

# **Eksamen IDR4000 1 Kvantitativ metode og statistikk**

Kandidatnummer 108

12/2/22

# Table of contents

<b>Forord</b>	<b>4</b>
<b>Deloppgave 1: Beskrivende statistikk, reliabilitet og validitet, verktøy for reproduserbar dataanalyse</b>	<b>5</b>
Introduksjon . . . . .	5
Metode . . . . .	5
Innhenting av datasett . . . . .	6
Resultater . . . . .	6
Gjennomsnitt ved 4mmol/L . . . . .	7
Watt ved 4mmol/L . . . . .	7
Watt . . . . .	8
Hjertefrekvens . . . . .	8
Diskusjon og reliabilitet . . . . .	8
<b>Deloppgave 2: Laborasjonsrapport</b>	<b>10</b>
Introduksjon . . . . .	10
Metode . . . . .	10
Resultater . . . . .	11
Diskusjon . . . . .	11
<b>Deloppgave 3: Vitenskapsfilosofi</b>	<b>12</b>
Falsifikasjonisme . . . . .	12
Hypotetisk deduktiv metode og abduksjon . . . . .	13
Replikasjonskrisen . . . . .	14
<b>Deloppgave 4: Studiedesign</b>	<b>16</b>
Introduksjon . . . . .	16
Metode . . . . .	17
Diskusjon . . . . .	18
Konklusjon . . . . .	19
<b>Deloppgave 5: Analysere repeterte målinger</b>	<b>21</b>
Introduksjon . . . . .	21
Metode . . . . .	22
Deltagere . . . . .	22
Intervensjon . . . . .	22

Resultater . . . . .	23
Kroppssammensetning . . . . .	23
Muskelstyrke . . . . .	25
Diskusjon . . . . .	26
Konklusjon . . . . .	27
<b>Referanser</b>	<b>28</b>

# Forord

Dette er min innlevering til eksamen i IDR4000-1 22H Kvantitativ metode og statistikk. Data og reproduserbare filer for deloppgavene finnes i min GitHub: <https://github.com/Molmshaug/innlevering-idr4000-qmd>

# **Deloppgave 1: Beskrivende statistikk, reliabilitet og validitet, verktøy for reproduserbar dataanalyse**

## **Introduksjon**

Laktatterskel, også kjent som anaerob terskel er den høyeste arbeidsintensiteten der mengden laktat(også kalt melkesyre) som produseres og fjernes er lik. Under fysisk aktivitet frigjøres energi i kroppen både gjennom aerobe og anaerobe prosesser. Ved arbeid over en viss intensitet produseres også laktat i tillegg til energi, desto høyere intensitet, desto større er laktatproduksjonen (Dahl, 2005). Laktatterskelen er altså den høyeste intensiteten en kan holde før konsentrasjonen av laktat overskrider fjerningen av laktat (Brooks, 1986).

Hensikten med denne oppgaven er å presentere målinger for reliabilitet av data innhentet på fysiologilab. Videre vil oppgaven ta for seg reliabiliteten for reproduserbarhet i datavitenskap. Laktatterskel kan være en gunstig indikator for evnen til å yte maksimalt i utholdenhetsidretter. Testing av laktatterskel kan utføres på flere ulike måter, men mest brukt er måling ved 4mmol/L. Videre finnes det ulike måter å justere selve testingen, ved eksempelvis å utføre test på sykkel, tredemølle eller rulleski. Testen kan deretter påvirkes av lengde på intervall og lokasjon for hvor prøvene blir tatt(Faude et al., 2009).

Denne oppgaven tar for seg resultater som ble gjennomført via testing på sykkel, og vil fokusere på watt og hjerterefrekvens ved 4mmol.

## **Metode**

I denne oppgaven er det brukt en kombinasjon av innhentede data fra idrettsstudenter ved HINN og et allerede eksisterende datasett "lactate threshold tests". Data som ble innhentet av studenter ble gjennomført i en periode på 1-4 testdager. Flere av studentene gjennomførte pre og post test samme dag. Under noen av testdagene var det feil med BIOSEN, som førte til unøyaktig målinger. Dataene som ble innhentet av studenter var fra tester som ble gjennomført samme dag. Før selve testen ble gjennomført fulgte testleder og assistent en planlagt protokoll.

Testleder og assistent gjorde klar biosen, nese-klypen, kalibrerte og satt samme munnstykket før testen. Før teststart ble forsøkspersoners navn, alder, fødselsdato, kjønn, høyde og vekt (med minst mulig tøy, 300g trekt fra) lagt inn på dataen. Deretter ble sykkel (Sport Excalibur) tilpasset forsøkspersonene og byttet til riktig pedal i forhold til sykkelens skoene. Forsøkspersonene ble informert om gjennomføringen av testen samt BORG skalaen som ble tatt i bruk under testen. Testen ble gjennomført på Sport Excalibur uten oppvarming. Hvert drag varte i 5 min og ved 4:30 ble forsøkspersonene spurt hvor på BORG skalaen personen befinner seg og det ble tatt laktat ved stikk i fingeren.

Ved måling av laktatprofil ble det angitt en gitt Watt (75-125) til å starte med. Testpersonen ble også i starten informert om at tråkkfrekvensen skulle ligge på mellom 70-100rpm. Det ble kjørt 5 minutters intervaller, hvor de første 1.5 minuttene ble kjørt uten noe ekstra utstyr, før deltakeren tok inn et munnstykke og satte på nese-klype. Det ble gjort jevnlig målinger, og disse hadde assistenten ansvar for å skrive ned fra 2.5 minutt til 4.5 minutt. Testpersonen ble bedt om å gjøre seg opp en mening om hvor slitne de følte seg ut ifra BORG-skala, før det så ble tatt en prøve i fingeren for å måle laktatverdiene. Etter at prøven ble tatt økte motstanden i Watt, og et nytt intervall startet. Dette gjentok seg til testpersonen nådde 4mmol/L.

## Innhenting av datasett

Datasettet ble logget for hånd og tastet inn i Excel. Datasettet er deretter lastet inn fra Excel til Rstudio ved bruk av readxl-pakken i Rstudio. Alle statistiske analyser er gjort i Rstudio (versjon 2022.07.2.576) for windows.

## Resultater

Kjønnsfordelingen mellom kvinner og menn var jevn, med 4 kvinner og 5 menn. Aldersspennet var fra 23 til 31 år, med en gjennomsnittsalder på 25.6 år ( $\pm 2.4$  år).

Table 1: Descriptive statistics

Sex	Age		Height		Weight		Participants
F	24.75	(1.50)	165.75	(2.99)	61.95	(12.20)	4
M	26.20	(3.11)	181.80	(3.11)	77.18	(3.73)	5

Standard deviation in parentheses

## Gjennomsnitt ved 4mmol/L

Gjennomsnitt av resultatene ved 4mmol laktat. Laktat eller melkesyre i blodet refereres til mangel på oksygentilførsel til deler av kroppen under arbeid (Husøy, 2022). Laktatterskel kan eksempelvis forklare hvilken intensitet du kan holde over en gitt periode, og unngå utmatelse.

Table 2: means at 4mmol/L

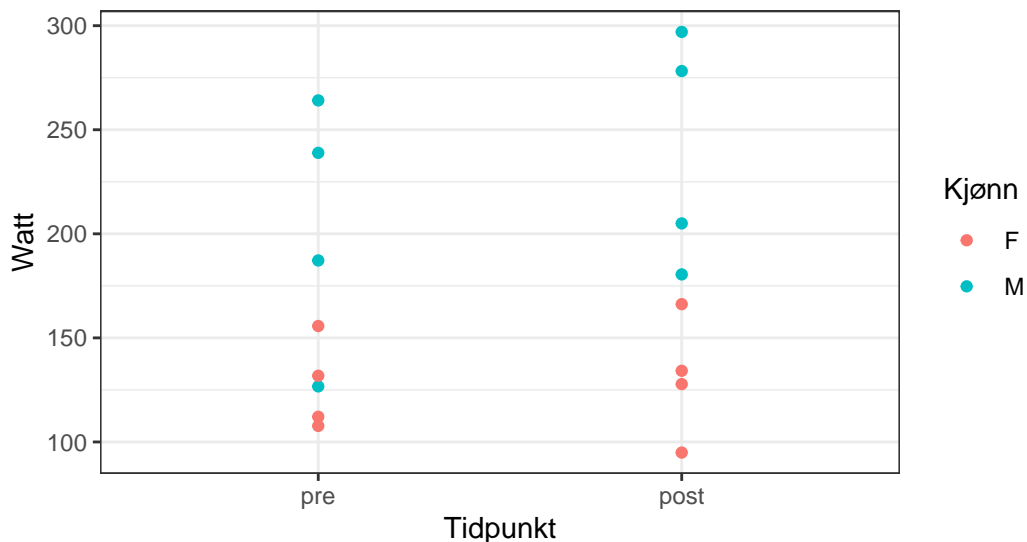
Timepoint	Watt 4mmol		HF 4mmol		Participants
pre	165.525	(59.17)	158.5714	(15.08)	9
post	185.475	(71.69)	161.0000	(11.11)	9

Standard deviation in parentheses

## Watt ved 4mmol/L

Watt produsert på terskel (4mmol/L viser fordeling basert på deltaker og kjønn, ved pre- og posttest.

