

# **Eksamen IDR4000**

110

12/2/22

# Table of contents

|   |           |
|---|-----------|
| <b>Preface</b>  | <b>4</b>  |
| <b>Deloppgave 1: Beskrivende statistikk, reliabilitet og validitet, verktøy for reproduserbar dataanalyse</b> | <b>5</b>  |
| Statistikk . . . . .  | 6         |
| Deskriptive data . . . . .  | 6         |
| Resultater . . . . .  | 7         |
| Konklusjon . . . . .  | 8         |
| <b>Deloppgave 2: Laborasjonsrapport</b>   | <b>10</b> |
| Introduksjon . . . . .  | 10        |
| Metode . . . . .  | 10        |
| Resultater . . . . .  | 11        |
| <b>Deloppgave 3: Vitenskapsfilosofi</b>   | <b>12</b> |
| 1. Falsifikasjonisme . . . . .  | 12        |
| 2. HD-metoden og abduksjon/Bayesianisme . . . . .   | 13        |
| 3. Replikasjonskrisen . . . . .   | 13        |
| <b>Deloppgave 4: Studiedesign</b>   | <b>16</b> |
| Introduksjon . . . . .  | 16        |
| Studiedesign . . . . .  | 17        |
| Statistiske analyser . . . . .  | 18        |
| <b>Deloppgave 5: Analysere repeterte målinger</b>   | <b>20</b> |
| Introduksjon . . . . .  | 20        |
| Metode . . . . .  | 21        |
| Deltakere og studieoversikt . . . . .   | 21        |
| Intervensjonen . . . . .  | 21        |
| Testing av muskelstyrke . . . . .   | 21        |
| Måling av muskelmasse og kroppsammensetning . . . . .   | 22        |
| Statistikk . . . . .  | 22        |
| Resultater . . . . .  | 23        |
| Kroppsammensetning . . . . .  | 23        |
| Muskelstyrke (benpress) . . . . .   | 23        |
| Diskusjon . . . . .   | 23        |



# Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

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

## *Laktatprofil løp*

Vi har gjennomført laktattesting av til sammen 7 masterstudenter i aldersgruppen 23-50 år i løping på tredemølle. Det er gjennomført to tester (Pre og Post) av alle, men det er ikke gjennomført intervensjon i perioden mellom testene. Testen ble gjennomført som en laktatprofil med en standardisert trappetrinnsprotokoll der farten økte med 1,5km/t per drag til forsøkspersonen passerte en laktatkonsentrasjon høyere enn 4mmol/L. Grunnet tekniske utfordringer, ble testprotokollen (fartsøkingsintervallet) endret etter testing av to individer. Disse to er imidlertid testet likt begge ganger selv om hastighetene ikke var lik på alle trappetrinn som resten av gruppen.

Det er tilstrebet å redusere antall ytre faktorer som kan påvirke testresultatene gjennom å standardisere forberedelser av testpersonene og testsituasjonen.

Utstyr benyttet til analyse har vært det samme gjennom all testing, og er kalibrert mellom hver test. Bionsen er benyttet for analyse av laktat og Jaeger Oxycon Pro analyserer ventilasjon, puls, O<sub>2</sub> forbruk.

Det er kjent at hydrerings- og ernæringstilstand påvirker testresultater. Redusert energitilgjengelighet (<30 kcal /kg fettfri masse/dag) forventes å påvirke metabolismen og prestasjon negativt. Inntak av karbohydrat i forkant av og under test, medfører økt glykogen tilgjengelighet. Lavt inntak kan medføre hypoglykemi og dårligere prestasjon. Koffeininntak i forkant av test har vist økt prestasjon, men hvis man er vant til å innta koffein daglig, bør dette også gjøres testdag, da abstinenser kan medføre hodepine og tretthet. Alkoholinntak kan påvirke hydreringstilstand og karbohydratmetabolismen og derved gi redusert prestasjon (Tanner & Gore, 2012).

Testpersonene i undersøkelsen er derfor forberedt og instruert i å trene, spise og sove likt før testene. Testutstyr som sko og tøy har også vært det samme under testing.

Ut over forberedelse av testperson, er det forhold ved testsituasjonen som kan påvirke resultatet. Temperatur og luftfuktighet bør tilstrebes å være lik under all testing. Andre kjente faktorer som kan påvirke testresultater er testpersonens kjennskap til testforløpet, tilbakemeldinger testpersonen får underveis, antall testpersoner i lokalet og kjønnsfordeling på

disse, “dagsform” eller mentalt overskudd hos testpersonen og om testpersonen hører musikk før og under testing (**halperin2015?**).

Testing er derfor tilstrebet gjennomført med: Testtidspunkt samme tid på døgnet. Samme testleder på pre og post test. Samme antall og kjønn tilstede ved alle tester. Ingen musikk før/under test. Tilnærmet lik tilbakemelding og engasjement fra testleder. Det er opplyst om videre hendelsesforløp før hver endring (forklart test, tempøkninger, tid igjen osv.) Det er ikke opplyst om puls, laktat og VO2. Notert ned evt. når og hvor mye forsøkspersonen drikker/spiser og repetert ved post-test

Møllen har vært innstilt på 1 % stigning, og det er kjørt 5 minutters drag med 1 minuts pause. Utgangshastighet var på 8,5 km/t, og denne ble økt med 1,5km/t per drag til FP målte over 4 mmol/L Lac.

Under testing, er data samlet inn via analysemaskin for laktat (Bionsen) og Jaeger Oxycon Pro, som analyserer ventilasjon, puls, O2 forbruk. Data er hentet ut i etterkant. Innsamlede data er i tillegg sikret løpende manuelt i plotteskjema. Alle data er etterfølgende lagt inn i et excelark (Data\_run\_2) før videre analyse.

## Statistikk

Statistisk analyse er utført i RStudio 2022.07.2. Det er beregnet mean, median, standardavvik og varians av maksimale verdier pre og post. Det er beregnet standardavvik, typical error, og variasjonskoeffesient av differansen.

