

# Mappeeksamen Kvantitativ metode og statestikk

Kristoffer Solum, Kandidatnummer:108

2021-12-01



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Arbeidskrav1</b>                            | <b>5</b>  |
| 1.1      | Introduksjon . . . . .                         | 5         |
| 1.2      | Metode . . . . .                               | 5         |
| 1.3      | Statistikk . . . . .                           | 6         |
| 1.4      | Resultat og diskusjon . . . . .                | 6         |
| <b>2</b> | <b>Vitenskapsfilosofi</b>                      | <b>9</b>  |
| 2.1      | Falsifikasjonisme . . . . .                    | 9         |
| 2.2      | HD-metoden og abduksjon/Bayesisme . . . . .    | 10        |
| 2.3      | Replikasjonskrisen . . . . .                   | 11        |
| <b>3</b> | <b>Study-Design</b>                            | <b>13</b> |
| 3.1      | Inntroduksjon . . . . .                        | 13        |
| 3.2      | Metode . . . . .                               | 13        |
| <b>4</b> | <b>Analyzing repeated measures experiments</b> | <b>17</b> |
| 4.1      | Inntroduksjon . . . . .                        | 17        |
| 4.2      | Metode . . . . .                               | 17        |
| 4.3      | Resultater . . . . .                           | 19        |
| 4.4      | Diskusjon . . . . .                            | 19        |



# Chapter 1

## Arbeidskrav1

### 1.1 Introduksjon

En av de mest sentrale faktorene for utholdenhets prestasjonen er det maksimale oksygenopptaket ( $VO_{2\text{maks}}$ ) Bassett [2000].  $VO_{2\text{maks}}$  er bestemt av flere sentrale faktorer: lungenes kapasitet til å føre oksygen fra blodet, hjertets pumpekapasitet, volum, og sammensetning av blodet, og musklenes evne til å bruke oksygenet Bassett [2000].

Reliabilitet kan bli definert som reproduserbarheten til testverdier, analyser eller andre målinger på gjentatte forsøk på de samme individene. Det finnes tre hoved variasjoner man må ta hensyn til når man snakker om reliabilitet, det er: deltakernes variasjon, endring i gjennomsnittet og rettest korrelasjon Hopkins [2000]. Deltakernes variasjon regnes ut gjennom standardavvik og standardfeil. Det kan komme fra biologiske forhold som kan variere mellom to tester. Utstyr kan og spille inn på dette, hvor det er støy ved målingene. Endring i gjennomsnitt kan man dele i to deler, det er tilfeldig endring og systematisk endring. Tilfeldig endring er støy i målinger og data innsamlingsfeil. Dette kan minimeres gjennom å ha mange tester, hvor da tilfeldige feilene vil spille mindre inn på resultatet. Systematisk endring er treningseffekten og læringseffekten man kan forvente mellom to tester, det handler om faktorene som spiller inn og kan gjøre test2, bedre enn test1. Korrelasjon i test2 handler om hvor godt test1 og test2 korrelerer, hvis man har bedre korrelasjon, har man og høyere reliabilitet mellom testene.

### 1.2 Metode

I denne rapporten skal reliabiliteten estimeres mellom to  $VO_{2\text{maks}}$  tester. Hvor test1 og test2 er gjennomført med en ukes mellomrom.

VO2maks testen ble gjennomført med en standard VO2maks protokoll. Hvor stigningen var konstant, 10,5% for guttene og 5,5% for jentene. Startfart var gitt på forhånd hvor alle startet på 8 km/t. For hvert minutt som gikk, økte farten med 1 km/t, og slik fortsatte det til utmattelse. Underveis i testen ble det gitt verbal oppmuntring fra testleder, det var og testleder som justerte farten underveis. Testen ble gjennomført på en woodway løpemølle (4FRONT, wisconsin). Hele testen ble gjennomført med kontinuerlig oksygenmåling hvert 30. sekund. Oksygenmålingene ble gjennomført med en Vyntus CPX, mixing chamber (Vyntus CPX, Jaeger-CareFusion, UK). Ved målinger på Vyntus CPX ble det automatisk kalibrert for gass, volum og luftfuktighet før hver test. Oppvarmingen før testen var valgfri, og varte i 15 min. Underveis i oppvarmingen ble testprotokollen forklart for utøver, og eventuelle spørsmål om testen ble avklart. Under testen ble det registrert ml/min VO2. Umiddelbart etter testen ble det spurt om Borg skala 6-20 (RPE: Borg [1970] ), det ble og notert ned hva siste belastning var og hvor lenge utøver løp på denne belastningen. Et minutt etter testen ble det målt laktat (La) gjennom et fingerstikk og analysert gjennom Biosen blodlaktatmåler (Biosen C-line, EKF Diagnostics, Barleben, Germany), HF ble og notert ned.

Test2 ble gjennomført på samme måte som test1, hvor samme protokoll ble gjennomført på nytt.

### 1.3 Statistikk

Alle statistiske beregninger ble gjort i RStudio (versjon RStudio 1.4.1717; R Foundation for Statistics Computing, Vienna, AT). Det ble regnet ut gjennomsnitt og standardavvik mellom test1 og test2. Det ble også regnet ut standardfeil for testen som et mål på reliabilitet.

### 1.4 Resultat og diskusjon

Resultatene for studien er vist i figur 1.

Standardfeilen i studien var 4.04 %.