Table 3: Watt at 4mmol/L



Viser watt på terskel (4mmol/L) ved pre- og posttest  
Menn markert i blått, kvinner markert i rødt

## Watt

Gjennomsnittet for deltakerne var 176 watt. Standardavviket var 19.1 og typical error 13.5. Koeffisienten var 7.70.

Avg watt	SD watt	Watt typical error	Watt coefficient variance	n
175.5	19.1155	13.5167	7.701824	9

SD = Standard deviation

## Hjertefrekvens

Gjennomsnittet for deltakerne var 160. Standardavviket var 4.32 og typical error 3.06. Koeffisienten var 1.91.

Avg HR	SD HR	HR typical error	HR coefficient variance	n
159.7857	4.320494	3.05505	1.911967	9

SD = Standard deviation; HR = heart rate

## Diskusjon og reliabilitet

Resultatet viste først «NA», fordi det lå «NA» i det eksisterende datasettet. Dette ble endret slik at resultat for differansen på watt ved 4mmol/L ble et standardavvik på 19,1, et gjennomsnitt på 176 og typical error på 13,5. Dersom resultatet omhandlet endringene i gjennomsnittet er under eller lik typical error, er ikke testen/resultatene til å stole på/reliable (Hopkins, 2000) (s.11). Ut fra dette utsagnet fra Hopkins (2000) kan vi si at våre resultater er til å stole på. Vi kan med fordel ha hatt en gjennomgang av testen med alle forsøkspersonene i forkant for å minske læringseffekten. Ut ifra Hopkins (2000) hadde flere forsøkspersoner i testen gitt et bedre resultat og testen kunne blitt mer reliabel (Hopkins, 2000) (s.10). Noen av studentene fikk unøyaktig målinger grunnet mangel på oksygenopptak, mangel på pulsbelt og feil med BIOSEN. Disse testresultatene ble ekskludert fra rapporten.

### Reliabilitet

Ut fra Hopkins (2000) kan det tenkes at de to viktigste synspunktene for feil ved måling er validiteten underveis i testen og pålitelighet til eventuelle retester (Hopkins, 2000). I tillegg uttaler Hopkins at de viktigste målingene etter en test er typiske feil, også kalt typical error, og endringene i gjennomsnittet mellom hver test (pre – post) (Hopkins, 2000). Videre påpeker



Hopkins (2000) at det er tre måter å måle reliabiliteten på; «within-subject variation», «change in the mean» og «retest correlation». Måling av påliteligheten av en test kan gjøres ved å utføre testen et bestemt antall ganger på et bestemt antall forsøkspersoner. Desto mindre feil en test har, jo mer målbart er den (Hopkins, 2000).

### **Hvilke mål av reliabilitet har vi brukt?**

Vi har tatt i bruk metodene fra Hopkins (2000) for å måle reliabiliteten. Den ene metoden vi har brukt kalles «within-subject variation». Denne metoden omhandler tilfeldige variasjoner og det finnes statistikk som fanger opp disse tilfeldige variasjonene av enkeltpersoners resultater i en retestet test. Dette omtales ofte som standardavvik og representerer typical error, i en måling (Hopkins, 2000). Utrekningen av typical error gjøres ved å dele standardavviket på kvadratroten av 2 (Hopkins, 2000). Det finnes ulike typer kilder til typical error i en måling, derav biologiske faktorer. Forsøkspersonene kan ha ulik fysisk- og mental tilstand under de ulike testene som kan være med på å påvirke resultatet. Hopkins påpeker også at det er ideelt at det i en test er mange forsøkspersoner og få prøver/re-tester (Hopkins, 2000).

Den andre metoden vi har brukt omhandler endring av gjennomsnittet. Endringene kan både være tilfeldige endringer eller systematiske endringer. Hopkins (2000) forteller at systematiske endringer i gjennomsnittet kan påvirkes av forsøkspersonenes økte læringseffekt fra første til andre gjennomføring av testen. Videre påpeker Hopkins (2000) at det er ideelt å utføre nok tester slik at læringseffekten eller andre systematiske endringer blir ubetydelig før intervensjonen. De systematiske endringene i gjennomsnittet kan påvirke enkeltindividers resultat som igjen kan øke typical error (Hopkins, 2000). Det er dermed viktig med god planlegging før en test for å eliminere så mye som mulig av læringseffekten. Det er flere faktorer som kan påvirke resultatene, hvor blant annet forsøkspersonenes motivasjon kan påvirke resultatene ved at de enten får mer eller mindre motivasjon. De kan også få en opplevelse av fatigue fra forrige test som kan være med å påvirke resultatene. Ulike testledere, ulikt utstyr og ulike omgivelser er også faktorer som kan påvirke resultatene .

Til slutt har vi metoden omhandlet retesting. Hopkins (2000) viser til tanken om å utføre testen to ganger på ulik tid på samme forsøksperson, for å finne ut om resultatet er tilnærmet likt (Hopkins, 2000). Desto nærmere resultatene er, jo mer reliabel er testen.

# Deloppgave 2: Laborasjonsrapport

## Introduksjon

Analyser av genuttrykk gjennom kvantitativ PCR i skjelettmuskulatur er rutine i treningsstudier (Kuang et al., 2018). Hvordan forsøkene gjennomføres og tolkes har stor betydning for reliabiliteten og reproduserbarheten til dataene, og det synes å være stor variasjon i protokoller og dataanalyser mellom ulike laboratorium og forskningsmiljøer. qPCR-målinger i idrettsvitenskapelig forskning er interessant å undersøke ettersom man kan se trenings-induserte endringer i genuttrykk i skjelettmuskulatur.

Prinsippet bak qPCR-testing er å følge prosessen til DNA polymerase-drevet amplifikasjon av DNA (PCR) i sanntid. Under PCR benyttes et termostabilt DNA-polymerase enzym til å syntetisere nye DNA-tråder komplementært til DNA-sekvensene man ønsker å undersøke. Enzymet blandes med DNA-templatet, primere og nukleotider (Kuang et al., 2018). Deretter følger 3 syklisk repeterte prosesser; deling av trådene (denaturering), primerbinding (annealing) og syntese av nye DNA-tråder (extending). Avslutningsvis vil sekvensen man har satt seg ut bli amplifisert til milliarder av kopier, også kalt PCR-amplikoner. Alternative metoder til qPCR-testing er nortern blotting, in situ hybridization og tradisjonell PCR. Sammenlignet med tradisjonell PCR, er qPCR i stand til å oppdage amplifikasjonen av PCR-amplikonet (sluttproduktet) på slutten av hver amplifikasjonssyklus (Kuang et al., 2018). Dette gjennom å bruke et “fluorescent dye system” og et “thermocycler with fluorescence-detection capability”. Sammenlignet med nortern blotting og in situ hybridization er qPCR enkelt og raskt å gjennomføre, og samtidig mer sensitiv og spesifikk.

## Metode

qPCR analyse ble gjennomført i laben sammen med labansvarlig. Labansvarlig hadde på forhånd forberedt fortynnet cDNA prøver som vi skulle bruke til analyse. qPCR-reaksjoner ble gjennomført med fortynnet cDNA og en blandet masterblanding. Masterblandingen bestod av med SYBR-green 2X mastermix (5 l), like deler primermix forward og reverse (1 l) og 2 l kjernefritt vann (H<sub>2</sub>O). 2 l med cDNA ble tilført mastermixen. Innholdet ble sentrifugert i 1 min på 1200rpm for å blande innholdet. Prøvene ble analysert i en sanntids PCR-system og med QuantStudio programvare (Applied Biosystems, Waltham, MA USA). PCR-systemet og QuantStudio analyserer antall sykluser som må til for at det fluoreserende signalet når

cyclisk terskel. Semi-kvantitative målinger ble gjort med 47s rRNA F1R1 som referansegen (Wackerhage, 2014).

## Resultater

Tabellen viser den sykliske terskelen for hvert målgjenstand per prøve. Hvert tall viser antall sykluser før den sykliske terskelen ble oppnådd. Et lavere tall indikerer et større genuttrykk. Til tross for at vi ikke testet for MyCH2X genuttrykk fikk vi en observasjon for MyCH2X på en prøve. Det er vanskelig å forklare hva årsaken til feil-observasjonen er, men vi mistenker at det skyldes feil-merkning under qPCR-syklusene.

Tabell 1: qPCR Results

Sample Number	MHC1 F1R1	MHC2A F2R2	MHC2X F2R2	MyHC2X	47s rRNA F1R1
33	10.47	13.04	17.24	NA	20.40
34	10.83	12.21	16.90	NA	20.84
37	12.39	11.43	15.45	NA	19.58
38	11.84	11.42	14.63	NA	19.99
39	13.41	11.54	14.36	NA	21.00
40	12.97	12.28	15.95	NA	20.88
43	11.44	11.28	16.00	NA	21.23
44	12.57	12.00	NA	15.26	20.58

## Diskusjon

Sammensetningen av muskelfibre i menneskelig skjelettmuskulatur kan måles ved bruk av immunhistokjemi. Dette er en teknikk som brukes på vevssnitt. En metode som har en tendens til å være både tidkrevende, teknisk utfordrende og komplisert på grunn av små mengder prøvevev er qPCR (Ellefsen et al., 2014). De tekniske utfordringene ved å gjennomføre en qPCR er mange og kan være malkvalitet, operatørvariabilitet, selve revers-transkripsjonen og subjektivitet i dataanalyse og rapportering (Bustin & Nolan, 2004). Dette gjør nøyaktig datatolkning vanskelig, som også ses i vårt tilfelle med en observasjon på MyCH2X som ikke skulle vært tilstede.