## Deskriptive data

Rows: 14

Columns: 12

```
$ FP          <dbl> 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 7
$ Gender      <chr> "Male", "Male", "Female", "Female", "Male", "Male", "Male", ~
$ Timepoint   <chr> "Pre", "Post", "Pre", "Post", "Pre", "Post", "Pre", "Post", ~
$ Age        <dbl> 24, 24, 50, 50, 23, 23, 29, 29, 25, 25, 23, 23, 23, 23
$ Height     <dbl> 195, 195, 169, 169, 175, 175, 180, 180, 191, 191, 171, 171, ~
$ Weight     <dbl> 86.0, 86.0, 56.1, 56.1, 81.9, 81.9, 93.4, 93.1, 80.7, 79.9, ~
$ start.load <dbl> 10.0, 13.0, 8.5, 8.5, 8.5, 8.5, 12.0, 10.0, 8.5, 8.5, 8.5, ~
$ end.load   <dbl> 19.0, 19.0, 11.5, 11.5, 16.0, 14.5, 15.0, 15.0, 14.5, 14.5, ~
$ VO2.max    <dbl> 5293, 5224, 2239, 2290, 4525, 3937, 4527, 4613, 4179, 4201, ~
$ lac.max    <dbl> 8.31, 5.33, 5.68, 5.20, 6.14, 5.00, 4.47, 4.12, 7.45, 5.41, ~
$ HR.max     <chr> "196", "201", "168", "169", "182", "180", "180", "175", "NA~
$ RER.max    <dbl> 1.00, 1.01, 0.99, 0.96, 1.01, 1.03, 0.96, 0.99, 1.10, 1.20, ~
```



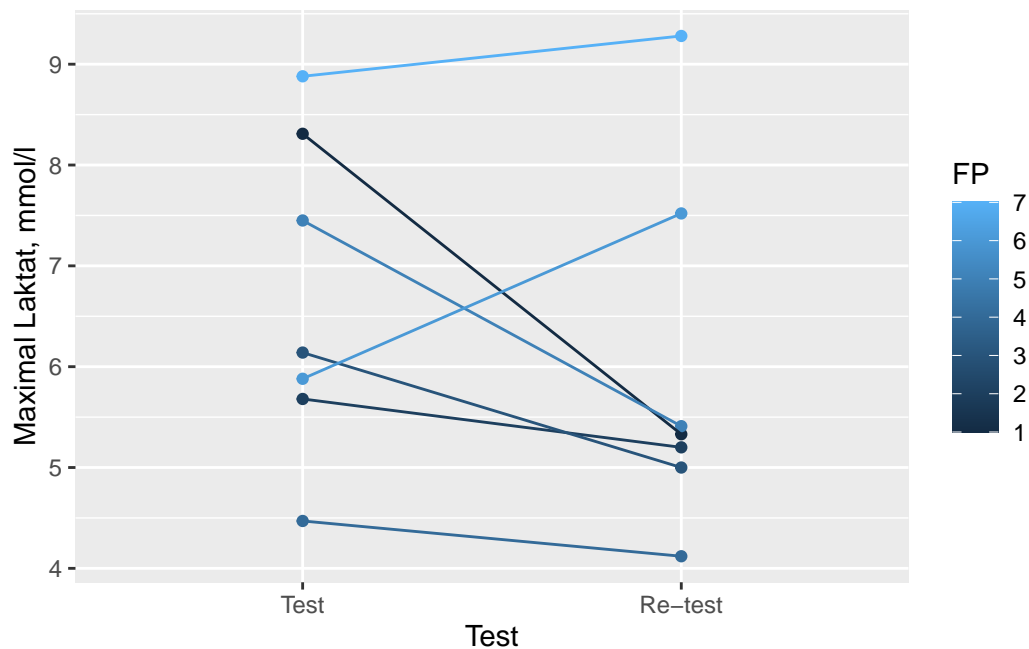
|      |      |      |       |     |
|------|------|------|-------|-----|
| Post | 1.78 | 0.08 | 1,025 | 2.8 |
| Pre  | 1.58 | 0.04 | 1,095 | 2.9 |

SD av maksimale verdier

## VARIANS

| Test | Max Laktat | Max RER | VO2max(ml min-1) | Max fart(km t-1) |
|------|------------|---------|------------------|------------------|
| Post | 3.17       | 0.01    | 1,049,849        | 8.0              |
| Pre  | 2.48       | 0.00    | 1,198,619        | 8.6              |

```
# A tibble: 1 x 4
  sd    te    m    cv
<dbl> <dbl> <dbl> <dbl>
1  1.53  1.08  6.33  17.1
```



## Konklusjon

Vi har valgt å gjøre beregninger ut fra høyeste målte laktat pr test. Test ble ikke gjennomført helt i henhold til protokoll, grunnet tidsnød i lab. Flere av forsøkspersonene ble testet rett



hverandre, med kun 10 minutter mellom testene. Protokollen ble endret underveis, og selv om dette ikke skal påvirke forholdet mellom pre- og post-testingen er det en svakhet ved forsøket. Konstant utskiftning av testpersonell også en sentral faktor til feilkilde. Når testene utføres av forskjellig testpersonell vil dette kunne gi utslag på resultatene, og være en feilkilde for spesielt “within-subject variation” (**hopkins2000?**). Eksempelvis vil resultatet på laktatprøvene avhenge av hvordan testeren tar selve prøven, hvor lang tid som brukes, og om prøven utføres korrekt. I vårt forsøk er Typical Error (TE) beregnet til 1,08. Typical Error vil indikere forskjellen i avviket fra gjennomsnittet i to gruppene. Standardavvik (sd) er beregnet til 1,53 og variasjonskoeffisient (cv) til 17%. Standardavviket er avviket fra gjennomsnittet. Variasjonskoeffesienten kan beskrives som standardavviket i forhold til gjennomsnittet i prosent. 17% anses som en stor variasjon, noe som tyder på lav reliabilitet.

### Referanser

Tanner, R. K., and C. J. Gore. 2012. Physiological Tests for Elite Athletes 2nd Edition. Book. Human Kinetics. <https://books.google.no/books?id=0OPIiMks58MC>

# Deloppgave 2: Laborasjonsrapport

## Introduksjon

Analyser av genuttrykk gjennom kvantitativ PCR i skjelettmuskulatur er rutine i treningsstudier (**kuang\_overview\_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\_overview\_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\_overview\_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 Kristian Lian. Kristian 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\_molecular\_2014?). @