Standardaviket i studien var 229.82

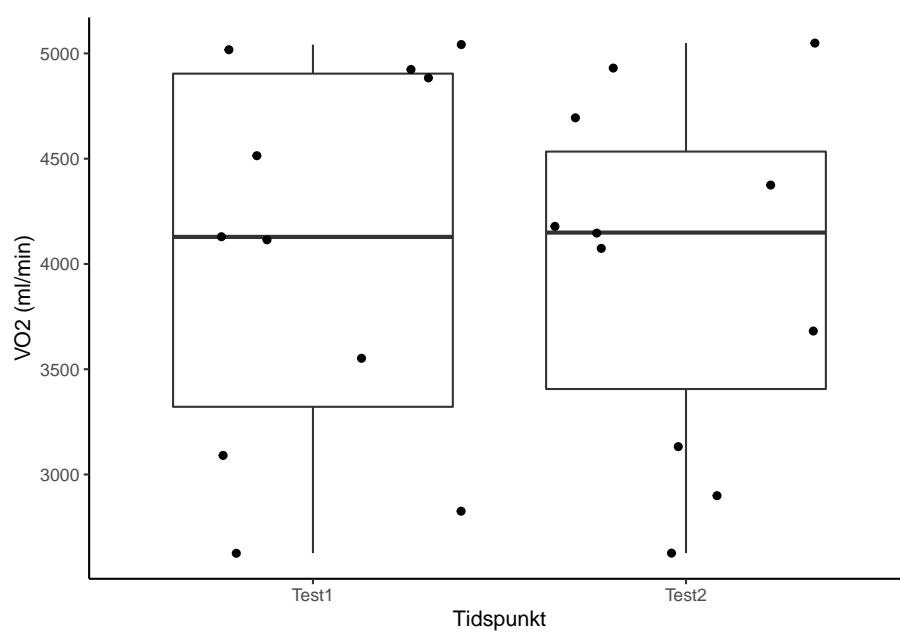


Figure 1.1: Figuren viser resultater maksimalt oksygenopptak (VO2maks) test fra test 1 og 2





## Chapter 2

# Vitenskapsfilosofi

### 2.1 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 handler enkelt forklart om hva han mente var en god vitenskapelig teori og hva som var en dårlig. I følge Popper ville det å si at noe var en god vitenskapelig teori bety at den var falsifiserbar Okasha [2016]. At en teori er falsifiserbar, betyr at det skal være mulig å påvise at den er gal eller uholdbar, og er det motsatte av å kunne verifisere. I følge popper var det en skarp linje mellom vitenskapelige og uvitenskapelige teorier, altså falsifiserbare og ufalsifiserbare teorier Okasha [2016]. Og at vitenskapelige teorier ikke kunne bli bekreftet, hverken godt eller dårlig de kan kun bli falsifisert og aldri bekreftet Okasha [2016]. Med dette mente popper at man kunne unngå induksjonsproblemet, som er hvordan man skal begrunne gyldigheten til teorier Okasha [2016]. For hvis man isteden bare prøvde å falsifisere dem var det ingen grunn til å bruke induksjon. Bakgrunnen for falsifiserbarhetskriteriet til popper var også at det skulle gi et svar på demarkasjonsproblemet, som er spørsmålet om hva som egentlig skiller vitenskap fra ikkevitenskap.

Den logiske strukturen til poppers falsifisering er at fra en teori kan vi utarbeide hypoteser, hvis disse hypotesene viser seg å være usanne via observasjoner og eksperimenter, vil dette bety at teorien er usann Okasha [2016]. En falsifiserbar teori vil da stride med noen mulige empiriske observasjoner, den vil også gi gode forutsigelser, altså forutsigelser som kan bli testet mot data og som ville falsifisere teorier hvis ikke dataen stemmer med forutsigelsen Okasha [2016]. En ufalsifiserbar teori vil derimot være forenlig med alle mulige empiriske observasjoner, og den vil ikke gjøre noen forutsigelser som kan testes mot dataen og

dermed kunne falsifisere teorien dersom dataen ikke passet med forutsigelsen Okasha [2016].

Andre vitenskapsteorier mener at det ikke er en skarp linje mellom vitenskapelige og uvitenskapelige teorier. Derimot er den viktigste forskjellen mellom godt og dårlig bekreftede teorier. Spesielt Okasha [2016] mente at svakheten i Poppers argument var klar. For målet med vitenskapen er ikke bare å motbevise teorier, men også anta hvilke teorier som er sanne eller mest sannsynlig sanne. Han foreslo at når vitenskapsmenn samler data vil kanskje målet deres være å vise at en spesiell teori, muligens fra deres erkerival, er feil. Men mer sannsynlig prøver de å bevise at deres egen teori er sann. Og for å kunne gjøre dette må de muligens ty til en form for induktivt argument. Så Poppers forsøk på å vise at vitenskap kan klare seg uten induksjon lykkes ikke. Slik jeg ser det veldig vanskelig å avgjøre hvem jeg syns har rett, jeg syns at popper har et veldig godt poeng med å ha et falsifiseringskriterium og et skarpt skille mellom hva som er vitenskap og ikke. Ved dette vil det bli veldig klart hva som er vitenskap og det kan være en stor fordel. Samtidig er det ikke helt riktig å dra all vitenskap som ikke kan falsifiseres under en kam og kalle det for pseudo- vitenskap, eller ikke vitenskap. Så jeg syns det er vanskelig å gjøre meg opp en klar mening om hvem som har rett når jeg syns begge to har gode meninger.

## 2.2 HD-metoden og abduksjon/Bayesisme

Hva er strukturen på et bekreftende vitenskapelig argument ifølge den hypotetisk deduktive metode? Forklar ut fra Hempels artikkel men bruk egne eksempler. Sammenlign også Hempels HD-metode med hvordan vitenskapelig bekreftelse fungerer ifølge enten abduksjon eller Bayesianisme.

Hypotetisk deduktiv metode, eller HD- metoden har en struktur som bekrefter vitenskapelige argumenter gjennom å få positive resultater når man tester de deduktive konsekvensene av teorien Hempel [1966]. Strukturen til HD- metoden er beskrevet av Hempel Hempel [1966], det første trinnet i HD- metodens strukturen er en «utdannet gjetning», her blir det formulert en hypotese eller teori. I det andre trinnet i strukturen blir det utledet empiriske konsekvenser fra teorien eller hypotesen. Testingen eller observasjoner av disse empiristiske konsekvensene skjer i trinn tre. Og i trinn fire ser man på om de empiriske konsekvensene viser seg å være riktige, og da blir teorien bekreftet til en viss grad, på induktivt vis. Hvis jeg skal forklare HD- metoden ved et eget eksempel kan jeg stille et spørsmål som «hvordan kan vi vite at kyssesyken er et virus?» I trinn to, vil i så fall da kyssesyken (a) være smittsom, men (b) ikke bli kurert av antibiotika. I trinn tre, vil observasjoner og eksperimenter vise at kyssesyken både er (a) og (b). Derfor vil vi i trinn fire kunne få en induktiv bekreftelse på at kyssesyken er forårsaket av et virus.