# Deloppgave 3: Vitenskapsfilosofi

## Falsifikasjonisme

*Hva er Poppers falsifiserbarhetskriterium og hvilket spørsmål skal dette kriterium gi svar på? Hvorfor mener andre vitenskapsfilosofer (f.eks. Okasha) at vi ikke trenger å svare på dette spørsmålet? Hvem synes dere har rett?*

Poppers falsifiserbarhetskriterium er basert på at en teori skal være falsifiserbar, men ikke falsifisert. Popper mente at dersom en teori skulle være vitenskapelig, måtte den også være mulig å avkrefte. Videre mente han at evidens som bekrefter en teori ikke eksisterer. Teorier som ble tilpasset data og som dermed ikke kunne motsies kalte Popper for «pseudo-science» («ikke-vitenskapelig vitenskap»). Popper kritiserte Karl Marx og Sigmund Freud sine teorier for å være «ikke-vitenskapelige». Karl Marx sine teorier ble kritisert av Popper for å tilpasses underveis. Dersom teoriene møtte motstand igjennom en test eller gjennom data så kunne teorien endres for å tilpasse seg motstanden i dataen. På denne måten kunne Marx fremdeles beholde teorien sin. Popper mente at Freuds teorier ikke var mulig å teste eller falsifisere og beskrev de dermed som myter og ikke som vitenskap. Poppers falsifiserbarhetskriterium skulle derfor gi svar på om en teori er vitenskapelig eller ikke. Utfordringen med å skille mellom vitenskapelige og «ikke-vitenskapelige» teorier kalte Popper for demarkasjonsproblemet (Okasha, 2016; Popper, 1969).

Flere filosofer har kritisert Poppers falsifiserbarhetskriterium for å være altfor forenklet. En av de er Okasha (2016) som skriver i *Philosophy of Science: A Very Short Introduction* at det er flere eksempler på det som Popper kalte for «pseudo-science» har ført til flere viktige vitenskapelige funn. For eksempel ved oppdagelsen av planeten Neptun. To forskere på midten av 1800-tallet John Couch Adams fra England og Urbain Leverrier fra Frankrike oppdaget planeten Neptun ved å gjøre nettopp det Popper kritiserte Marxister for å gjøre. Adams og Leverrier forsket videre til tross for at Newtons gravitasjonsteori var feil i spådommen om Uranus sin bane i verdensrommet. I likhet med Marx og Freud tilpasset Adams og Leverrier sin teori og konkluderte med at det måtte finnes en ny planet som ikke var enda oppdaget. Neptun ble like etter oppdaget som bekreftet Newtons gravitasjonsteori nok en gang. (Okasha, 2016). Okasha mente med blant annet dette eksempelet at Poppers forsøk på å avgrense vitenskap fra «pseudo-vitenskap» ikke kan stemme. Jeg mener selv at Poppers falsifiserbarhetskriterium har hatt en stor verdi for vitenskapsteori. Det utfordrer til kritisk tenkning og filosofering med en enkel og forståelig metode. Videre mener jeg at Okasha har rett i at Poppers falsifiserbarhetskriterium blir for enkelt og at det i tilfeller mener jeg at det

blir riktig og si at en teori alltid er feil eller gal (Okasha, 2016). Det finnes en slags objektivitet ved at en teori skal kunne falsifiseres, men at den viktigste forskjellen ligger i hvor godt eller dårlig en teori er bekreftet. Det vil ikke være riktig å avskrive enhver teori som møter litt motstand, men jeg mener det bør være en viss balanse mellom det å avkrefte og bekrefte teorier. Dette ville også avskrevet flere viktige vitenskapelige funn. Spørsmålet om vi skal benytte oss av Poppers falsifiserbarhetskriterium blir derfor erstattet av spørsmålet om teorier er gode eller dårlige og hvorvidt det finnes empirisk data og statistikk for å bekrefte eller avkrefte teorier.

## Hypotetisk deduktiv metode og abduksjon

*Hva er strukturen på et bekreftende vitenskapelig argument ifølge den hypotetisk deduktive metode? Forklar ut fra Hempels artikkel, men bruk egne eksempler.*

Hypotetisk deduktiv metode (HD-metoden) starter med at vi formulerer en teori eller hypotese. Deretter deduserer vi empiriske konsekvenser som følge av teorien eller hypotesen som testes gjennom eksperiment eller empiriske observasjoner. Dersom deduksjonen viser seg å stemme bekreftes teorien induktivt, til en viss grad. Det sies at teorien bekreftes til viss grad på grunn av at teoriene aldri kan bekreftes til å være 100% sanne. Hempel (1966) beskriver dette med at fremtidig relevant data kan gi nye funn enn det som tidligere har kommet frem. Dette kan beskrives med et eksempel fra styrketrening. For å oppnå økt maksimal styrke er det bedre å trene med få repetisjoner og tunge vekter enn å trene med mange repetisjoner på en lettere vekt. Nåværende forskning viser at dette stemmer, men vi kan ikke bevise at dette er sant ved 100% av tilfellene. Det kan komme forskning frem i tid som viser helt andre resultater enn det vi har i dagens litteratur. Med en slik induktiv påstand vil vi aldri være helt sikker på at det stemmer, men økt vitenskap og mer forskning vil enten styrke eller svekke påstanden. På denne måten kommer vi også frem til svakhetene ved Hempel sin HD-metode.

Andre faktorer kan også påvirke resultatet til påstanden. For eksempel fysiske forutsetninger hos populasjonen det forskes på, forskjellig næringsinntak, søvnmengde og -rytme osv. Alle disse faktorene kan spille inn på effekten av styrketrening enten det er med tunge vekter og få repetisjoner eller om det er med lette vekter og mange repetisjoner. HD-metoden er alltid åpen for at det er andre forklaringer og slutter seg derfor bare til en teoris styrke eller svakhet ut fra det man har testet empirisk (Hempel, 1966). HD-metoden er også svak ved spørsmål om sannsynlighet eller prevalens, ettersom at det ikke er mulig å dedusere noe om frekvens. Et eksempel kan være følgende teori: Sannsynligheten for å få et hoftebrudd øker i takt med alderen. Hvis man i dette tilfellet sier at sjansen for å få et hoftebrudd øker med 2% hvert år etter fylte 60 år så vil man i teorien ha en sannsynlighet for å få et hoftebrudd som øker lineært med alderen. Svakheten med HD-metoden i et slikt tilfelle er at det finnes flere teorier/faktorer som kan påvirke risikoen for å få et hoftebrudd enn bare økt alder, for eksempel aktivitetsstatus, kjønn, tidligere skadehistorikk og funksjonsnivå for å nevne noen. Dette kan påvirke frekvensen for å få hoftebrudd etter fylte 60 år slik at sannsynligheten ikke

er slik at den økes 2% per år. I dette tilfellet tar ikke HD-metoden hensyn til at ulike faktorer kan påvirke dataen.

Abduksjon er en annen teori som ligner den hypotetisk deduktive metoden. Abduksjon er også kjent som «slutning til den beste forklaringen». Den er ment å løse problemene til HD-metoden. Den logiske strukturen i abduksjon er lignende den i HD-metoden, men skiller seg fra at den slutter seg til den beste forklaringen. Den beste forklaringen anses å være god eller bedre når den enten forklarer flere forskjellige data, har høyere forklaringskraft, utgår fra færre årsaker/faktorer eller at den er tilsvarende en annen forklaring, men enklere. Abduksjon krever med andre ord at man sammenligner flere teorier for deretter å lande på den teorien som best forklarer teorien/hypotesen. Der abduksjon stiller flere teorier opp mot hverandre og sammenligner for å finne den teorien som passer best, tar HD-metoden kun for seg en teori om gangen for så å undersøke om dataen/funnet induktivt støtter teorien.

## Replikasjonskrisen

*Hva mener Alexander Bird er forklaringen til at mange resultat i noen vitenskaper ikke repliseres? Oppsummere Birds argument for dette. Sammenlign også Birds forklaring med noen av de andre forklaringene som Bird diskuterer i seksjon 4. Har Bird rett i at hans forklaring er bedre?*

Bird (2020) beskriver at vitenskapen innen psykologi og medisin har vært utsatt for en krise. En krise kalt replikasjonskrisen. Dette er fenomen der flere tidligere vitenskapelige studier er forsøkt replisert, men der resultatene ikke har stemt overens med tidligere studier. Årsaken har blitt forsøkt forklart ved å peke på dårlig gjennomførte studier, tvilsomme forskningspraksiser, at det finnes en bias mot å publisere negative resultater og til og med forfalskning av resultater. Replikasjonskrisen dreier seg også om frykten for å grave opp i tidligere artikler og kritisere disse, ettersom dette ville svekke tilliten til vitenskapen i offentligheten.