## Resultater

Table 1 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         |

# Deloppgave 3: Vitenskapsfilosofi

## 1. Falsifikasjonisme

Poppers falsifiserbarhetskriterium er enkelt forklart at vitenskapelige teorier må kunne falsifiseres, altså motbevises, for å kunne kategoriseres som vitenskapelige. Bakgrunnen for dette er at oppfatningen av et sterkt induktivt argument ikke stemmer, og at deduksjon er argumentet som burde brukes i vitenskap (**Popper1969?**). Vitenskapelige teorier skal ifølge Popper altså ikke underbygges av data, blant annet fordi man da kan modifisere en teori for å få den til å stemme med dataene. Dermed kan man alltid justere teorien etter dataene slik at den blir nærmest umulig å tilbakevise. Om en teori ikke er falsifiserbar, altså ikke kan tilbakevises, er den uvitenskapelig ifølge Popper, også kjent som pseudovitenskap. Okasha og andre vitenskapsfilosofer er uenig i Poppers syn, og mener at en sterk vitenskapelig teori trenger bekreftelse, noe som krever induktive argumenter. Eksempelvis tester kanskje forskere en teori gjennom en studie eller eksperiment, men hvis dataene her gir grunnlag for å endre eller modifisere hypotesen, er det kanskje nettopp det som behøves for å gjøre videre fremskritt. Et eksempel på dette er Adams og Leverriers oppdagelse av Uranus, som tok utgangspunkt i Newtons teorier. De fant senere ut at den ikke var helt korrekt på et område, som igjen førte til oppdagelsen av en ny planet (**okasha2016?**). Ifølge Popper er dette pseudovitenskap, mens Okasha mener dette i høyeste grad er vitenskap. Det legges mer vekt på hvor godt en teori er bekreftet, og skille mellom hva som er vitenskapelig og ikke lagt like mye vekt på. Selv om både Popper og Okasha har flere gode argumenter for sine syn, må jeg nok si meg mest enig med Okasha. For det første vil det være svært vanskelig å komme utenom induktiv metode i forskning på veldig mange områder. I et perfekt scenario hadde utelukkende bruk av deduktiv metode vært det beste, da konklusjonen er garantert og risikoen for feil liten. Men i realiteten er det vanskelig å komme uten induktivismen. Det ville også vært uheldig om forskere måtte forkaste alle teorier hvis den ble motbevist i et tilfelle, mens den er bekreftet i x-antall andre tilfeller. Her har man jo også eksempelvis meta-analyser som sammenfatter forskning og studier på et område, viser hovedtrekkene og luker ut de svakeste studiene. Hvis vitenskapen har rom for å være oppdagende, gir det også rom for nye oppdagelser, selv om teorien som var utgangspunktet ikke var helt korrekt på alle områder. (**okasha2016?**)

## 2. HD-metoden og abduksjon/Bayesianisme

Strukturen på et bekreftende vitenskapelig argument ifølge den hypotetisk deduktive metoden (HD-metoden) er delt inn i en oppdagende- og en begrunnende del. Den oppdagende delen kan betegnes som en kvalifisert gjetning, og ifølge Hempel som sto bak metoden, er det ikke noen rett eller gal måte å gjøre dette på. Dette begrenses bare av vitenskapsmenns oppfinnsomhet og fantasi, og det trengs ingen videre argumenter eller begrunnelse i dette stadiet (Dellsén, 2022). Men, for at en teori skal være mer enn en kvalifisert gjetning, kreves det en begrunnende del. Teorien, eller hypotesen, må utlede noen empiriske konsekvenser som kan bekrefte (eller avkrefte) teorien (**hempel1966?**). Et eksempel på dette kan være hypotesen om at to terskeløkter på samme dag gir bedre anaerob laktattterskel enn en terskeløkt om dagen. Ut fra dette kan man dedusere empiriske konsekvenser, som igjen kan etterprøves. I dette eksempelet vil de empiriske konsekvensene være at utøveren får en høyere terskelfart, lavere laktat på samme belastning sammenlignet med før studien og høyere fart på samme laktatverdi. Bekreftes disse, bekreftes teorien induktivt, til en viss grad. Den kan ikke garantere at teorien stemmer grunnet at den er bekreftet induktivt, men jo mer eller fler empiriske konsekvenser som bygger opp under teorier, jo sterke er teorien bekreftet. I praksis deler vi inn HD-metoden i 4 trinn (**hempel1966?**). Trinn 1 er å formulere hypotesen, som nevnt tidligere en kvalifisert gjetning. Trinn 2 er å utlede de empiriske konsekvensene som medfølger teorien, som altså må stemme om teorien skal være sann. I trinn 3 må man teste og etterprøve de empiriske konsekvensene, som oftest gjennom en studie eller et eksperiment. Trinn 4 er der man eventuelt får den induktive bekræftelsen. Viser det seg at de empiriske konsekvensene stemmer, bekreftes teorien induktivt til en viss grad. Abduksjon er en annen form for vitenskapelig bekræftelse som er ment for å løse svakhetene til HD-metoden, også kjent som slutningen til den beste forklaringen (Dellsén, 2022). Her sammenlignes ulike teorier som skal forklare relevante data, der den beste teorien velges. Kriterier som legges vekt på i abduksjon er forklaringskraft og enkelhet. Med forklaringskraft menes hvor godt teorien forklarer dataene, i likhet med HD-metoden. En teori må kunne forklare eksempelvis et fenomen bedre enn andre teorier. Men fins det flere teorier som har lik forklaringskraft, er det den enkleste teorien som foretrekkes i abduksjon. I praksis betyr dette at hvis to teorier forklarer empiriske data like godt, er det den enkleste teorien som er den beste. Denne tankegangen baserer seg på ideen fra William Ockham, og den såkalte Ockhams barberkniv. Ockhams barberkniv går ut på at unødvendige momenter i en forklaring elimineres. Fordelen med dette er at unødvendige momenter i en forklaringsmodell kan ha en høyere sannsynlighet for å være feil, nettopp fordi de er unødvendige.