En annen form for vitenskapelig bekreftelse som er interessant å sammenligne med den hypotetiske deduktive metoden er bayesianisme. Bayesianisme er en

metode som brukes for å øke tillitten til en teori eller hypotese ettersom mer informasjon og bevis blir tilgjengelige Okasha [2016]. For å få til dette brukes det sannsynlighet, ifølge bayesianisme må en rasjonell person meninger om hvor sannsynlig forskjellige påstander er. Dette betyr at om en forsker er 50% sikker på at en teori er riktig må man også være 50% sikker på at teorien er usann. Sannsynligheten kommer ut i fra at forskere bruker den bayesiske kondisjonaliseringsregelen, som sier at når en har hentet inn data(D), burde sannsynlighetsverdiene for en teori(T) oppdateres gjennom å kondisjonalisere D. Og T burde da aksepteres som bekreftet når sannsynligheten er tilstrekkelig høy Okasha [2016]. Et eksempel på bayesianisme i praksis blir beskrevet av Okasha [2016] som følger, hvis vi ser for oss at en T har et testbart utsagn (U), og forskeren i utgangspunktet har tillit til at både T og U er sant. Vi antar at både T og U tar «non- extreme values», som vil si at de verken er en eller null. Hvis vi deretter ser for at forskeren finner ut at U er definitivt sant, vil det være logisk at forskeren får en enda større tro på at T også er sann. Med andre ord en forsker som oppdager at teorien deres har gitt en sann spådom vil nødvendigvis øke deres tillitt til teorien så lenge de adlyder kondisjonaliseringsregelen. Så det faktum at suksessfulle utsagn ofte fører til at forskere får mer tiltro til deres egen teori, har en ryddig forklaring på det bayesianske synet på vitenskapelig bekreftelse Okasha [2016]. Hovedforskjellen mellom hypotetisk deduktiv metode og bayesisme er da at HD- metoden tar en utdannet gjetting i forhold til observasjonene den får inn, mens bayesisme bare blir styrket mer og mer etter som den får inn flere resultater og tilslutt blir bekreftet når sannsynligheten er høy nok. Begge metodene tar også en bekreftelse ut fra resultatene de kommer frem til, men HD- metoden har på forhånd en hypotese den tester ut ifra, som eksemplet mitt viste hvor hypotesen var at kysseesyken var et virus. I HD- metoden blir da eksperimentene og forskningen testet for å bekrefte eller motbevise dette, mens i bayesismen vil da teorien til forskerne kunne endres mye underveis ettersom hvilke resultater de kommer frem til.

## 2.3 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?

Begrepet replikasjonskrisen er brukt til å beskrive både at det er enormt mange felt av studier det ikke er mulig å replisere, og at disse feltene står ovenfor en krise som en konsekvens av dette Bird [2020]. Ut av 1576 forskere i en undersøkelse mener hele 52% at det er en signifikant krise når det kommer til replikasjon Baker [2016]. En spesiell bekymring av replikasjonskrisen er at det vil muligens øke mistillit til vitenskapen som tilhører helt andre felt enn der selve replikasjonskrisen oppstår Bird [2020].

Grunnen til at mange resultater innenfor forskingen ikke kan repliseres er på

grunn av basefrekvensfeilen, her er det to typer feil, akseptere en usann teori og avise en sann teori. Bird Bird [2020] mener Basefrekvensfeilen vil gjerne oppstå når man trekker en slutning om sannsynligheten for en bestemt forekomst av et generelt fenomen, for eksempel om en person har en sykdom. Feilslutningen oppstår når forskeren fokuserer utelukkende på et eller annet fremtredende bevis mens han overser hvor mange forekomster av dette fenomenet vil skje uansett, uavhengig av beviset. Dette fører til feilaktige konklusjoner når bevisene er sterkt, men ikke helt korrekt, for eksempel når bevisene er en slags test for et fenomen, og selve fenomenet er sjeldent. For å demonstrere dette bruker Bird et eksempel hvor passasjerer på et fly bli undersøkt for å være terrorister, hvor testen er 95% riktig. I dette tilfelle vil testen si at av 100 personer som ikke er terrorister vil 95 være ikke terrorist og 5 vil være terrorister. I så fall om en person ikke består testen vil den si at han er en terrorist. Dette mener Bird var tullete etter som det absolutt ikke var en så høy sannsynlighet for at passasjerer på fly er terrorister. Hvis denne testen blir trodd vil det i så fall bli akseptert en usann teori. Hvis mange av hypotesene vi tester er usanne, da er sannsynligheten til hypotesene som bekreftes i mange vitenskaper ganske lav Bird [2020]. Andre forklaringer for replikasjonskrisen som Bird diskuterer er at enkelte har assosiert replikasjonskrisen med den lave statistiske kraften fra mange av studiene. Populasjonen altså antall deltakere, styrer i stor grad styrken til forskningen, og gir en større sannsynlighet for å ikke få en type II feil, altså et falsk negativt resultat Bird [2020]. Ifølge Bird er dette likevel ikke en god forklaring fordi man fremdeles får en høy sannsynlighet for type II feil selv med høy statistisk styrke Bird [2020]. For å understreke dette viser Bird frem en utregning som tilsier at det fremdeles er 31% sjanse for type II feil selv med høyere statistikk enn i «vanlig» vitenskap Bird [2020]. Bird mener det dermed vil være en stor fordel for vitenskapen å få opp den statistiske styrken men det vil likevel ikke gi en løsning på krisen.

Basert det man har sett av studier er det vanskelig å gi en fast konklusjon men, dårlig praksis og juks kan være en forklaring på replikasjonskrisen Bird [2020]. Derfor vil det ikke være gode nok argumenter til å mene at en dårlig moral innenfor vitenskapen kan forklare hele krisen Bird [2020]. Slik jeg tolker det er det ganske tydelig at Bird mener hans egne argumenter og forklaring gir mest styrke til teorien om replikasjonskrisen. Og en større forståelse av statistikk og sannsynlighet vil være viktig for alle som driver med vitenskap Bird [2020].