Alexander Bird forklarer i sin artikkel i British Journal for the Philosophy of Science (Bird, 2020) det han mener er en av utfordringene til at vitenskaper ikke repliseres. Bird mente at basisratefeilen (base rate fallacy) er en sannsynlig årsak til replikasjonskrisen. Han beskrev at basisratefeilen sannsynlig vil oppstå dersom man trekker en slutning om sannsynlighet for en gitt forekomst av et generelt fenomen. Feilen som gjøres er at man fokuserer spesifikt på resultatet av en test, men glemmer å ta over seg hvor hyppig forekomsten av gitt fenomen skjer uavhengig av testen som gjennomføres. Dette gir utslag på at en for eksempel konkluderer med at sannsynligheten for å få en sjelden sykdom er større enn det den i realiteten er, fordi man kan feiltolke falske positive eller falske negative resultater (type-I-feil og type-2-feil). I vitenskapen i dag er resultater typisk regnet som signifikant eller ikke signifikant basert på P-verdien. Grensen for P-verdien er som regel satt på 0.05 (5%) og blir en slags pekepinn på om noe er sant eller usant basert på om resultatet er over eller under grensen. P-verdien sier noe om sannsynligheten for at et resultat er falskt positivt (type-1-feil). Et problem med å bruke P-verdi er da kan man konkludere på en befolkning, men P-verdien i studiet vil bare si

noe om utvalget som det forskes på. Type-2-feil, kalles også falsk negativ, og betyr at man forkaster en alternativ forklaring, selv om den egentlig stemmer.

Av andre forklaringer som kan ha bidratt til replikasjonskrisen skriver Bird (2020) om at flere studier har lav statistisk, publiseringsbias og tvilsom forskningspraksis og svindel. Flere har kommet med forklaringen om at lav statistisk styrke er en av årsakene til replikasjonskrisen. Studier med få forsøkspersoner blir spesielt trukket frem, da det er vanskeligere å replisere resultater gjort på små populasjoner. Publiseringsbias eller publiserings-skjevhet er det at det er for mange studier med positive resultater er publisert, når man sammenligner med studier med negative resultater. Det er et stort press fra flere kanter som alle ønsker å publisere nye funn og positive resultater, og ikke negative resultater. Dette kan føre til økt ønske om å replisere studier med positive funn og at igjen kan øke forekomsten av at falske positive resultater blir publisert. Tvilsom forskningspraksis er også et av områdene det gjøres mye feil i. Dette kan være alt fra ubevisst skjevhet og p-hacking, der man leter bevisst etter positive resultater i egen studie, til helt bevisst forfalskning av data. Presset på å publisere er så høyt at dette kan føre til at forskningen gjennomføres etter standarden den burde. Det er også store økonomiske faktorer som påvirke ønsket om å publisere. Det kan være årsaker der større studier koster langt mer penger å gjennomføre og derfor tar man til takke med mindre studier og publiserer heller resultater og funn hyppigere.

Det er lett å si at Bird har rett til at basisratefeilen er en av årsakene til replikasjonskrisen. Dette er heller ikke hele årsaken da vi vet et at det finnes skjevhet der positive eller negative resultater ikke publiseres like ofte. Ønsket om å finne helt nye funn er så stort at forskere kan ignorere bevisst og ubevisst sannsynligheten for at de har falske positive resultater. Det er veldig mange studier som er gjennomført med lav statistisk styrke som følge av få forsøkspersoner, dette fører til at det er vanskelig å replisere disse studiene. Det økonomiske presset eller insentivene som finnes er også store, gjerne for å finne positive resultater, som kan gi disse resultatene for mye vekt, men også føre til økt hyppighet av forfalskning og svindel.

# Deloppgave 4: Studiedesign

## Introduksjon

Styrketrening er den viktigste intervensjonen for å øke muskelmasse blant mennesker. Teorien sier at volumet med styrketrening spiller en signifikant rolle i kroniske muskulære endringer og tilpasninger som for eksempel økt styrke og økt muskelmasse (Kraemer & Ratamess, 2004). Ivaretagelse av muskelstyrke og muskelmasse er viktig for alle. For populasjonen i sin helhet er det viktig med tilstrekkelig styrke for å kunne utføre daglige oppgaver uten utfordringer. Å være fysisk sterk har også flere positive helseeffekter, det er til og med god evidens på at økt muskulær helse gir lavere dødelighet (Ruiz et al., 2008).

Det er tidligere foreslått at styrketrening med høyt treningsvolum, med moderat til høy intensitet og korte hvilepauser gir først og fremst muskelbyggende effekt i form av hypertrofi, deretter økt utvikling av styrke (Baechle et al., 2008). På den andre siden vil et lavere treningsvolum, men tyngre vekter/økt intensitet og lengre hvilepauser primært være et program for å øke muskelstyrke først og fremst. Dette er i tråd med hvordan kroppsbyggere trener i faser der de ønsker å maksimere muligheten for hypertrofi og i studier som rapporterer om økt hypertrofi i tråd med at treningsvolumet øker (Schoenfeld et al., 2017).

I denne oppgaven skal det undersøkes hvordan ulike mengder og ulike intensiteter med styrketrening påvirker utviklingen av styrke og muskelvekst hos trente individer. Det er valgt ut fem artikler for gjennomgang av disse undersøkelsene (Brigatto et al., 2022; Mangine et al., 2015, 2016; Schoenfeld et al., 2019, 2015). For å finne frem til nettopp disse 5 studiene ble det brukt et generelt litteratursøk etter originale studier rettet mot treningsvolum og intensitet, begrenset til de siste 10 årene. 2 av studiene undersøkte lavt treningsvolum versus høyt treningsvolum, på ulike intensitetsområder (Schoenfeld et al., 2019, 2015). I de to neste artiklene undersøkte forfatterne (Mangine et al., 2015, 2016) effekten av et visst treningsvolum versus treningsintensitet i to ulike studier, med litt forskjellige målemetoder. I det siste av fem studier så Brigatto et al. (2022) på effekten av ulike treningsvolum for å undersøke nærmere dose-respons forholdet til styrketrening. Oppgaven inkluderer en vurdering av studiedesign, likheter og ulikheter i studiene og styrker og svakheter med de forskjellige studiene.



## Metode

Schoenfeld et al. (2015) brukte et randomisert studiedesign der 24 menn ble parett etter styrke ved pre-test for deretter å bli tilfeldig plassert i en av to grupper. De hadde en gjennomsnittsalder på 23 år, hadde ingen nåværende muskel- og skjelettskader og var «rene utøvere». Alle deltagerne var erfarne i styrketrening, det vil si at de har trent minimum 3 økter i uken det siste året og regelmessig trent knebøy og benkpress. Studien bestod av en gruppe som trente på lav intensitet (30-50% av 1RM) til utmattelse og den andre gruppen trente 8-12 repetisjoner på høy intensitet (70-80% av 1RM). Begge gruppene trente samme øvelser, men med ulike vekter og repetisjoner. Tester ble gjennomført før og etter (pre og post) intervensjonsperioden (8 uker) med måling av maksimal styrke i form av 1RM i Knebøy og benkpress og måling av muskulær utholdenhet i benkpress ved maks antall repetisjoner utført på 50% av 1RM. Det ble også målt tykkelse av muskulatur ved å bruke ultralyd. Tykkelsen ble målt på triceps, biceps og quadriceps. En parett t-test ble brukt for å undersøke forskjeller fra før og etter intervensjon. Multipel regresjonsanalyse med mål post-intervensjon som avhengig variabel og pre-test verdiene som kovariater ble brukt for å vurdere forskjeller mellom gruppene. Forutsetningene i regresjonsanalysene ble kontrollert. Signifikansnivået ble satt til  $P < 0.05$ .

Mangine et al. (2016) brukte et randomisert studiedesign der 33 fysisk aktive menn ble rekruttert for å sammenligne to ulike intervensjoner. Alle deltagerne med erfaring i styrketrening tilsvarende minimum 2 år (gjennomsnitt på 5.7 års erfaring), uten noen nåværende muskel- og skjelettskader. Intervensjonen var et 8 ukers styrketreningsprogram etter et 2 ukers styrketreningsprogram likt for alle deltagerne. Gruppene ble delt i to grupper der en gruppe gjennomførte trening på høy intensitet og lavt treningsvolum, og den andre gruppen gjennomførte trening på høyt volum og moderat intensitet. Utover dette har ikke forfatterne utdypet metoden de brukte for å rekruttere deltagere. Deltagerne ble testet i 1RM i knebøy og benkpress, med måling av hastigheten på stangen i den konsentriske fasen. Nøyaktige samme maksimale vekt ble brukt på pre- og post-tester, slik at forfatterne kunne utforske om hastigheten endret seg på samme vekt etter intervensjonsperioden. Kraftproduksjon ved isometrisk kontraksjon i andredraget (midten av lårene, i markløft/frivending) ble også målt ved hjelp av en kraftplate og en fastspent stang. Dataen ble analysert med kovariansanalyse (ANCOVA) etter intervensjonsperioden. Dataen fra pre-test ble brukt som kovarians. En parett t-test ble brukt for å undersøke om det var en signifikant endring fra pre- til post-test. Forskjellen mellom gruppene ble analysert ved å regne ut effektstørrelsen. Signifikansnivået ble satt til  $P < 0.05$ .