## 3. Replikasjonskrisen

Alexander Bird mener at basefrekvensfeilen (The base rate fallacy) er årsaken til at mange resultat i noen vitenskaper ikke repliseres. Basefrekvensfeilen oppstår ofte i scenarioer der sannsynlighet for visse fenomener er inne i bildet, som følge av at forskere fokuserer ensrettet

på tilstedeværelsen av bevis og neglisjerer sannsynligheten for at fenomenet ville oppstå uten beviset (**bird2021?**). Dette fører, ifølge Bird, til feilaktige konklusjoner da bevisene er sterke, men feilaktig korrelert med forekomsten. Bird bruker eksempler der sannsynligheten for å ha f. eks en sykdom er lav, eksempelvis 0,01%, og at testen at testen for å påvise sykdommen er 95% sikker. Hvis man da tester positivt skulle man kanskje ved første øyekast si at det er 95% sannsynlighet for at personen har sykdommen, men sannheten er faktisk at sannsynligheten for at personen som testet positivt faktisk har sykdommen er 1,9%. Dette fordi sannsynligheten for å få et falskt positivt svar er mye høyere enn et korrekt positivt svar når sykdommen er så sjelden som 0,01% (**bird2021?**). Hvordan basefrekvensfeilen fører til replikasjonskrisen ifølge Bird er ved at mange hypoteser som klassifiseres som riktige gjennom forskningen er falske positive (type-1 feil). I tillegg trekker Bird fram et poeng om at den statistiske signifikansen holder et for lavt nivå. Hvis vi for eksempel tar utgangspunkt i at 90% av testede hypoteser er usanne, er basefrekvensen til de sanne hypotesene så lav som 10%. Birds analyse baserer seg i stor grad på Bayesiansk sannsynlighetsregning. I mange forskningsfelt er 0,05 satt som P-verdi for statistisk signifikans (**bird2021?**), og kan enkelt forklart sies å være sannsynligheten for å få type-1 feil. Statistisk styrke er derimot sannsynligheten for å ikke få et falskt negativt resultat, type-2 feil (Dellsén, 2022). Kravet til statistisk signifikans er ofte satt til 80%. Regner vi ut sannsynligheten for eksempelet over er sannsynligheten for statistisk signifikans kun 64%. Altså kan replikasjonskrisen forklares som en kombinasjon av at flertallet av hypotesene er usanne kombinert med at kravet til statistisk signifikans er for lavt. I sin artikkel diskuterer også Bird flere andre mulige forklaringer på replikasjonskrisen. Blant disse finner vi lav statistisk styrke, tvilsomme forskningspraksiser, at negative resultater ikke publiseres, dårlige incentiver og til og med forskningsjuks (**bird2021?**). At lav statistisk styrke, som gir større rom for type-2 feil er forklaringen, er Bird uenig i. Han peker på at problemet ikke forsvinner selv om man øker den statiske styrken, som i eksempelet i forrige avsnitt. Med tvilsomme forskningspraksiser mener Bird fabrikkering og/eller falsifisering av data, f. eks manipulering av P-verdier. Han anerkjenner at problemet eksisterer, men stiller likevel spørsmålsteget ved hvor stor grad dette kan være årsaken til falske positive hypoteser og replikasjonskrisen. Det finnes bevis, men ikke tilstrekkelig til å trekke en tydelig konklusjon (**bird2021?**). Det kan også være andre ting som ikke er direkte fusk med dataene, men eksempelvis små studier med få forsøkspersoner som gir lav statistisk styrke (Dellsén, 2022), noe som gjør at resultatet er mindre reproducerbart enn ved større studier med flere forsøkspersoner. Det er også en skeivfordeling mellom andelen positive og negative resultat som publiseres, og andelen av publisert forskning har en overvekt av positive funn. Dette er kanskje ikke så rart ettersom det er det man ønsker å finne, men dette kan bidra til å gi et falskt bilde på virkeligheten. Et negativt resultat er også et resultat, og når dette ikke publiseres kan det være med å bygge opp under den såkalte replikasjonskrisen. Jeg mener Bird argumenterer godt for sitt synspunkt på replikasjonskrisen, og er i mer eller mindre grad enig med han. Årsakene til at studier mislykkes å repliseres synes å være forårsaket av basefrekvensfeil heller enn at studiene var feil eller dårlig gjennomført (**bird2021?**). I tillegg legger han ikke skjul på at andre mulige forklaringer også spiller en rolle, men presiserer samtidig tydelig at disse i seg selv ikke gir en tilstrekkelig forklaring. Men på en annen side, er det også noen gode argumenter som taler imot. Spesielt når det gjelder at lav statistisk styrke ser det kanskje

ut til at P-grenseverdi på 0.05 er betryggende, mens mange mener den burde være betydelig lavere. Selv er Bird åpen for at den kanskje burde vært så lav som 0.005 på flere forskningsfelt. Størrelsen på studier trekkes også fram til en årsak for replikasjonskrisen, noe begge sider av debatten virker å anerkjenne. Om Bird har rett i at hans forklaring er bedre er vanskelig å kunne si helt sikkert, men argumentene underbygges godt og han neglisjerer ikke argumentene til den andre siden. Jeg vil derfor kunne påstå at hans forklaring er bedre.

Kilder: • Popper, K. (1969). *Conjectures and Refutations: The Growth of Scientific Knowledge* (3. utg.). London: Routledge • Okasha, S. (2016). *Philosophy of Science: A Very Short Introduction* (2nd ed.). Oxford: Oxford University Press. • Hempel, C.G. (1966). *Scientific Inquiry: Invention and Test*. I C. G. Hempel, *Philosophy of Natural Science* (s.3-18). Princeton: Prentice Hall. • Bird, A. (2018). “Understanding the Replication Crisis as a Base Rate Fallacy”, *British Journal for the Philosophy of Science*. DOI: 10.1093/bjps/axy051 • Dellsén, F. (2022). Power-point «KvantMet2022, Vitenskapsfilosofi». Canvas. o 3. KvantMet2022 Vitenskapelige teorier.pdf o 4. KvantMet2022 Vitenskapelige argument.pdf o 5. KvantMet2022 Abduksjon og sannsynlighet.pdf o 6. KvantMet2022 Sannsynlighet og replikasjonskrisen.pdf

# Deloppgave 4: Studiedesign

## Introduksjon

I denne oppgaven har jeg valgt ut 5 studier for å se på studiedesign og statistiske analyser, vurdere styrker og svakheter rundt dette og gi en anbefaling på hvordan framtidige studier bør utformes på bakgrunn av dette. Felles for studiene som er valgt ut, er at de sammenligner effekten av å bolklegge utholdenhetstrening (BP; blokk-periodisering) sammenlignet med tradisjonell strukturering av utholdenhetstrening (TP; tradisjonell-periodisering). Studiene har sett på påvirkningen på ulike utholdenhetsparameter som Vo2-maks, laktatterskel gjennom laktatprofil, submaksimal og maksimal aerob effekt og også prestasjonstester av ulik karakter. I denne oppgaven fokuserer jeg på effekten av BP på Vo2-maks. Studiene er naturligvis ikke helt likt utformet, men felles er at det bolklegging av intensiv trening, dvs intensitetssone 4/5 på Olympiatoppens intensitetsskala for utholdenhetstrening (**OLT2021?**) (**OLT\_\_int2021?; mølmen2019?**). (Obs; studiene deler ofte opp i 3 intensitetssoner istedenfor 3, så % av makspuls bør benyttes for å sammenligne intensitet mellom studiene). Kontrollgruppene som har benyttet seg av tradisjonell trening har i de fleste studiene hatt 2 (til 3) ukentlige intervalløkter med intensiv trening.