## Chapter 3

# Study-Design

### 3.1 Introduksjon

I denne oppgaven har jeg funnet 5 originalartikler [Billat et al., 2000, 2001, Dupont et al., 2002, Thevenet et al., 2007, Wakefield and Glaister, 2009] som jeg skal sammenligne og se på hvordan design disse studiene har og hvilke statistiske analyser de bruker. Gjennom denne oppgaven ønsker jeg å fremheve hva studiene gjør bra og hva de eventuelt kan gjøre bedre.

### 3.2 Metode

Formålet i alle studiene handler i hovedsak om å undersøke sammenhengen mellom intensitet ved kortintervaller og hvor mye tid det gir over en gitt prosent av det maksimalt oksygenopptak (vo2 maks) deltakerne oppnår ved forskjellige protokoller. Studien til Thevenet et al. [2007] og Wakefield and Glaister [2009] sammenligner begge effekten av treningsintensitet på tid over 95% av VO2 maks, Thevenet et al. [2007] har en hypotese som tilsier at en økning i hastigheten fra 100% av farten ved vo2maks til 115% av farten ved vo2maks under intervallene ville også gi en økning i tiden over 90- og 95% av vo2maks. Mens Wakefield and Glaister [2009] har en motsatt hypotese om at tid på eller over 95% av vo2maks ville være lenger ved løp på 105% av farten ved vo2 maks enn ved 115% av farten ved vo2 maks. Studien til Billat et al. [2000] og Billat et al. [2001] til forskjell intervaller hvor de ser på hvordan man kan fremkalle mest tid ved 100% av vo2maks, deriblant hvilken intensitet og intervallform som er best for dette. Mens Dupont et al. [2002] passer bra inn med alle de andre artiklene over og ser på både tid ved vo2maks og tiden over 90% av vo2maks ved kortintervaller 15/15.

Alle studiene presenterte en hypotese. Hypotesen de kommer med er begrunnet av tidligere forskingsdata. Som et eksempel begrunner Wakefield and Glaister [2009] sin hypotese ved å referere til data i studien til Thevenet et al. [2007]. Gjennom å presentere denne dataen fra tidligere forskning skapes det en logisk linje gjennom introduksjonen til hypotesen. Dette blir gjort på en spesielt god måte i artikkelen til Thevenet et al. [2007] hvor artikkelen starter bredt med å fortelle om vo2maks, og at det er ansett som en viktig faktor innenfor langdistanseløping. Dette blir bygget opp ved å referere til tidligere artikler som har kommet frem til dette. Videre blir det beskrevet treningsmetoder for å forbedre vo2maks, hvor mye tid på en høy % av vo2maks blir beskrevet som en viktig faktor spesielt tid over 90- og 95% av vo2maks. Etter dette beveger introduksjonen i artikkelen seg over på å fortelle om intervalltrening og spesielt kortintervaller og effekt på å kunne arbeide på en høy % av vo2 maks. Artikkelen forteller videre om andre forskingsartikler som har sett på kortintervaller og tid over 90 og 95% men avdekker samtidig kunnskapshull i disse artiklene. Dette gjør at Thevenet et al. [2007] formulerer en hypotese basert på informasjonen om vo2maks og kortintervaller som skal forsøke å tette disse kunnskapshullene som ble beskrevet i introduksjonen.

Dermed funker denne introduksjonen som en trakt som starter med å fortelle om et tema bredt, og smaler seg etter hvert inn på mindre temaer og spesifikke problemer den ønsker å svare på, samtidig som den har en rød trå eller logisk linje gjennom introduksjonen og frem til hypotesen. De andre artiklene følger i likhet denne samme strukturen i sin innledning som leder til en hypotese.

Studiedesignet i alle disse studiene var randomisert kontrollert undersøkelse (RCT). I en RCT metode der sammenlignes flere grupper. Gruppene som sammenlignes burde være så like som mulig med tanke på alt som kan påvirke utfallet. Dette oppnås ved å tilfeldig fordele eller randomisere deltakerne. I disse testene ble deltakerne randomisert i grupper som skulle utføre forskjellige intervaller eller så ble det randomisert rekkefølgen de skulle gjennomføre intervallene i. Denne randomisering er den metoden som med størst sannsynlighet gir sammenlignbare grupper.

Siden alle studiene på forhånd hadde spesifisert en primær hypotese stemmer dette også godt med designet i en clinical trial. En god regel for clinical trial, er å etablere flere hypoteser på forhånd som gir mening, men å spesifisere en hypotese som den primære, denne må kunne bli statistisk testet uten argument om at hypotesen må endres (Hulley [2013] s53). Mer viktig vil en primær hypotese hjelpe studien å fokusere på dens hovedoppgaver og bidra til en ren base for kalkulasjoner av utvalgsstørrelsen (Hulley [2013] s53).

Alle studiene hadde en god beskrivelse av forsøkspersonene som var med i testen. De viste alle frem med standardavvik antall deltakere, kjønn, alder, høyde og vekt. Alle studiene fortalte også at alle deltakere var godt trente personer, at de hadde meldt seg som frivillige og at de var fullt informert over test prosedyren og hadde underskrevet på dette, foresatte underskrev for deltakere under 18 år. Alderen på deltakerne i de forskjellige studiene varierte mye hvor Thevenet et al.

[2007] hadde den yngste snittalderen på  $16,1 \pm 1$  mens Billat et al. [2001] hadde den eldste snittalderen på  $51 \pm 6$ . Ingen av disse to studiene forsvarte hvorfor de valgte å teste disse aldersgruppene på henholdsvis yngre og litt eldre utøvere. I testen til Thevenet et al. [2007] ble det forklart at deltakeren alle var rekruttert fra samme treningsklubb, og i studiene til både Dupont et al. [2002] og Wakefield and Glaister [2009] var deltakeren idrettsstudenter. I de to resterende artiklene var det ikke oppgitt hvor deltakerne var rekruttert fra Billat et al. [2001] og Billat et al. [2000]. Alle studiene hadde forholdsvis få deltakere med mellom 7-9 personer, og ingen av studien hadde noen forklaring på hvorfor de hadde det antallet deltakere. Dette er litt underlig ettersom en studie med flere deltakere ville gitt en større statistisk styrke, fordi det da vil være fler observasjoner som påvirker gjennomsnittet.