Brigatto et al. (2022) brukte et randomisert studiedesign der 27 mannlige deltagerne ble parett etter styrkenivå og deretter tilfeldig plassert i en av tre like store grupper. Der gruppene gjennomførte henholdsvis, 16, 24 og 32 ukentlig sett med trening per muskelgruppe. Deltagerne hadde minimum 1 års erfaring med styrketrening tre ganger per uke og ble rekruttert gjennom universitetets treningsstudio. Det ble på forhånd gjort beregninger på utvalgsstørrelsen med basert på en pilotstudie, med en effektstørrelse på 0.75, alfanivå på  $P < 0.05$  og statistisk styrke på 0.80. Deltagerne ble testet med 1RM i knebøy og benkpress og ved å måle

muskeltykkelse ved bruk av ultralyd på biceps, triceps og vastus lateralis. Det var en utdannet tekniker som utførte alle ultralydene. Forfatterne bekreftet normaliteten og homogeniteten til variansene ved å bruke Shapiro-Wilk og Levene tester. En variansanalyse med gjentatte mål (ANOVA) ble brukt for å sammenligne 1RM og hypertrofi før og etter intervensjonen på tvers av studiegruppene.

Schoenfeld et al. (2019) brukte et randomisert studiedesign med 45 mannlige deltagere som meldte seg frivillig. Deltagerne var studenter på et universitet som var erfarne innen styrketrening med gjennomsnittlig erfaring på 4.4 år. Det ble på forhånd gjort beregninger på utvalgsstørrelsen med en effektstørrelse på 0.25, med en alpha-verdi på 0.05 og en statistisk styrke på 0.80. Disse beregningene identifiserte at studien hadde behov for minimum 36 deltagere. Øvrige deltagere var rekruttert i tilfelle noen trakk seg underveis. Deltagerne ble deretter tilfeldig delt in i 3 like store grupper der en gruppe gjennomførte 1 sett per øvelse per økt, en gruppe trente 3 sett per øvelse per økt og den siste gruppen trente 5 sett per øvelse per økt. Det ble målt muskeltykkelse med ultralyd på biceps, triceps og quadriceps av en enkelt utdannet tekniker på alle deltagere. Maksimal styrke ble testet med 1RM i knebøy og benkpress og muskulær utholdenhet i benkpress ble testet med 50% av 1RM til utmattelse. Statistiske analyser ble gjennomført ved å bruke både frekventistisk og Bayesiansk tilnærming. Den første brukte en ANCOVA på endringen fra pre-test til post-test, med separate grupper som faktor og med verdiene fra pre-test som kovarians. I den Bayesianske tilnærmingen ble det brukt en JZS Factor ANCOVA med standardiserte skalaer. Signifikansnivået ble satt til  $P < 0.05$ .

Mangine et al. (2015) brukte et randomisert studiedesign med 33 fysisk aktive menn. Minimum 2 års erfaring med regelmessig styrketrening var et av inklusjonskriteriene (samme studiepopulasjon som Mangine et al. 2016). Alle deltagere gjennomførte samme treningsprogram på 2 uker før de ble tilfeldig delt inn i en av 2 like store grupper. Den ene gruppen trente med et høyt volum på 10-12 repetisjoner på 70% av 1RM, mens den andre gruppen trente på en høy intensitet på 3-5 repetisjoner på 90% av 1RM. Begge gruppene trente i 8 uker. Deltagerne ble testet og målt med maksimal styrke med 1RM i knebøy og benkpress, ultralydmålinger på triceps, quadriceps og pectoralis major for å måle muskeltykkelse og DXA skann for å måle kroppsmasse og endringer. Alle DXA undersøkelser ble utført av samme utdannede radiograf. Blodprøver ble tatt ved 4 tidspunkt på pre- og post-testing, med gjentakende tidspunkt på dagen. Dataen ble analysert med kovariansanalyse (ANCOVA) etter intervensjonsperioden. Dataen fra pre-test ble brukt som kovarians. En paret t-test ble brukt for å undersøke om det var en signifikant endring fra pre- til post-test. Forskjellen mellom gruppene ble analysert ved å regne ut effektstørrelsen. Signifikansnivået ble satt til  $P < 0.05$ .

## Diskusjon

Alle studiene gjennomførte et randomisert studiedesign, men Brigatto et al. (2022) skriver sin studie som en randomisert longitudinell studie. Ettersom ingen av studiene har en kontroll-

gruppe, men sammenligner to eller flere ulike intervensjoner er det heller ingen av forfatterne som hevder at de gjennomfører en randomisert kontrollstudie. Alle studiene er gjennomført med en 8 ukers treningsperiode. Antall forsøkspersoner er fordelt fra 18 til 45 forsøkspersoner. Alle forsøkspersoner er unge menn i aldersgruppen 18-35 år og alle er kjent med styrketrening, med litt ulike definisjoner på antall år for å regnes som kjent med styrketrening. Alle studiene beskriver hvordan forsøkspersonene er rekruttert, Mangine et al. (2016) har en begrenset beskrivelse av rekruttering av forsøkspersoner. Tre av fem studier er helt randomiserte og de siste studiene bruker stratifisert blokkert randomisering (Hulley, 2013).

Alle fem studier bruker 1RM målinger i knebøy og markløft, noe som bidrar til en enkel sammenligning av forskjellige studier til tross for ulike spørsmål mål med studiene. Utover det varierer målemetoden litt fra studie til studie. Det er ikke i noen av studiene nevnt noe om teknikker som gjennomfører ultralyd- eller andre målinger er blindet. Det er også relativt store variasjoner i hvordan målinger med ultralyd er gjennomført. Liggende og stående målinger er ikke nødvendigvis sammenlignbare, og det ville vært en fordel i denne analysen at alle målingene ble gjennomført i samme utgangsstilling. Øvelsene som brukes i de forskjellige studiene samsvarer i all hovedsak godt i de forskjellige studiene, men kun en eller to øvelser som varierer fra studie til studie. Felles for studiene er inklusjon av typiske øvelser som brukes i styrketreningsprogram og av kroppsbyggere (Baechle et al., 2008).

To av de fem studiene beskriver at de på forhånd har beregnet utvalgsstørrelsen som er nødvendig for å trekke konklusjoner, men de øvrige tre studiene beskriver ikke dette i sin statistikk. De forskjellige studiene bruker ANCOVA der data samlet fra pre-test ble brukt som kovarians. Brigatto et al. (2022) bruker ANOVA for å sammenligne 1RM og hypertrofi før og etter intervensjonen på tvers av studiegruppene. Tre av fem studier beskriver at de også brukte en parret t-test for å undersøke om det var en signifikant endring fra pre- til post-test innad de forskjellige gruppene.

Alle studiene svarer på egen problemstilling gjennom analyse av endringer i pre og post-verdier, både innad i egne grupper og på tvers av ulike grupper. Alle studier, bortsett fra Mangine et al. (2015) går grundig gjennom studiene sine begrensninger, sistnevnte beskriver ikke begrensninger studiet kan ha hatt. Alle artikler er spesifikke på sine konklusjoner til sin populasjon (unge, trente menn), men ikke alle går dypere gjennom mulig inferens mot øvrig populasjon.

## Konklusjon

Alle studiene adresserte hypotesen deres i konklusjonen og diskuterer resultatene som førte til dette. Schoenfeld et al. (2015), Schoenfeld et al. (2019) og Brigatto et al. (2022) foreslår alle videre forskning innenfor teamet. Alle spesifiserer også behovet for studier på treningsvolum og effekt av styrketrening mot en bredere befolkning, for eksempel kvinner og eldre. Mangine et al. (2015) og Mangine et al. (2016) er to studier som kommer fra nøyaktig samme populasjon og lik intervensjon, men med ulike mål og hypoteser. Ingen av de studiene kommer med forslag til videre forskning og har heller ikke et stort søkelys på inferens i sin artikkel. Videre studier

undersøke og sammenligne effekten unge, trente menn har på styrketrening med eldre og/eller kvinner. Dette kan i fremtiden gi oss mer kunnskap om treningsprogrammering og spesifikk inkludering av trening på en gitt intensitet eller et gitt treningsvolum.

# Deloppgave 5: Analysere repeterte målinger

## Introduksjon

For en tid tilbake var kroppsbygging, styrketrening og hypertrofi et domene forbeholdt sterke menn, og kun menn. Først i OL i Sydney i 2000 fikk kvinner konkurrere på lik linje menn i olympisk vektløfting (Stone et al., 2006). Siden den tid har søkelyset på styrketrening økt og det har blitt en av de vanligste formene for trening i det norske samfunn ifølge Levekårsundersøkelsen i 2019 (Statistisk sentralbyrå, 2019). Styrketrening har visst seg å ha en veldig positiv effekt på blant annet aldersbetinget muskeltap (Hurley et al., 2011) og generell dødelighet (FitzGerald et al., 2004). For å finne ut mer om hvordan styrketrening påvirker kroppen og helse forskes det på faktorer som påvirker effekten av styrketrening. Dette kan være faktorer som treningsvolum, intensitet, hyppighet og øvelsesutvalg. Ved å endre disse variablene kan man endre det mekaniske og metabolske stresset på musklene og kroppen utsettes for (Ratamess et al., 2009; Toigo & Boutellier, 2006).