Studiene jeg har valgt ut er;

1. «Effects of 12 weeks of block periodization on performance and performance indices in well-trained cyclists» (Rønnestad, Ellefsen, et al., 2014)
2. «Block periodization of high-intensity aerobic intervals provides superior training effects in trained cyclists» (Rønnestad, Hansen, et al., 2014)
3. «5-week block periodization increases aerobic power in elite cross-country skiers» (**5-weekb?**)
4. «No Differences Between 12 Weeks of Block- vs. Traditional-Periodized Training in Performance Adaptations in Trained Cyclists» (Almquist et al., 2022)
5. «Improved VO2max and time trial performance with more high aerobic intensity interval training and reduced training volume: a case study on an elite national cyclist» (Støren et al., 2012)

\*Studie 1 og 2 er basert på samme forskningsprosjekt av samme forskere og utvalg. Dog har studie 1 sett på effekten av BP over en 12-ukersperiode, mens studie 2 har sett på effekten etter 4 uker.



## Studiedesign

Det kan brukes flere former for studiedesign når man skal forske på effekten av BP på utholdenhetsprestasjon og parametere. Studiedesign kan deles inn i flere ulike kategorier, men innenfor idrett er intervensjons- og observasjonsstudier mye brukt (**thiese2014?**). Intervensjonsstudier benytter seg som navnet tilsier av en intervensjon og ser på effekten av denne, mens i en observasjonsstudie påvirker ikke forskerne deltakerne, men ser på eventuell sammenheng mellom faktorer og utfall (**thiese2014?**). Studiene er enten retrospektive, altså bruker data som allerede er innsamlet, og forskerne kan dermed ikke påvirke studien i denne fasen. Fordeler med denne formen kan være at den er kostnadsbesparende og i mange tilfeller kan man bruke data fra en stor populasjon, men samtidig kan det være unøyaktighet i datainnsamlingen og forskerne får ofte ikke påvirket datainnsamlingen som det ønsker (Hulley, 2013; Thiese, 2014). Ved prospektivt studiedesign følges deltakerne derimot framover i tid, og forskerne kan dermed designe forsøkene så de blir best mulige.

Innenfor idrettsforskning blir studier ofte designet som en randomisert-kontrollert studie (randomized controlled trial; RCT) for å sammenligne effekten til det som forskes på opp mot en kontrollgruppe før og etter en intervensjon (**thiese2014?**). Ved at deltakerne er randomisert sikrer forskerne mot bias, altså at deres eventuelle forutinntatthet påvirker studien. I tillegg kan forsøkene standardiseres for å sikre reproduserbarhet [ (Hulley et al., 2014). Av studier plukket ut til denne oppgaven er samtlige intervensjons- og prospektive studier. Dog er studie 1 til 4 RCT-er, mens studie 5 er en casestudie. Alle studier er pre-post studier, altså ser de på effekten etter intervensjon sammenlignet før (**thiese2014?**). De 4 første studiene har et to-gruppe design, med en BP-gruppe og en TP-gruppe (kontroll). Kontrollgruppen har gjennomført det som karakteriseres som tradisjonell utholdenhets trening, ofte referert som «best practice» i meta-analyser (**mølmen2019?**; **thiese2014?**), men med ulik utforming i studiene. Studiene kan derfor også kategoriseres som en form for aktivt kontrollerte studier da også kontrollgruppen har gjennomført en intervensjon (Hulley et al., 2013). Studiene gjennomførte hardøktene som høyintensiv intervalltrening, intensitetssone 4/5 (**OLT\_int2021?**) > 88% av HR-maks. Studie 4 hadde i tillegg intervalløkter på moderat intensitet (OLT-sone 3) kalt MIT for å se om dette ga effekt på submaksimale prestasjonsparametere (Almquist et al., 2022). [-@ Rønnestad et al., 2012] og 2 gjennomførte en 4-ukers syklus der uke 1 inneholdt 5 HIT-økter, mens uke 2-4 inneholdt 1 HIT-økt. (**almquist2022?**) hadde også en syklus på 4-uker over en 12-ukers periode, men med mindre volum anliggende både lav, moderat og høy intensitet. Uke 1 inneholdt 4 moderat-intensitet økter (MIT) av *412 minutter*, *3 lav-intensitets økter (LIT)* i uke 2 og *4 HIT økter* av 55 minutter i uke 3. Uke 4 var designert til restitusjon. Rønnestad et al., 2015 hadde en intervensjon bestående av 5 HIT-økter første uken, 3 i uke 3, mens resterende uker hadde 1 HIT-økt for BP-gruppen. Øktene ble gjennomført på 88-100% av HR-maks (Rønnestad et al., 2015) ved 5x6min og 6x5min økter. (**støren2012?**) skiller seg et fra de andre studiene ikke bare gjennom å være en case-studie, men også gjennom intervensjonen som varte 1 år. Forsøkspersonen gjennomgikk 2 blokkperioder på 9 og 10 dager med henholdsvis 14 og 15 HIT-økter. Øktene besto av 4x4minutter på 90-95% av HR-