I alle studiene ble det brukt p-verdier som hvor signifikantnivået var  $P < 0.05$ . Ingen av studiene brukte andre tester som t-test eller effektstørrelse som kunne ha bidratt til å styrke resultatet ettersom dette er en test som kan styrke opp om p-verdien. Hvordan testene ble gjennomført var svært forskjellige Wakefield and Glaister [2009] gjennomførte testen på mølle med 1% stigning mens resten av testene ble gjennomført på løpe bane. Alle studiene hadde intervaller som en av testene, her var alle kortintervaller med mellom 15- 30sek arbeidstid. Før intervallene gjennomførte alle studiene vo2maks test, som ble bestemmende for farten de skulle løpe på under intervallene. Dupont et al. [2002] hadde den raskeste hastigheten med 15/15s intervaller på 140% av farten ved vo2maks, mens Thevenet et al. [2007] hadde intervaller med lavest hastighet 30/30sek med 100% av farten ved vo2maks. Under alle testene ble gassutveksling målt, dette innebar variabler som ventilasjon, pustefrekvens, oksygen, og  $co_2$ . I tillegg målte alle studiene laktat mens Billat et al. [2001] og Billat et al. [2000] var de eneste som ikke målte hjertefrekvens. Av disse målingene var oksygenmålingen den eneste variabelen som direkte svarte på hypotesen i alle studiene.

| Studie                        | Deltakere                               | Studiedesign   |
|-------------------------------|---|--|
| Thevenet et al. [2007]        | N= 9<br>Kjønn= M<br>Alder= $16,6 \pm 1$ | RCT<br>Løpetest: 400m løpebane<br>Intervall 1: 30/30sek 100% fart ved VO2maks<br>Intervall 2: 30/30sek 110% fart ved VO2maks                                   |
| Wakefield and Glaister [2009] | N= 7<br>Kjønn= M<br>Alder= $22 \pm 5$   | RCT<br>Løpetest: Løpemølle 1% stigning<br>Intervall 1: 20/20sek på 105 og 115%<br>Intervall 2: 25/20sek på 105 og 115%<br>Intervall 3: 30/20sek på 105 og 115% |

| Studie                  | Deltakere                         | Studiedesign  |
|-------------------------|-----------------------------------|---|
| Billat et al.<br>[2001] | N= 7<br>Kjønn= M<br>Alder= 51±6   | RCT<br>Løpetest: 400m løpebane<br>Intervall 1: 15/15sek på 90-80%<br>Intervall 2: 15/15sek på 100-70%<br>Intervall 3: 15/15se på 60 x 110%                    |
| Billat et al.<br>[2000] | N= 8<br>Kjønn= M<br>Alder= 34±6   | RCT<br>Løpetest: 400m løpebane<br>Intervall 1: 30/30 sek på 100%<br>Intervall 2: 30/30sek på laktattersekel   |
| Dupont et al.<br>[2002] | N= 9<br>Kjønn= M<br>Alder= 25±2,8 | RCT<br>Løpetest 200m løpebane<br>Løpetest 1: 15/15sek på 110%<br>Løpetest 2: 15/15sek på 120%<br>Løpetest 3: 15/15sek på 130%<br>Løpetest 4: 15/15sek på 140% |

Alle studiene adresserte hypotesen deres i konklusjonen, og diskuterte rundt resultatene de fikk. I studiene til Billat et al. [2000] og Thevenet et al. [2007] foreslo studiene at ytterligere forskning krevdes for å kunne fastslå resultatet helt. Dette var fordi forskingen deres ikke så på langtidsvirkningen av intervallene og det mente de var nødvendig for å kunne fastslå om den typen intervall som prestere best i deres studie faktisk ville være den beste for utvikling av kardiovaskulære systemer over tid. I tillegg var det interessant å se at alle studiene antatt Dupont et al. [2002] kun snakket om sine egne resultater i konklusjonen. Mens Dupont et al. [2002] refererte til andre studier i sin egen konklusjon.

Totalt sett syns jeg at alle studiene gjør en grei oppgave med statistiske tester. Likevel savner jeg at alle studiene kunne forklart antall deltakere, dette er fordi fler deltakere ville gitt en mye bedre statistisk styrke til studien og det er merkelig av forfattere og ikke annerkjenne dette. De kunne også gjøre flere statistiske tester, for å styrke resultatet de kommer med i konklusjonen og for å vise bedre frem at det resultatet de kom frem til faktisk stemmer.



## Chapter 4

# Analyzing repeated measures experiments

### 4.1 Innføring

Hensikten med denne rapporten er å undersøke effekten av styrketrening på muskelstyrke og muskelvekst. Forskning på styrketrening har vist at det forbedrer muskelstyrke, muskel masse, beintetthet og bindevevstykkelse [Kraemer et al., 2002]. For å designe et treningsprogram for styrketrening krever det at man tar en del variabler til betraktning, dette inkluderer, treningsfrekvens, intensitet og volum av programmet [Hass et al., 2001]. Innenfor forskningen på styrketrening er det stor forskjell hva er fokuset, og noe av det mest interessante er å se på forskjellen mellom antall sett, som har blitt gjort i flere studier [Krieger, 2010]. Tidligere har det blitt argumentert for at ett sett (single-sett) per øvelse er alt som er nødvendig for hele populasjonen og at man ikke får større utbytte av flere sett [Carpinelli and Otto, 1998b]. Mens flere andre studier har argumentert for at det er et større utbytte av flere sett (multiple-sett) per øvelse Galvão and Taaffe [2005a]; Humburg et al. [2007b]. Ettersom det er motstridende artikler innenfor dette området ønsker jeg i denne studien å undersøke om det er en større effekt i å trene multiple-sett mot single-sett når det kommer til utvikling av muskelstyrke og muskelvekst.