Mange i dagens samfunn lever i konstant tidsnød og dette kan hindre trening. Forskere ser derfor på hvor lav dose med trening er nødvendig for å oppnå ønsket respons (Choi et al., 2017). Schoenfeld et al. (2019) så på forskjellen i muskelvekst ved å utføre studie der de sammenlignet 1 sett mot tre sett i et styrketreningsprogram. I denne studien fant de ingen signifikant forskjell mellom gruppene som gjennomførte ett sett og gruppen som gjennomførte tre sett. De fant også at mengden trening følger en dose-respons-kurve der et høyere treningsvolum fører til økt hypertrofi, men ikke nødvendigvis like god utvikling i styrke. Flere andre studier har også sett på effekten av å trene ett sett versus tre sett (Galvão & Taaffe, 2005; Hass et al., 2000; Krieger, 2009; Radaelli et al., 2014), men resultatene fra disse studiene har vært sprikende.

I lys av sprikende resultater i forskning og et ønske om å videre belyse temaet om hvilken dose trening som må til for å oppnå en signifikant økning i styrke og hypertrofi så vil denne studien se på effekten av ett sett versus tre sett over en treningsperiode på 12 uker. Vi ligger opp til en hypotese om at gruppen som trener tre sett vil ha en bedre fremgang på styrke og hypertrofi etter 12 uker.

## Metode

### Deltagere

41 mannlige og kvinnelige deltagere ble rekruttert til studiet. Inklusjonskriteriene for å være med var at deltagerne måtte være mellom 18 og 40 år, ikke-røykende, ha en treningshistorie med minst en økt i uka med styrketrening det siste året og være fri for skader. Av de opprinnelige 41 deltagerne ble sju ekskludert på bakgrunn av at de ikke klarte å fullføre 85% av planlagte treningsøkter av ulike grunner: smerter i underekstremitet under trening ( $n = 5$ ), skade som ikke var relatert til studiet ( $n = 1$ ), klarte ikke å gjennomføre studieprotokoll ( $n = 1$ ).

### Intervensjon

Intervensjonen bestod av et 12-ukers styrketreningsprogram for hele kroppen, med trening 2 til tre ganger i uken. Øvelsene på bein ble gjennomført unilateralt for å kunne måle intern differensiering av ulike treningsvolum. Det ene benet trente ett sett, mens det andre benet trente tre sett. Hvilket ben som trente ett og tre sett ble randomisert. Før alle treningsøktene gjennomførte deltagerne en standardisert oppvarming med 5 minutter på ergometersykel (12-14 på Borg-skala) og 10 repetisjoner av 4 øvelser med kroppsvekt: individuelt tilpassede armhevinger, situps, rygghev og knebøy. Oppvarmingen ble fullført med ett sett med 10 repetisjoner på 50% av 1RM i hver øvelse i programmet.

Treningsprogrammet som alle deltagere gjennomførte var likt, eneste forskjell var antall sett som ble gjennomført på høyre eller venstre side. Øvelsene som ble gjennomført ble gjennomført i en standardisert rekkefølge: ettbeins beinpress, knefleksjon og kneekstensjon. Det benet som bare gjennomførte ett sett gjennomførte sitt sett mellom andre og tredje sett for motsatt ben. Etter beinøvelsene gjennomførte deltagerne øvelser på overkroppen i følgende rekkefølge: benkpress med manualer, nedtrekk og skulderpress eller sittende roing (annenhver økt). Treningsintensiteten var progressiv gjennom intervensjonsperioden, fra 10RM de første to ukene, til 8RM de neste tre ukene og 7RM de siste sju ukene. For å bidra til god restitusjon ble deltagerne gitt en standardisert drikk bestående av 0.15g/kg kroppsvekt med protein, 11.2g/kg kroppsvekt med karbohydrater og 0.5g/kg kroppsvekt med fett. ### Testing #### Styrketester Styrketester ble gjennomført ved pre og post. Isokinetiske og isometriske unilaterale øvelser på kneekstensjon ble gjort med dynamometer (Cybex 6000, Cybex International, Medway USA). Maksimalt isokinetisk (isok) dreiemoment ble målt på tre ulike vinkelhastigheter (60o, 120o og 240o s<sup>-1</sup>). Isometrisk (isom) styrke ble målt i 30o graders vinkel (full ekstensjon = 90o). Deltagerne fikk 2 forsøk på «isok60» og «isom60» og tre forsøk på «isok120» og «isok240», høyeste verdi ble registrert. 1RM ble målt i unilateral beinpress og kneekstensjon.

Oppvarming var standardisert med 5 minutter på ergometersykel, standardiserte oppvarmingsrepetisjoner i dynamometeret. I 1RM testen ble det gjennomført spesifikk oppvarming til hver øvelse med 10, 6 og 3 repetisjoner på 50, 75 og 85% av forventet 1RM. Deretter

ble 1RM funnet ved at vekten økte gradvis helt til deltagerne ikke lenger klarte å fullføre en hel repetisjon. Den høyeste vekten deltagerne løftet i de ulike øvelsene ble registrert som 1RM. Deltagerne fikk 4-6 forsøk. Resultatene ble regnet om til en kombinert score som et gjennomsnitt av alle styrketestene.

## Måling av kroppssammensetning

Ved pre og post-test ble det målt kroppssammensetning med «dual-energy X-ray absorptiometry» (DXA) (Lunar Prodigy, GE Healthcare, Oslo, Norway), i henhold til standardisert protokoll. Deltagerne fikk instruksjon om å komme fastende (2 timer) og frastå fra tung fysisk aktivitet (48 timer) før måling.

## Statistikk

Statistisk analyse er gjennomført i RStudio (versjon 2022.07.2.576) for Windows. Deskriptiv statistikk er skrevet som gjennomsnitt og standardavvik (mean(SD)), oppgitt med og uten prosent. Statistiske modeller og tabeller er skrevet på engelsk for å unngå begrepsblandinger. For å tolke effekten av treningsvolum (ett eller tre sett) på muskelhypertrofi og muskelstyrke ble det brukt en parett t-test til å regne ut p-verdi på endringsscore fra pre- til post-test. Statistisk signifikans er satt ved  $P < 0.05$ .

## Resultater

### Kroppssammensetning

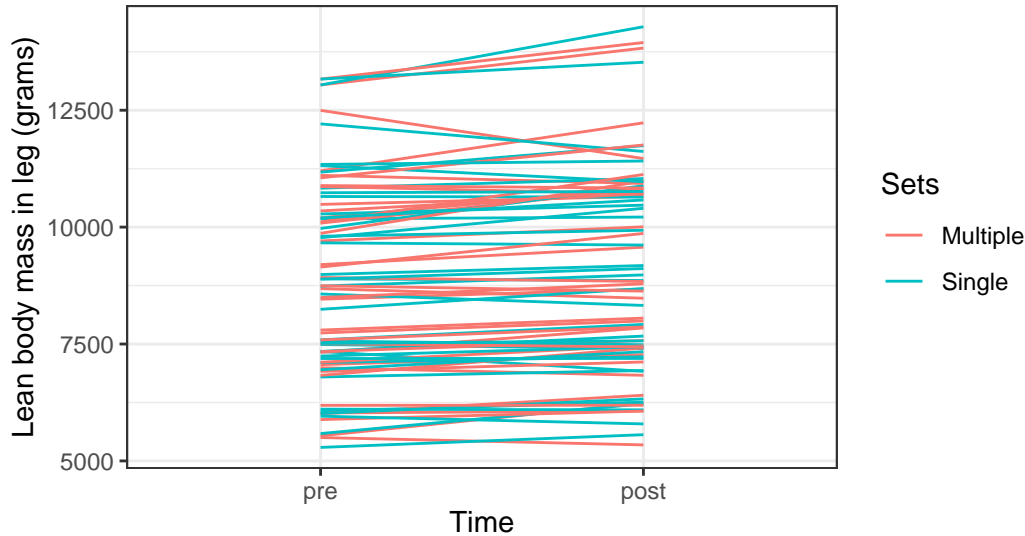
Resultatene i studien viser begge gruppene at 12 uker med styrketrening førte til en signifikant endring i fettfri masse i bein som gjennomførte tre sett 3.32% ( $\pm 4.39\%$ ) sammenlignet med bein som gjennomførte ett sett 2.04% ( $\pm 3.71\%$ ). Statistisk signifikansverdi  $p < 0.05$ .

Table 1: Average change in lean body mass in legs  
Measured in grams and percentage

Sets	Mean(SD) in grams	Mean(SD) in %
3 sets	289(423)	3.32(4.39)
1 sets	174(342)	2.04(3.71)

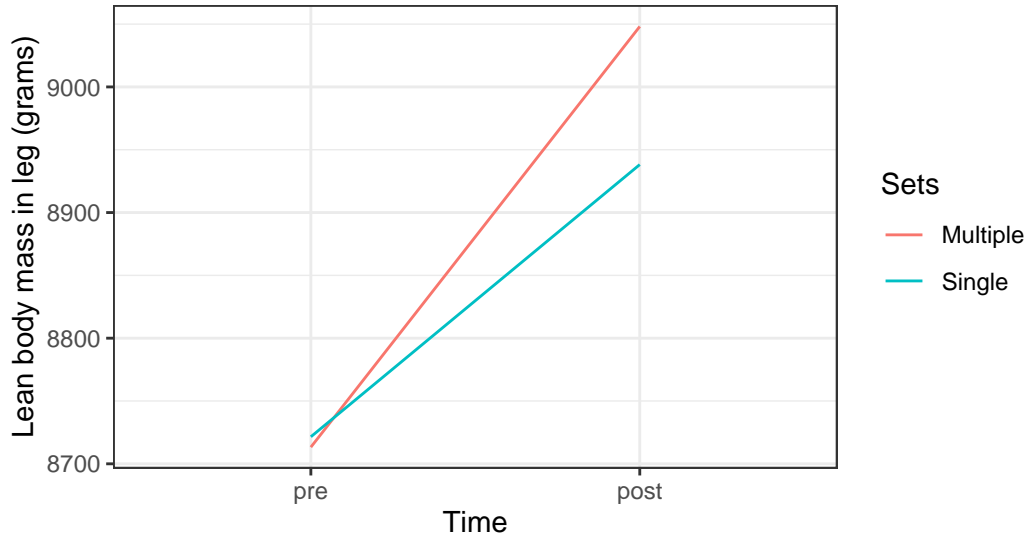
Mean (SD) increases in lean body mass for groups performing multiple sets and single sets.

