifiserte Australske laboratorier fastslo ett gjennomsnitt på 2.2% for standardfeil (Halperin, Pyne, and Martin 2015). Data fra det Australske institutt for sport har også fastslått at en standardfeil på omtrent 2% er riktig for både maksimal og submaksimal O₂ (Clark et al. 2007; Robertson et al. 2010; Saunders et al. 2009). Dette indikerer at med godt kalibrert utstyr og med utøvere som er godt vant med testingen vil en standardfeil på 2% for det biologiske, og analytiske være riktig (Halperin, Pyne, and Martin 2015). Vår standardfeil på 4.04% kan derfor tenkes å være et bilde hvordan det kan se ut med få deltakere, med ulikt utgangspunkt, men også uten skikkelig standardisering av treningshverdagen i forkant av testene. Det kan også tenkes at med et varierende nivå hos deltagerne kan enkelte oppleve en treningseffekt av test 1. Samtidig som andre kanskje ble slitne av å få en test inn i treningshverdagen.

Referanser

Lab rapport

Introduksjon

Vitenskapsteori

Vitenskapsteori er læren om filosofien, metodene, strukturene og betydningene vitenskapen har. Vitenskap har eksistert i alle år, og har vært uttrykket gjennom mer eller mindre strukturerte rammer. Ut ifra disse metodene, strukturene, og filosofien, har det kommet frem forskjellige teorier og metoder, som den moderne vitenskapen bygges på.

Falsifikasjon

Falsifiseringsprinsippet Karl Raimund Popper (1902-1994), var en Østerisk-Britisk filosof. Popper hadde siden høsten 1919 bearbeidet spørsmålet om hva som skulle til for å kunne danne et skille mellom vitenskap og pseudo-vitenskap. På den tiden(starten av og et stykke inn på 1900tallet) var det en aksept for at skilnaden mellom vitenskap og pseudo-vitenskap, var den empiriske fremgangsmetoden(Popper, 1969,s.33). Den empiriske fremgangsmetoden går fra observasjoner, eller eksperimenter til å danne teorier ved hjelp av induktive argumenter. Et induktivt argument bygger på at premissene som er satt er sanne, og ved nok observasjoner av disse premissene, vil konklusjonen sannsynligvis være sann, eller man kan si at man har styrket bekræftelse av en teori. Denne tankerekkefølgen ble kritisert av David Hume (1711- 1776) allerede på 1700tallet. Hume stilte spørsmål for hvordan man kan slå fast at induksjon er pålitelig, og svaret ble ved hjelp av et argument. Hvilket argument skulle brukes? Induktiv metode kan ikke garanteres for ved et deduktivt argument, da det ikke finnes et deduktivt argument for fremtiden. Et induktivt argument bygger på erfaringer, men kan heller ikke slå fast induksjonens sikkerhet. Dette var kronglete å holde styr på i hodet, men var et fundamentet i kritikken av induksjonismen langt inn på 1900tallet.

Popper som i utgangspunktet var ute etter å skille vitenskap fra pseudo-vitenskap viste om denne svakheten ved induktive argumenter som var det som var det opprinnelige skille. Ved hjelp av datidens største og mest fremgangsrike vitenskapsteorier lette Popper etter forskjeller. Einsteins relativitetsteori, marx historiske teorier, Freuds psyko-analyser, og Alfred Adlers individuelle psykologi var de mest omtalte teoriene på den tiden. Når han etter hvert

oppdaget at flere av punktene i de forskjellige teoriene ble bekreftet, ble han også mens tiden gikk mindre og mindre imponert av Marx, Freud og Adler. Det var her han startet å se etter forskjeller. De andre teoriene hadde mange bekreftelser, og hvert eneste tilfelle kunne på en eller annen måte forklares ved hjelp av teoriene. Einsteins relativitetsteori kunne derimot motsies ved hjelp av en eksakt måling. Ut av dette vokste Poppers falsifiserbarhetskriterium. (Popper, 1969,s.34-35)

Falsifiserbarhetskriteriumet vokste altså frem av at en teori ikke burde bekreftes, men avkreftes, eller falsifiseres som det kalles vitenskapelig. Teorien må derfor være så konkret at den kan måles, og dermed falsifiseres. Et eksempel kan være: «det kommer til å regne i morgen». Dette utsagnet vil i løpet av morgendagen kunne falsifiseres, hvis det blir oppholdsvær. Om det skulle regne, vil det komme en ny morgendag, der teorien kan falsifiseres. Dette vil skille pseudo-vitenskap og vitenskap i og kalles for Poppers demarkasjonskriterium. Vitenskap vil derfor være alle teorier som kan falsifiseres. Når noe er falsifisert vil det vokse frem en ny falsifiserbar teori, og vitenskapen som en enhet vil etter hvert nærme seg en sannhet. Med dette kriterium mente også popper at induksjonsproblemet også var løst. Dette mente han ved at alle vitenskapelige teorier er deduktive argumenter som falsifiserer, og at vitenskapen dermed ikke behøver bruke induktive argumentasjon. Denne tankerekkefølgen gjør at man ikke kan stole på at noen ting vil være sant, eller ikke engang at det kan argumenteres for at det er sannsynlig å være sant.