### 4.2 Metode

#### 4.2.1 Forsøkspersoner

I denne studien var det med 41 deltakere med kvalifikasjonskrav mellom 18 og 40, både kvinner og menn og å ikke røyke. Deltakere ble ekskludert hvis de

hadde intoleranse til lokal bedøvelse, en treningshistorikk på mer enn en ukentlig styrketreningsøkt i løpet av de siste 12 måneder, nedsatt muskelstyrke på grunn av nåværende eller tidligere skade, og inntak av medisiner som kunne påvirke styrketreningen. 7 deltakere ble utelukket fra studien ettersom de ikke oppfylte kravet om å fullføre minst 85% av de planlagte treningsøktene. Beinøvelser ble utført på ett og ett bein for å kunne tillate for individuelle forskjell i treningsvolum. For hver deltaker ble det randomisert tildelt for hvert bein enten ett sett (single- sett) eller tre sett (multiple- sett). Muskelstyrken ble målt ved baseline og etter treningsintervensjonen. Muskelbiosien ble målt fra begge bein (vastus lateralis) ved baseline og etter 12 uker med trening i uthvilt tilsand.

### 4.2.2 Treningsprotokol

Før alle treningsøktene ble det utført en standardisert oppvarmingsrutine som inneholdt 5 min sykling og 10 repetisjoner med kroppsvekt av pushups, situps, rygghev og squats og 10 repetisjoner på 50% av 1RM på hver øvelse de skal trene. Beinøvelsene ble utført i følgende rekkefølge: ettbeinsbeinpress, kne fleksjon og kneekstensjon utført som enten single- sett eller multiple- sett. Etter beinøvelsene utførte deltakerne to sett av bilateral benkpress, nedtrekk og enten skulderpress eller sittende roing. Pausene mellom settene var på 90- 180 sekunder. Intensiteten på treningsøktene ble gradvis økt gjennom treningsperioden, deltakerne utførte 10RM de første 2 ukene, etterfulgt av 8RM i 3 uker og 7 uker med 7RM. Treningsøktene med maksimal innsats hadde minst 48t mellom og treningsøkter og submaksimale økter hadde 24t mellom. For å hjelpe med restitusjonen ble en standardisert drikke gitt etter hver øvelse.

Maksimal styrke ble beskrevet som 1RM i ettbeinsbeinpress og kneekstensjon. Testen for hver øvelse startet med en standardisert spesifikk oppvarming før 1RM ble funnet ved å øke motstanden progressivt til deltakeren ikke lenger klarte å løfte vekten.

Kroppssammensetning til deltakerne ble bestemt før og etter treningsperioden ved bruk av en DXA- scan. Før DXA målingene måtte deltakerne være fastende i 2timer og måtte unngå hard fysisk aktivitet 48t før.

### 4.2.3 Statestikk

All deskriptiv data er presentert som et gjennomsnitt av prosentvis ending med standardavvik. P- verdier er regnet ut ved en ANCOVA modell, på endringsscoren fra post til pre. Statistiske tester ble utført i RStudio (versjon RStudio 1.4.1717; R Foundation for Statistics Computing, Vienna, AT)

### 4.3 Resultater

Totalt, førte 12 uker med styrketrening til en signifikant økning i muskelstyrke på  $31 \pm 14\%$  for multiple- sett og en økning på  $25 \pm 13\%$  for multiple- sett ( $P=0.001$ ) styrkeøkningen mellom gruppene var signifikant på ( $P= 0.02988$ ). Økningen i muskelvekst var også signifikant på  $3,3 \pm 4\%$  for multiple- sett og  $2 \pm 4\%$  for single- sett ( $P=0.001$ ) men her var det ikke signifikant økning mellom gruppene ( $P=0.193$ , fig 4.2) .

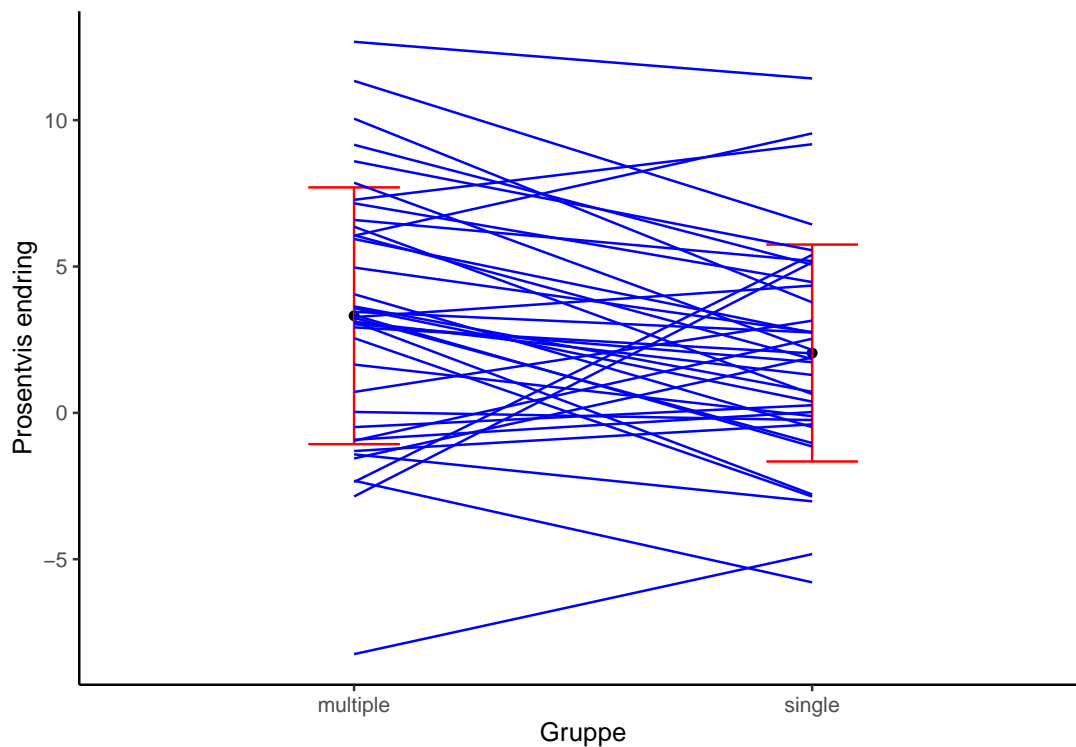


Figure 4.1: Figuren viser prosentvis økning i muskelvekst mellom multiple- sett og single- sett ved posttest etter 12 uker med styrketrening ( $P= 0.193$ ) (#fig:Prosentvis økning figur dexta)

### 4.4 Diskusjon

I denne studien viste det at multiple- sett styrketrening førte til bedre økning i muskelstyrke enn single- sett trening. Dette resultatet stemmer overens med andre studier som har funnet at flere sett er bedre enn en [Galvão and Taaffe, 2005c, Humburg et al., 2007a].