Table 2: Change in lean body mass



Shows the difference between multiple and single sets for each participant, sorted by colour.

Table 3: Change in lean body mass



Show the difference between multiple and single sets for each group, sorted by colour



## Muskelstyrke

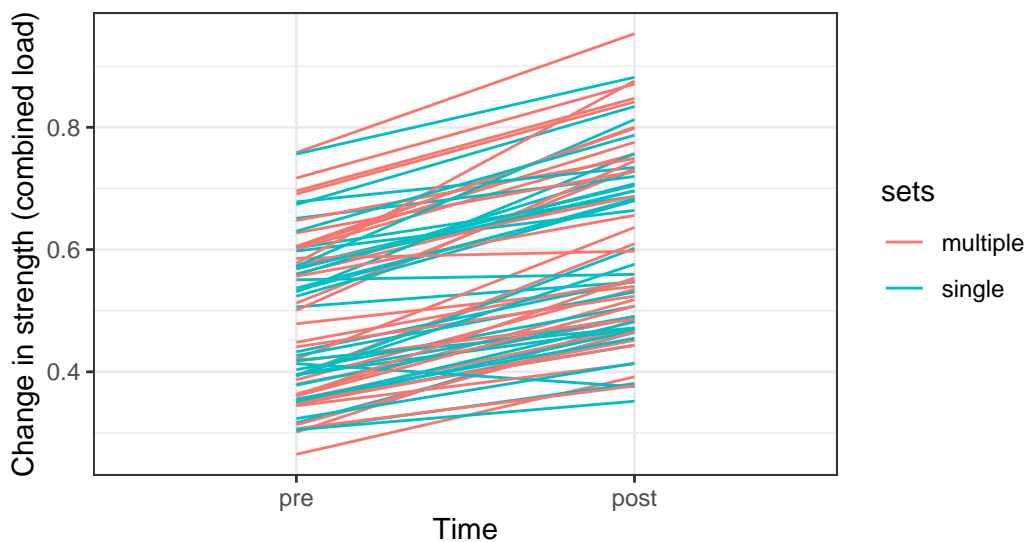
Resultatene av testene gjennomført på muskelstyrke viser at begge gruppene økte i styrke fra pre til post. Resultatene viser at bein som gjennomførte ett sett hadde en prosentvis økning i styrke på 24.5% ( $\pm 12.9\%$ ), bein som gjennomførte tre sett hadde en prosentvis økning i styrke på 31% ( $\pm 14.2\%$ ). Styrketestene er regnet om til en kombinert score (combined load) som et gjennomsnitt av alle styrketestene. Kombinert score viser et skåringstall uten måleenhet. Resultatene viser en signifikant forskjell i endring i muskelstyrke mellom gruppene. Statistisk signifikansverdi  $p < 0.05$ .

Table 4: Average change in strength from 3 and 1 sets  
Measured in percentage(SD)

Sets	% change(SD)
3 sets	31(14.2)
1 sets	24.5(12.9)

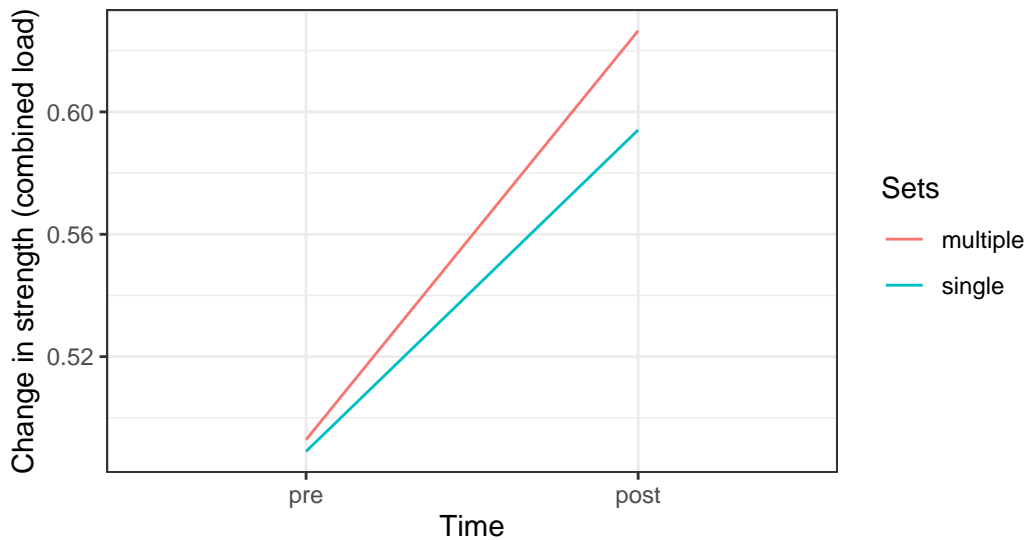
Average change strength in % (SD for 3 sets and 1 set).

Table 5: Change in strength



Shows the difference between multiple and single sets on strength change, for each participant.

Table 6: Change in strength



Shows the mean difference between multiple and single sets on strength change, per group.

## Diskusjon

Hovedfunnene i den gjennomførte studien viser oss at to til tre økter med styrketrening i uka over en 12-ukers treningsperiode ga økt styrke og muskelhypertrofi i beina. Å trene tre sett gir bedre effekt i form av økt styrke og hypertrofi sammenlignet med å trene ett sett med styrketrening over en 12-ukers treningsperiode. Resultatene samsvarer med det store mengder litteratur viser, (Galvão & Taaffe, 2005; Krieger, 2009; Krieger, 2010; Schoenfeld et al., 2016) at tre sett har større effekt enn ett sett med styrketrening på muskelstyrke og muskelmasse.

Brigatto et al. (2022) fant også en sammenheng med at flere sett var gunstig for å oppnå større grad av muskelvekst og økt muskelstyrke. Forfatterne fra denne studien finner også en statistisk signifikant økning på høyt volum og økt styrke i bein, sammenlignet med lavere volum. Dette er også i tråd med Hass et al. (2000) som finner signifikant økning med tre sett styrketrening sammenlignet med ett sett styrketrening på øvelser på underkroppen, men ikke på øvelser på overkroppen. Dette kan antyde at tre et høyere treningsvolum er å foretrekke for å øke styrke i underkroppen, men at kan være tilstrekkelig med ett sett på overkroppen. Dette stemmer også overens med konklusjon fra andre studier (Rønnestad et al., 2007; Schoenfeld et al., 2015). Det vil derfor være av interesse å reprodusere dette studiet, der man i tillegg til å trene forskjellig på høyre og venstre ben, også gjør det samme på høyre og venstre arm. Dette vil undersøke om man ser samme effekten som nevnte studier også med dette studiedesignet.

Når vi ser på økningen i muskelmasse, var denne også signifikant med et høyere treningsvolum.

Dette i tråd med funnene til Schoenfeld et al. (2019), som konkluderte med at muskelhypertrofi følger et dose-respons-forhold, der et økt treningsvolum førte til mer muskelhypertrofi sammenlignet med et mindre volum. Galvão & Taaffe (2005) kunne ikke trekke samme konklusjon på deres studie der de sammenlignet ett sett og tre sett på eldre mennesker, der de konkluderte med at trening med ett sett per øvelse er tilstrekkelig for å bedre fysisk prestasjonssevne. Studien vi har gjennomført skiller seg ut ved at hver enkelt deltager er sin egen kontroll ved at de trener høyre og venstre ben forskjellig.

## **Konklusjon**

Resultatene i studien stemmer med hypotesen om at 3 sett gir større effekt på styrke og muskelhypertrofi sammenlignet med ett sett styrketrening.