Dette er en av punktene som kritiseres av andre filosofer. Altså er det ikke mulig å si at noe er mer sant en annet. Popper kritiseres av andre av at det ikke er forskjell på godt bekreftede og dårlig bekreftede teorier. Det gjør også det mulig å sette opp nye teorier, eller endre teorier sånn at en teori ikke er falsifisert, da man ikke har noen form for hierarki over hvilke teorier som er best bekreftet. Et annet argument som ofte trekkes frem i denne sammenhengen er at det ikke trenger å være et klart skille mellom hva som er ekte vitenskap, og hva som er pseudo-vitenskap. Det vil i mine øyne ikke være så viktig da hvor godt noe er bevist eller bekreftet bør veie tyngre enn at det er formulert en hypotese som kan motsies. I mine øyne er en god vitenskapelig teori noe som er godt bekreftet. Et argument mot denne definisjonen på god vitenskapelig teori, er at den er en lite konkret og veldig vag definisjon, som er vanskelig å standardisere, og dermed dårlig i praksis. Denne definisjonen på en god teori, betyr at i praksis at teorien er bygget i bunn på induktive argumenter. Det er flere forskjellige metoder å benytte seg av induktiv argumentasjon, fire kjente måter er: Hypotetisk deduktiv metode, naiv induktivism, abduksjon, og bayesianisme

HD-metoden og abduksjon

En av de som støttet induktiv argumentasjon for vitenskapelig teorier var Carl Hempel (1905- 1997), han var naturlig nok (delte nesten et helt århundre sammen, med motsatte meninger) en av Poppers kritikere. Han kritiserte også naiv induktivism, som baserte seg rundt to påstander. Den første er at en teori kan støttes induktivt når den kan trekkes ut fra et utvalg, og den andre er at teorien da bør formuleres etter at alle dataene er innsamlet. Hempel hang seg spesielt opp i det siste og mente at det var umulig å trekke noen konklusjoner før tidens slutt, da det alltid vil dukke opp nye data. Ei heller ville det være mulig å innhente all data som eksisterer i dag. (Hempel 1966, s. 11) Hempel var av en annen oppfatning av hvordan en teori kunne forklares eller bekreftes. Hempel mente dette var den hypotetisk deduktive metoden.

Den hypotetisk deduktive metoden baserer seg på fire steg. Det første steget er steget der det dannes eller formuleres en hypotese, teori etc. Dette steget gjøres ved hjelp av et «educated guess», eller «utdannet gjetning» som oversettelsen blir. Det neste steget er deduksjon, som i praksis betyr å utlede empiriske konsekvenser. Her legges grunnlaget for de neste to stegene da disse empiriske konsekvensene deduseres. I steg tre blir disse deduserte empiriske konsekvensene testet ut, som dirket leder til steg fire. Hvis disse empiriske konsekvensene viser seg å være sanne er denne teorien induktivt bekreftet, eller induktivt styrket bekreftelse av teorien. (Hempel, 1966, s.12-13). Den hypotetisk deduktive metoden kan forklares som en vei med to kjøreretninger, fra data til teori går det induktive argumentet, som er bygd på deduktivt argument som går motsatt vei, altså fra teori til data. Problemet knyttet til Hypotetisk deduktiv metode er at ikke alt kan deduseres, samt at enklere teorier kan passe like bra eller bedre. Om man følger Ockhams barberkniv ønsker man å skjære bort alt unødvendig, da den enkleste teorien som oftest vil være rett.

Abduksjon er en liknende modell til den hypotetisk deduktive metoden, men har et par vir og vendinger som gjør dem klart forskjellige allikevel. Abduksjon følger ordets rette betydning og betyr at man samler seg til eller slutter seg om, den beste teorien som forklarer et fenomen. Er det ingen annen teori som gir like god eller bedre forklaring, har teorien induktivt styrket bekreftelse eller er bekreftet. I Abduksjon vil enkelhet tas hensyn til og hvis tre teorier forklarer det samme fenomenet vil det enkleste induktivt bekreftes.