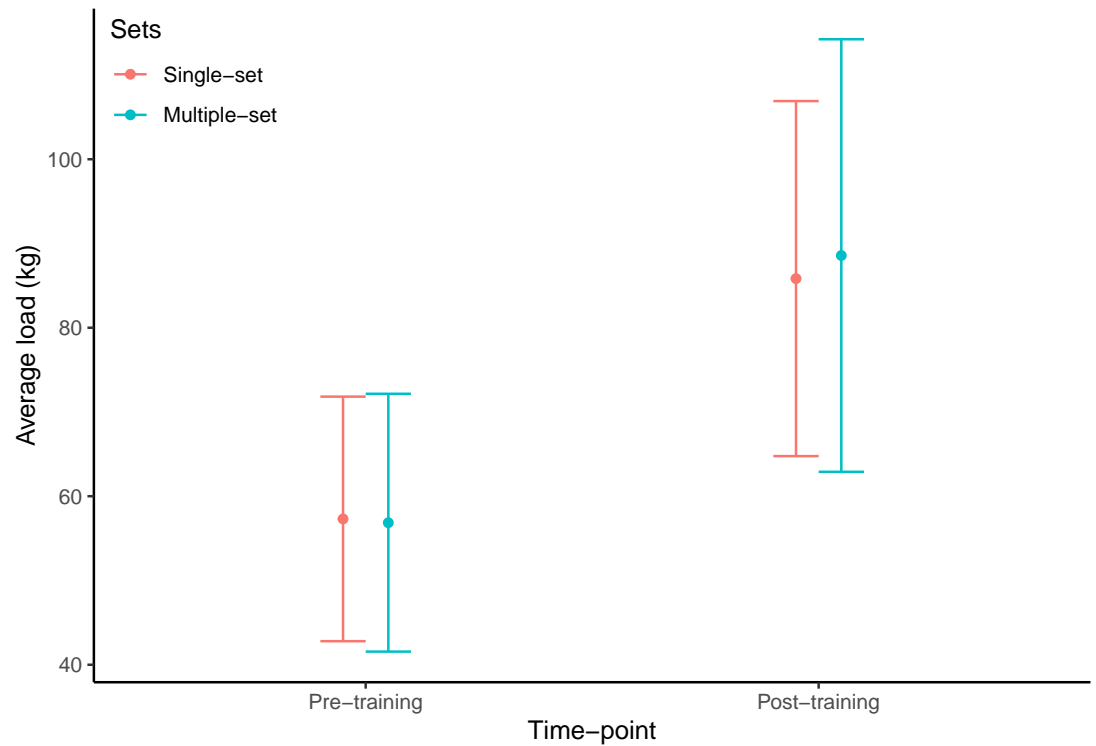


Figure 4.2: Figuren viser økning i muskelstyrke mellom multiple- sett og singe-sett ved pretest og posttest etter 12 uker med styrketrening ( $P = 0.029$ ,  $\text{diff} = -0.029$ )

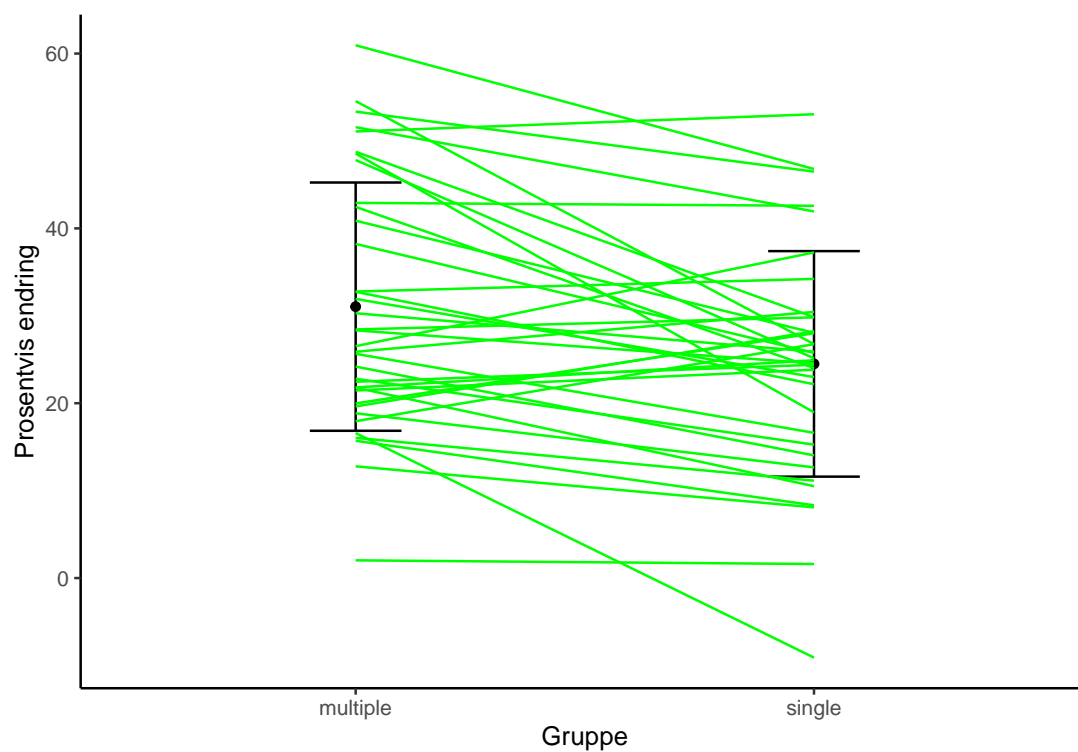


Figure 4.3: Figuren viser prosentvis økning i muskelstyrke mellom multiple- sett og singe- sett ved posttest etter 12 uker med styrketrening (P- 0.029) (#fig:Prosentvis økning figur styrke)