maks. Studien hadde et within-subject design ettersom forsøkspersonen både gjennomførte BP- og TP-trening. Fordelen med dette er at man kan se effektene på samme forsøksperson, uten at individuelle forskjeller i treningsrespons påvirker. Dette med forutsetning at eventuell treningseffekt fra en periode påvirker den neste (Hulley et al., 2014). For framtidige studier synes randomiserte kontrollerte studier å gi gode data på effekten av BP vs. TP for effekt på Vo2-maks. Støren et al. Legger heller ikke skjul på at RCT var å foretrekke, men man kan likevel trekke fram noen styrker ved å bruke en case-studie. Blant annet var BP-periodene i studie 5 vesentlig mer intensiv, og forsøkspersonen viste periodevis tegn til overtreningssyndrom (Støren et al., 2012). Balansegangen mellom hva utøver kan akkumulere av trening før eventuelle symptomer på overtreningssyndrom oppstår er en hårfin balansegang, og her spiller også «high-risk high-reward» prinsippet inn. Jo tøffere intervensjon en utøver klarer å akkumulere, jo større effekt kan potensielt oppnås. Men dette er vanskeligere å få til med et utvalg på flere utøvere, da individuelle forskjeller og også etiske dilemmaer spiller inn. Man kan også i framtidige studier se på muligheter for bruk av crossover design, som også har blitt gjennomført tidligere av studier som ikke er tatt med i denne oppgaven (Garcia-Pallarés et al., 2010; McGawley et al., 2017; Manchado et al., 2017). Fordelene med dette kan være å se respons av BP og TP i samme individ, slik som i Støren et al., men her får man sett på flere individer samtidig. En utfordring med dette kan dog være om forsøkspersonene tar med seg eventuelle effekter fra f. eks BP inn i TP, slik at man ikke får sett effektene forårsaket kun av intervensjonen man nettopp har vært gjennom (Hulley et al., 2013). Et annet moment som kan være interessant å se på i framtidige studier er en kontrollgruppe med mer utforskende design. I studiene som er med i denne oppgaven har samtlige hatt tilnærmet samme volum av trening på lav intensitet, utenom Støren et al. Selv om dette naturligvis forsvares med at det er effekten av selve bolkleggingen man vil se på, bør man likevel spørre seg om det gir svar på om bolklegging er det beste designet på treningsplaner for utholdenhetsutøvere. Tidligere studier har blitt kritisert for å ha et for ensidig treningsprogram i kontrollgruppa (Kiely, 2012), noe Rønnestad et al. tok til betraktning i sin studie i 2015 (studie 3). Her hadde kontrollgruppa 3 HIT-økter den ene uka under intervensjonen for å variere fra de 2 ukentlige HIT-øktene de resterende ukene. Jeg mener dette er et steg i en interessant retning, og her vil jeg også strekke det enda lenger i framtidige studier. Ettersom BP i studiene har hatt et stimulus fra blokkperiodiseringen, burde de også strengt tatt få større adaptasjoner enn TP ettersom volumet på lav intensitet er det samme. Det vil alltid være et dilemma i utholdenhetstrening hva man skal vektlegge i treningsarbeidet, og en hardbolk vil gjøre at treningsvolumet utenom må justeres deretter. Derfor vil det i praksis si at TP kan ha et større treningsvolum, som ville gitt et større stimuli.

## Statistiske analyser

Valget av statistiske analyser ser man gjenspeiles av studiedesignet i de respektive studiene. Ettersom studie 1-4 var RCT-studier med 2 grupper, brukes verktøy for å se korrelasjon mellom gruppene for å lete etter statistisk signifikans. Studiene gjennomført av Rønnestad et al. (1,

2 og 3) regnet ut effektstørrelsen (ES) mellom gruppene som Cohen's d for å finne statistisk signifikans. Kriteriene satt som grenser for å tolke av ES var følgende; 0.0–0.2 triviell, 0.2–0.6 liten, 0.6–1.2 moderat, 1.2–2.0 stor, og >2.0 veldig stor (Hopkins et al., 2009). Almquist et al. brukte derimot en mikset linjer modell med grupper definert som «fixed effects» og rettet ved bruk av Pre-verdier som et kovariat med SPSS v. 25 software. ANOVA ble brukt i studiene til Rønnestad et al., men i noe ulik grad. I studie 1 og 3 ble en to-veis repeterende analyse av varians for å evaluere differanser innad og mellom gruppene fra Pre til Post. I studie 1 og 2 ble uparrede students t-test brukt for å sammenligne gruppene ved baseline og gruppenes treningsvolum. Valg av statistiske tester påvirker resultatene i stor grad, og valget av tester bør overveies godt av forskerne. Med t-test kan man sammenligne om to grupper har forskjeller med statistisk signifikans seg imellom, mens med ANOVA kan man sammenligne 3 eller flere grupper for å se etter signifikans. Testene blir gjort opp mot hypotese (null-hypotese). Den åpenbare fordelten med t-test er mindre grad av risiko for error ettersom det kun er 2 grupper, mens med ANOVA blir risikoen for dette større. Samtidig er ANOVA bedre egnet for flere grupper, da den reduserer risikoen for error som vil forekomme ved en t-test på flere enn 2 grupper. Samtlige studier utenom case-studien til Støren et al. hadde en alpha-level på  $P < 0,05$ .

# Deloppgave 5: Analysere repeterte målinger

## Introduksjon

Adaptasjoner i menneskekroppen til styrketrening henger sammen med en rekke faktorer, både biologiske, genetiske og ytre faktorer. Av disse finner vi treningsvolum, intensitet, valg av øvelser, hvile mellom sett, hurtigheten på utførelsen på øvelsene og hyppigheten på treningen (Ratamess et al., 2009). Genetiske disposisjoner synes å ha en betydelig innvirkning på responsen av trening (Yang et al., 2003). Styrketrening har en rekke prestasjonsmessige fordeler i idretter som stiller krav til styrke, men også utholdenhetsidrett (Rønnestad et al., 2014; Beattie et al., 2014), men også i et folkehelseperspektiv (Westcott, 2012). Styrketrening er ikke bare nyttig for prestasjon i idrett, men også for folkehelsen. Aldersbetinget muskeltap har en rekke godt dokumenterte negative helsekonsekvenser (Hurley, 2000), og styrketrening synes å ha god effekt for å motvirke dette (Westcott, 2012). Men selv om gevinsten av styrketrening virker å være stor, kan det for mange være vanskelig for mange å få dette inn i en hektisk hverdag. Derfor er det både interessant og viktig at forskere ser på sammenhengen mellom treningsvolum, muskelmasse og styrke. Dette for å eksempelvis finne ut hva som minimum må til for å få effekt.

En del studier har funnet at et minimalt treningsvolum gir adaptasjoner på muskelmasse og styrke som moderat-volum (Ostrowski et al. 1997; Cannon & Marino, 2010; Mitchell et al. 2012), mens flere metaanalyser favoriserer et moderat volum (Rhea et al. 2003; Krieger, 2009, 2010; Schoenfeld et al. 2016). Rønnestad et al. (2007) gjennomførte en lignende studie som denne, og sammenlignet 1 vs. 3 sett for bein og overkropp. Funnene her var sprikende, 3 sett ga signifikant mer effekt på 1RM og muskeltverrsnitt i underkropp, mens det var ingen endring og ingen signifikant endring for henholdsvis 1RM og muskeltverrsnitt målt i overkropp. Med treningsvolum menes mengde styrketrening, altså hvor mye og hvor ofte. Muskelvolum/masse måles i fettfri masse (lean body mass), og måler muskelvolum i tillegg til annen fettfri masse (bein, sener, leddbånd, indre organer etc.). I tillegg har man muskelstyrke som er hvor stor kraft en muskel eller muskelgruppe kan produsere.