Replikasjonskrisen

En viktig faktor i vitenskapens unione søken mot et bedret kunnskapsbilde er åpenhet, det skal kunne testes av andre forskere at man får samme data fra samme forsøk. Dette er for å kunne styrke eller svekke et fenomens teori, og kanskje kunne oppdage en ny teori. For at dette skal kunne skje må forskningen

være repliserbar. Dette er ikke vært gjennomført godt nok i de siste årene. Et økt fokus har gjort at det har kommet tall på problemene. 1576 forskere fikk spørsmål om det var en signifikant krise når det kommer til reproduserbarhet i forskning, i en undersøkelse for journalen Nature, der 52% av deltakerne var enige i det (Baker, 2016).

Forklaringen på hva som står bak denne krisen er bestående av en hel haug med små og store bidragsytende faktorer. Den minst beskuende og tillitsfulle forklaringen ligger i statistikken. Sann at replikasjonskrisen kan være et produkt av to statistiske faktorer. De fleste hypoteser vi tester er usanne, og nivået for statistisk signifikans er for lavt ($p < 0,05$). Det er tvilsomme forskningspraksiser, som f.eks tuklinger med data, p-hacking, fjerning av materiale. Det er mange små studier, som fører til lav statistisk styrke. Det publiseres ikke like ofte studier med kjedelige eller negative resultater (Bird, 2018). Det er også kraftig motivasjon for å få frem data med overraskende, eller nyttig resultat, som kan føre til at det blir gjort små ting som gjør at data ikke kan repliseres (Romero, 2017).

For meg høres dette ut som et kombinert problem. Kanskje er verden sann at vi ikke presterer den beste forskningen i ett frittstående konkurranse miljø. Det er enklere å få navnet opp og frem med noe kontroversielt enn med noe som støtter en allerede eksisterende teori. I det idrettsvitenskaplige fagfeltet er det ofte små utvalg i intervensjoner som gjennomføres, og med relativt kort varighet i forhold til adaptasjonene som skal måles på, som gjør at det krever at mye skal stemme for å få et signifikant resultat, som kan forsterke publikasjonsskjevheten, og dårlig forskningspraksis for å kunne få signifikante resultater.

Bassett, D. R., and E. T. Howley. 2000. "Limiting factors for maximum oxygen uptake and determinants of endurance performance." *Medicine and Science in Sports and Exercise* 32 (1): 70–84. <https://doi.org/10.1097/00005768-200001000-00012>.

Borg, G. A. 1982. "Psychophysical bases of perceived exertion." *Medicine and Science in Sports and Exercise* 14 (5): 377–81.

Clark, Sally A., P. C. Bourdon, W. Schmidt, B. Singh, G. Cable, K. J. Onus, S. M. Woolford, T. Stanef, C. J. Gore, and R. J. Aughey. 2007. "The effect of acute simulated moderate altitude on power, performance and pacing strategies in well-trained cyclists." *European Journal of Applied Physiology* 102 (1): 45–55. <https://doi.org/10.1007/s00421-007-0554-0>.

Halperin, Israel, 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. <https://doi.org/10.1123/ijsp.2014-0566>.

Hempel, Carl G. 1966. *Philosophy of Natural Science*. Prentice Hall.

Hill, A. V., and H. Lupton. 1923. "Muscular Exercise, Lactic Acid, and the Supply and Utilization of Oxygen." *QJM* os-16 (62): 135–71. <https://doi.org/10.1093/qjmed/os-16.62.135>.

- Hopkins, W. G. 2000. "Measures of reliability in sports medicine and science." *Sports Medicine (Auckland, N.Z.)* 30 (1): 1–15. <https://doi.org/10.2165/00007256-200030010-00001>.
- Robertson, Eileen Y., Philo U. Saunders, David B. Pyne, Robert J. Aughey, Judith M. Anson, and Christopher J. Gore. 2010. "Reproducibility of performance changes to simulated live high/train low altitude." *Medicine and Science in Sports and Exercise* 42 (2): 394–401. <https://doi.org/10.1249/MSS.0b013e3181b34b57>.
- Saunders, P. U., R. D. Telford, D. B. Pyne, A. G. Hahn, and C. J. Gore. 2009. "Improved running economy and increased hemoglobin mass in elite runners after extended moderate altitude exposure." *Journal of Science and Medicine in Sport* 12 (1): 67–72. <https://doi.org/10.1016/j.jsams.2007.08.014>.
- Tønnessen, Espen, Erlend Hem, Ida Svendsen, Eirik V. Larsen, Martin Skaugen, and Elisabeth Solbakken. 2017. "Utholdenhetstester Ved Olympiatoppen." https://www.olympiatoppen.no/fagomraader/trening/testing/testing_av_utholdenhet/media53703.media.