# Referanser

- Baechle, T. R., Earle, R. W., & Conditioning Association (U.S.), N. S. & (Eds.). (2008). *Essentials of strength training and conditioning* (3rd ed). Human Kinetics.
- Bird, A. (2020). Understanding the replication crisis as a base rate fallacy. *The British Journal for the Philosophy of Science*, 000–000. <https://doi.org/10.1093/bjps/axy051>
- Brigatto, F. A., Lima, L. E. de M., Germano, M. D., Aoki, M. S., Braz, T. V., & Lopes, C. R. (2022). High Resistance-Training Volume Enhances Muscle Thickness in Resistance-Trained Men. *The Journal of Strength & Conditioning Research*, 36(1), 22–30. <https://doi.org/10.1519/JSC.00000000000003413>
- Brooks, G. A. (1986). The lactate shuttle during exercise and recovery: *Medicine & Science in Sports & Exercise*, 18(3), 360–368. <https://doi.org/10.1249/00005768-198606000-00019>
- Bustin, S. A., & Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J Biomol Tech*, 15(3), 155–166.
- Choi, J., Lee, M., Lee, J., Kang, D., & Choi, J.-Y. (2017). Correlates associated with participation in physical activity among adults: A systematic review of reviews and update. *BMC Public Health*, 17(1), 356. <https://doi.org/10.1186/s12889-017-4255-2>
- Dahl, H. A. (2005). *Klar - ferdig - gå! Grunnbok i aktivitetsfysiologi* (2. opplag). Cappelen Akademisk Forlag.
- Ellefsen, S., Vikmoen, O., Zacharoff, E., Rauk, I., Slettaløkken, G., Hammarström, D., Strand, T. A., Whist, J. E., Hanestadhaugen, M., Vegge, G., Fagernes, C. E., Nygaard, H., Hollan, I., & Rønnestad, B. R. (2014). Reliable determination of training-induced alterations in muscle fiber composition in human skeletal muscle using quantitative polymerase chain reaction: Muscle fiber typing using qPCR. *Scand J Med Sci Sports*, 24(5), e332–e342. <https://doi.org/10.1111/sms.12185>
- Faude, O., Kindermann, W., & Meyer, T. (2009). Lactate Threshold Concepts. *Sports Medicine*, 39(6), 469–490. <https://doi.org/10.2165/00007256-200939060-00003>
- FitzGerald, S. J., Barlow, C. E., Kampert, J. B., Morrow, J. R., Jackson, A. W., & Blair, S. N. (2004). Muscular Fitness and All-Cause Mortality: Prospective Observations. *Journal of Physical Activity and Health*, 1(1), 7–18. <https://doi.org/10.1123/jpah.1.1.7>
- Galvão, D. A., & Taaffe, D. R. (2005). Resistance Exercise Dosage in Older Adults: Single-Versus Multiset Effects on Physical Performance and Body Composition: RESISTANCE-TRAINING VOLUME AND PERFORMANCE. *Journal of the American Geriatrics Society*, 53(12), 2090–2097. <https://doi.org/10.1111/j.1532-5415.2005.00494.x>
- Hass, C. J., Garzarella, L., De Hoyos, D., & Pollock, M. L. (2000). Single versus multiple sets in long-term recreational weightlifters: *Medicine & Science in Sports & Exercise*, 235. <https://doi.org/10.1097/00005768-200001000-00035>

- Hempel, C. G. (1966). *Philosophy of natural science*. Prentice Hall.
- Hopkins, W. G. (2000). Measures of Reliability in Sports Medicine and Science. *Sports Medicine*, 30(1), 1–15. <https://doi.org/10.2165/00007256-200030010-00001>
- Hulley, S. B. (Ed.). (2013). *Designing clinical research* (4th ed). Wolters Kluwer/Lippincott Williams & Wilkins.
- Hurley, B. F., Hanson, E. D., & Sheaff, A. K. (2011). Strength Training as a Countermeasure to Aging Muscle and Chronic Disease: *Sports Medicine*, 41(4), 289–306. <https://doi.org/10.2165/11585920-000000000-00000>
- Husøy, A.-M. (2022). Laktat. In *Store medisinske leksikon*. <https://sml.snl.no/laktat>
- Kraemer, W. J., & Ratamess, N. A. (2004). Fundamentals of Resistance Training: Progression and Exercise Prescription: *Medicine & Science in Sports & Exercise*, 36(4), 674–688. <https://doi.org/10.1249/01.MSS.0000121945.36635.61>
- Krieger, J. W. (2009). Single versus multiple sets of resistance exercise: A meta-regression. *The Journal of Strength & Conditioning Research*, 23(6), 1890–1901. <https://doi.org/10.1519/JSC.0b013e3181b370be>
- Krieger, J. W. (2010). Single vs. Multiple sets of resistance exercise for muscle hypertrophy: A meta-analysis [Journal Article]. *J Strength Cond Res*, 24(4), 1150–1159. <https://doi.org/10.1519/JSC.0b013e3181d4d436>
- Kuang, J., Yan, X., Genders, A. J., Granata, C., & Bishop, D. J. (2018). An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research. *PLOS ONE*, 13(5), e0196438. <https://doi.org/10.1371/journal.pone.0196438>
- Mangine, G. T., Hoffman, J. R., Gonzalez, A. M., Townsend, J. R., Wells, A. J., Jajtner, A. R., Beyer, K. S., Boone, C. H., Miramonti, A. A., Wang, R., LaMonica, M. B., Fukuda, D. H., Ratamess, N. A., & Stout, J. R. (2015). The effect of training volume and intensity on improvements in muscular strength and size in resistance-trained men. *Physiol Rep*, 3(8), e12472. <https://doi.org/10.14814/phy2.12472>
- Mangine, G. T., Hoffman, J. R., Wang, R., Gonzalez, A. M., Townsend, J. R., Wells, A. J., Jajtner, A. R., Beyer, K. S., Boone, C. H., Miramonti, A. A., LaMonica, M. B., Fukuda, D. H., Ratamess, N. A., & Stout, J. R. (2016). Resistance training intensity and volume affect changes in rate of force development in resistance-trained men. *Eur J Appl Physiol*, 116(11-12), 2367–2374. <https://doi.org/10.1007/s00421-016-3488-6>
- Okasha, S. (2016). *Philosophy of science: a very short introduction* (Second edition., Vol. 67). University Press.
- Popper, K. R. (1969). *Conjectures and refutations: The growth of scientific knowledge* (3rd ed. (revised)). Routledge & K. Paul.
- Radaelli, R., Wilhelm, E. N., Botton, C. E., Rech, A., Bottaro, M., Brown, L. E., & Pinto, R. S. (2014). Effects of single vs. Multiple-set short-term strength training in elderly women. *AGE*, 36(6), 9720. <https://doi.org/10.1007/s11357-014-9720-6>
- Ratamess, N. A., Alvar, B. A., Evetoch, T. K., Housh, T. J., Kibbler, B., & Kraemer, W. J. (2009). Progression Models in Resistance Training for Healthy Adults. *Medicine & Science in Sports & Exercise*, 41(3), 687–708. <https://doi.org/10.1249/MSS.0b013e3181915670>
- Rønnestad, B. R., Egeland, W., Kvamme, N. H., Refsnes, P. E., Kadi, F., & Raastad, T. (2007).

- Dissimilar effects of one- and three-set strength training on strength and muscle mass gains in upper and lower body in untrained subjects. *Journal of Strength and Conditioning Research*, 21(1), 157–163. <https://doi.org/10.1519/00124278-200702000-00028>
- Ruiz, J. R., Sui, X., Lobelo, F., Morrow, J. R., Jackson, A. W., Sjostrom, M., & Blair, S. N. (2008). Association between muscular strength and mortality in men: Prospective cohort study. *BMJ*, 337(jul01 2), a439–a439. <https://doi.org/10.1136/bmj.a439>
- Schoenfeld, B. J., Contreras, B., Krieger, J., Grgic, J., Delcastillo, K., Belliard, R., & Alto, A. (2019). Resistance Training Volume Enhances Muscle Hypertrophy but Not Strength in Trained Men. *Medicine & Science in Sports & Exercise*, 51(1), 94–103. <https://doi.org/10.1249/MSS.0000000000001764>
- Schoenfeld, B. J., Ogborn, D., & Krieger, J. W. (2017). Dose-response relationship between weekly resistance training volume and increases in muscle mass: A systematic review and meta-analysis. *Journal of Sports Sciences*, 35(11), 1073–1082. <https://doi.org/10.1080/02640414.2016.1210197>
- Schoenfeld, B. J., Ogborn, D., & Krieger, J. W. (2016). Dose-response relationship between weekly resistance training volume and increases in muscle mass: A systematic review and meta-analysis [Journal Article]. *J Sports Sci*, 1–10. <https://doi.org/10.1080/02640414.2016.1210197>
- Schoenfeld, B. J., Peterson, M. D., Ogborn, D., Contreras, B., & Sonmez, G. T. (2015). Effects of Low- vs. High-Load Resistance Training on Muscle Strength and Hypertrophy in Well-Trained Men. *Journal of Strength and Conditioning Research*, 29(10), 2954–2963. <https://doi.org/10.1519/JSC.0000000000000958>
- Statistisk sentralbyrå. (2019). *Styrketrening stadig mer populært*. <https://www.ssb.no/kultur-og-fritid/artikler-og-publikasjoner/styrketrening-stadig-mer-populaert>
- Stone, M. H., Pierce, K. C., Sands, W. A., & Stone, M. E. (2006). Weightlifting: A Brief Overview. *J Strength Cond*, 28(1), 50. [https://doi.org/10.1519/1533-4295\(2006\)28%5B50:WABO%5D2.0.CO;2](https://doi.org/10.1519/1533-4295(2006)28%5B50:WABO%5D2.0.CO;2)
- Toigo, M., & Boutellier, U. (2006). New fundamental resistance exercise determinants of molecular and cellular muscle adaptations. *Eur J Appl Physiol*, 97(6), 643–663. <https://doi.org/10.1007/s00421-006-0238-1>
- Wackerhage, H. (Ed.). (2014). *Molecular exercise physiology: An introduction*. Routledge, Taylor & Francis Group.