Ettersom forskningsresultatene er sprikende angående kravet til volum av styrketrening for å få en økning i muskelmasse og styrke, vil denne studien undersøke dette videre. Ved å sammenligne 1 vs. 3 sett under styrketrening for underkropp over en 12-ukers periode vil vi se om et minimalt volum kan måle seg med et moderat volum. Ettersom majoriteten av tidligere forskning tyder på at et moderat volum er overlegent et minimalt, tar vi utgangspunkt i

en hypotese om at 3 sett vil gi et signifikant forbedring av muskelstyrke og muskelmasse sammenlignet med 1 sett.

## **Metode**

### **Deltakere og studieoversikt**

Det ble rekruttert 41 mannlige og kvinnelige deltakere til studien med kriterium om å være ikke-røyker og å være mellom 18 og 40 år gammel. Eksklusjonskriterium var hemmet muskelstyrke som følge av tidligere skade, intoleranse for lokal anestesi, ha gjennomført mer enn en ukentlig styrketreningsøkt siste året og bruk av medisiner som kan påvirke adaptasjoner av treningen. På bakgrunn av å ha gjennomført mindre enn 85% av treningsprogrammet ble 7 deltakere ekskludert fra studien. Årsakene til dette var ubehag eller smerter i underekstremiteter under trening (n=5), fulgte ikke treningsprotokollen (n=1) og skade urelatert til studien (n=1). Ved studiestart var det ingen forskjeller i maksimal frivillig kontraksjon (MFK) normalisert til kroppsmasse eller antropometri mellom inkluderte og ekskluderte deltakere. Ved prøvetaking i uke 2 av biopsi og blod avsto en av de inkluderte deltakerne dette, og i tillegg ble det ikke tatt blodprøve av 3 andre deltakere ved ulike tidspunkt grunnet feil under prøvetaking. Alle inkluderte deltakere hadde tidligere erfaring med idrettsaktiviteter som eksempelvis lagidretter, langrenn og gymnastikk. Antall deltakere som rapporterte om fysisk trening utenom studien under intervensjonen (median = 2 ukentlige økter fra 0,5 til 4) var 20.

### **Intervensjonen**

Den 12 uker lange intervensjonen ble gjennomført mellom september og november. Intervensjonen besto av fullkroppss styrketrening, og benøvelsene ble gjennomført separat for hvert enkelt ben. Dette for å kunne muliggjøre differensiering av treningsvolum på samme deltaker. Hvert enkelt ben til hver deltaker ble tilfeldig tildelt enten single-set eller 3-set programmet, ergo hver deltaker gjennomførte begge protokoller. Testing av muskelstyrke ved målt før og etter intervensjonen, i tillegg til uke 3, 5 og 9. Muskelbiopsi ble tatt fra begge ben fra vastus lateralis i uke 0 i hvilt tilstand, før og 1 time etter femte treningsøkt og etter intervensjonen i uke 12 i hvilt tilstand. Kroppssammensetning ble målt før etter intervensjonen. Deltakerne loggførte inntak av kosttilskudd fra uke 6, inntaket ble veid og loggført i 4 til 5 sammenhengende dager inkludert en helgedag.

### **Testing av muskelstyrke**

Ved bruk av et dynamometer (Cybex 6000, Cybex International, Medway, MA, USA) ble det gjennomført styrketest av isometrisk og isokinetisk unilateral kneesktensjon. Det ble foretatt målinger av muskelstyrke, muskeltverrsnitt, kroppssammensetning, hormoner i tillegg til

muskelbiopsi i forbindelse med studien. Med kneet justert med rotasjonsaksen til dynamometeret gjennomførte deltakerne test av maksimalt isokinetisk dreiemoment ved 60°, 120° og 240° s<sup>-1</sup> hastighet. Deltakerne ble gjennom submaksimale gjennomføringer gjort kjent med alle testens hastigheter på forhånd, og ble gitt 2 (60° s<sup>-1</sup>) og 3 forsøk (120 og 240° s<sup>-1</sup>) under selve testingen der den høyeste verdien for hver hastighet ble brukt i statistiske analysene. MFK-dreiemoment ble målt ved knevinkel på 30° (full ekstensjon på 90°) etter isokinetisk testing. Instruksjon lød å presse med maksimal kraft i 5 sekunder. Med 30 sekunders hvile mellom ble deltakerne gitt 2 forsøk, der det beste ble brukt for videre analyser.

Testing av 1RM i unilateral beinpress og kneekstensjon startet med spesifikk oppvarming i hver øvelse med 10, 6 og 3 repetisjoner på henholdsvis 50, 75 og 85% av antatt 1RM. 4 til 6 forsøk ble gitt for å finne 1RM, og høyeste vekt som ble løftet med full «range of motion» ble satt som 1RM etter å ha økt motstanden progressivt.

Ved baseline ble både 1RM, isokinetisk og isometrisk styrketester gjennomført i 2 omganger med minst 4 dager mellom hver omgang, og de beste verdiene ble brukt i de statistiske analysene. Fra testene var det igjen minst 48 timer før og etter intervensjonen. Alle 3 testene utgjorde sammen en kombinert score for muskelstyrke.

## Måling av muskelmasse og kroppssammensetning

Muskeltverrsnitt av knestrekere (vastus lateralis, medialis, intermedius og rectus femoris) ble målt før og etter intervensjonen ved bruk av MRI i samsvar med produsentens protokoll (S-Scan, Esaote Europe B.V., Maastricht, the Netherlands). Bildene ble analysert anonymt av samme forsker ved bruk av OsiriX (v.5.6, Pixmeo Sarl, Bernex, Switzerland). Tverrsnittet ble målt på nøyaktig samme sted for hver enkelt deltaker på hver test ved bruk av minst 4 påfølgende bilder. Kroppssammensetning ble fastslått før og etter intervensjon ved bruk av dual-energy X-ray absorptiometry (DXA) (Lunar Prodigy, GE Healthcare, Oslo, Norway). Det skilte 2 dager fra siste styrketest til måling av kroppssammensetning og deltakerne ble bedt om å faste og unngå krevende fysisk aktivitet siste 2 timer før måling av MRI og DXA.

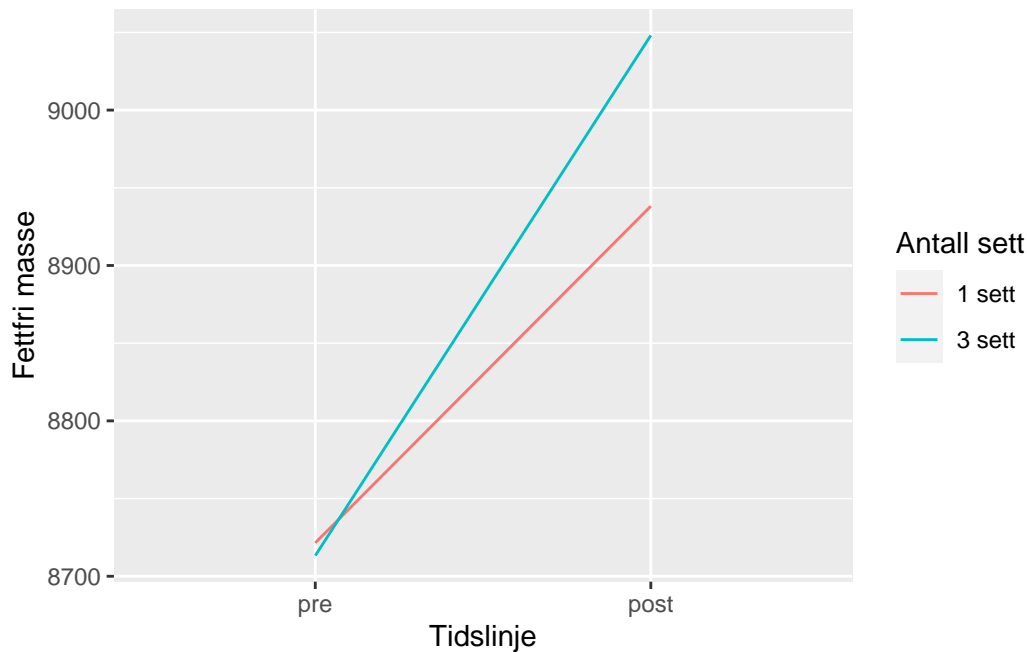
## Statistikk

RStudio (versjon 2022.07.1) for Windows ble brukt for statistiske analyser. Et mål på forandring (change score) fra pre- til post-test ble regnet ut for å måle treningseffekten. En gjennomsnittlig baseline ble regnet ut fra pretest og første treningsøkt under intervensjonen (for øvelsen benpress), og ut fra dette ble endringen til post-scoren regnet ut ved bruk av en parret t-test. Endring i fettfri masse tok utgangspunkt i pre og post-testing, og ble også regnet ut ved bruk av en parret t-test. Kravet for statistisk signifikans (alpha-level) var  $P < 0,05$ .

## Resultater

### Kroppssammensetning

Sammenlikning med bruk av parret t-test av endring i fettfri masse fra pre til post mellom gruppene med 1 sett og 3 sett. Økningen i 3 sett-gruppen var signifikant større enn gruppen med 1 sett ( $t(33) = 2.1875$ .  $P = 0.0359$ ).  $P < 0.05$ .



Figur 1: Endring i fettfri masse fra pre til post for gruppene med 1- og 3-sett.

### Muskelstyrke (benpress)

Sammenlikning med bruk av parret t-test av endring i maksimal muskelstyrke (benpress) fra pre til post mellom gruppene med 1 sett og 3 sett. Økningen i 3-sett gruppen var signifikant større enn gruppen med 1 sett ( $t(29) = 2.1366$ .  $P = 0.0412$ .  $P < 0.05$ ).

## Diskusjon

Ut fra resultatene ser vi at økningen i både fettfri masse (muskelmasse) og muskelstyrke i benpress økte signifikant mer for ben som gjennomførte moderat volum sammenlignet med ben som gjennomførte minimalt volum. Dette bekrefter altså hypotesen om at et moderat

volum er overlegent et minimalt volum, ved at 3 sett ga signifikant økning i muskelmasse og muskelstyrke i forhold til 1 sett.

Resultatene i denne studien bygger opp under funnene til (Rønnestad et al., 2007; Rhea et al. 2003; Krieger, 2009, 2010; Schoenfeld et al. 2016)



# Kilder

- Almquist, N. W., Eriksen, H. B., Wilhelmsen, M., Hamarsland, H., Ing, S., Ellefsen, S., Sandbakk, Ø., Rønnestad, B. R., & Skovereng, K. (2022). No Differences Between 12 Weeks of Block- vs. Traditional-Periodized Training in Performance Adaptations in Trained Cyclists. *Frontiers in Physiology*, 13. <https://www.frontiersin.org/articles/10.3389/fphys.2022.837634>
- Dellsén, F. (2022). *Vitenskapsfilosofi 2022.docx* 3. *KvantMet2022 Vitenskapelige teorier.pdf* 4. *KvantMet2022 Vitenskapelige argument.pdf* 5. *KvantMet2022 Abduksjon og sannsynlighet.pdf* 6. *KvantMet2022 Sannsynlighet og replikasjonskrisen.pdf*.
- Hulley, S. B. (Ed.). (2013). *Designing clinical research* (4th ed). Wolters Kluwer/Lippincott Williams & Wilkins.
- Rønnestad, B. R., Ellefsen, S., Nygaard, H., Zacharoff, E. E., Vikmoen, O., Hansen, J., & Hallén, J. (2014). Effects of 12 weeks of block periodization on performance and performance indices in well-trained cyclists. *Scandinavian Journal of Medicine & Science in Sports*, 24(2), 327–335. <https://doi.org/10.1111/sms.12016>
- Rønnestad, B. R., Hansen, J., & Ellefsen, S. (2014). Block periodization of high-intensity aerobic intervals provides superior training effects in trained cyclists. *Scandinavian Journal of Medicine & Science in Sports*, 24(1), 34–42. <https://doi.org/10.1111/j.1600-0838.2012.01485.x>
- Støren, Ø., Bratland-Sanda, S., Haave, M., & Helgerud, J. (2012). Improved V[Combining Dot Above]O<sub>2</sub>max and Time Trial Performance With More High Aerobic Intensity Interval Training and Reduced Training Volume: A Case Study on an Elite National Cyclist. *The Journal of Strength & Conditioning Research*, 26(10), 2705–2711. <https://doi.org/10.1519/JSC.0b013e318241deec>
- Thiese, M. S. (2014). Observational and interventional study design types; an overview. *Biochemia Medica*, 24(2), 199–210. <https://doi.org/10.11613/BM.2014.022>