Likevel samsvarer ikke resultatet i denne studien med [Carpinelli and Otto, 1998c]. Der kom de frem til at det ikke var noen signifikant forskjell i størrelsen på styrke økningen mellom 1- sett og multiple sett. I studien til [Carpinelli and Otto, 1998a] så de på studien til [Reid et al., 1987] som fant ut at den gjennomsnittlige økningen i styrke var på 17,7% for single- sett gruppen og på 17,9% multiple- sett. Sammenlignet med vårt resultat som var på henholdsvis 25% og 31% på single- sett og multiple- sett. Disse forskjellene i konklusjon kan muligens bli forklart av [Galvão and Taaffe, 2005b] konkluderer med at styrketrening med single- sett øvelser er nok til å signifikant øke muskel funksjon, selv om muskelstyrke vil bli bedre ved et høyere treningsvolum som multiple- sett gir.

#### 4.4.1 Konklusjon

Totalt sett viste denne studien at multiple- sett ga en bedre økning i muskelstyrke sammenlignet med single- sett. Selv om begge gruppene hadde økning vil det kunne være mer hensiktsmessig å trene styrketrening med multiple- sett.

# Bibliography

Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533 (7604):452–454, 05 2016. doi: 10.1038/533452a. URL <https://www.nature.com/articles/533452a>. Bandiera\_abtest: a Cg\_type: Nature Research Journals Number: 7604 Primary\_atype: News Publisher: Nature Publishing Group Subject\_term: Peer review;Publishing;Research management Subject\_term\_id: peer-review;publishing;research-management.

David R. Jr Edward T. Howley Bassett. Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Medicine Science in Sports Exercise*, 32(1):70, 01 2000. URL [https://journals.lww.com/acsm-msse/Fulltext/2000/01000/Limiting\\_factors\\_for\\_maximum\\_oxygen\\_uptake\\_and.12.aspx](https://journals.lww.com/acsm-msse/Fulltext/2000/01000/Limiting_factors_for_maximum_oxygen_uptake_and.12.aspx).

V. Billat, J. Slawinski, V. Bocquet, P. Chassaing, A. Demarle, and J. Koralsztein. Very short (15 s - 15 s) interval-training around the critical velocity allows middle-aged runners to maintain  $\dot{V}O_2$  max for 14 minutes. *International Journal of Sports Medicine*, 22(03):201–208, 12 2001. doi: 10.1055/s-2001-16389. URL <http://www.thieme-connect.de/DOI/DOI?10.1055/s-2001-16389>.

Véronique L. Billat, Jean Slawinski, Valery Bocquet, Alexandre Demarle, Laurent Lafitte, Patrick Chassaing, and Jean-Pierre Koralsztein. Intermittent runs at the velocity associated with maximal oxygen uptake enables subjects to remain at maximal oxygen uptake for a longer time than intense but sub-maximal runs. *European Journal of Applied Physiology*, 81(3):188–196, 01 2000. doi: 10.1007/s004210050029. URL <http://link.springer.com/10.1007/s004210050029>.

Alexander Bird. Understanding the replication crisis as a base rate fallacy. *The British Journal for the Philosophy of Science*, pages 000–000, 12 2020. doi: 10.1093/bjps/axy051. URL <https://www.journals.uchicago.edu/doi/10.1093/bjps/axy051>.

G. Borg. Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, 2(2):92–98, 1970. PMID: 5523831.

- Ralph N. Carpinelli and Robert M. Otto. Strength training. *Sports Medicine*, 26(2):73–84, 08 1998a. doi: 10.2165/00007256-199826020-00002. URL <https://doi.org/10.2165/00007256-199826020-00002>.
- Ralph N. Carpinelli and Robert M. Otto. Strength training. *Sports Medicine*, 26(2):73–84, 08 1998b. doi: 10.2165/00007256-199826020-00002. URL <https://doi.org/10.2165/00007256-199826020-00002>.
- Ralph N. Carpinelli and Robert M. Otto. Strength training. *Sports Medicine*, 26(2):73–84, 08 1998c. doi: 10.2165/00007256-199826020-00002. URL <https://doi.org/10.2165/00007256-199826020-00002>.
- Grégory Dupont, Nicolas Blondel, Ghislaine Lensel, and Serge Berthoin. Critical velocity and time spent at a high level of for short intermittent runs at supramaximal velocities. *Canadian Journal of Applied Physiology*, 27(2):103–115, 04 2002. doi: 10.1139/h02-008. URL <http://www.nrcresearchpress.com/doi/10.1139/h02-008>.
- Daniel A. Galvão and Dennis R. Taaffe. Resistance exercise dosage in older adults: Single- versus multiset effects on physical performance and body composition. *Journal of the American Geriatrics Society*, 53(12):2090–2097, 2005a. doi: 10.1111/j.1532-5415.2005.00494.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.2005.00494.x>. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.2005.00494.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.2005.00494.x).
- Daniel A. Galvão and Dennis R. Taaffe. Resistance exercise dosage in older adults: Single- versus multiset effects on physical performance and body composition. *Journal of the American Geriatrics Society*, 53(12):2090–2097, 2005b. doi: 10.1111/j.1532-5415.2005.00494.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.2005.00494.x>. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.2005.00494.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.2005.00494.x).
- Daniel A. Galvão and Dennis R. Taaffe. Resistance exercise dosage in older adults: Single- versus multiset effects on physical performance and body composition. *Journal of the American Geriatrics Society*, 53(12):2090–2097, 2005c. doi: 10.1111/j.1532-5415.2005.00494.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.2005.00494.x>. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.2005.00494.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.2005.00494.x).
- C. J. Hass, M. S. Feigenbaum, and B. A. Franklin. Prescription of resistance training for healthy populations. *Sports Medicine (Auckland, N.Z.)*, 31(14):953–964, 2001. doi: 10.2165/00007256-200131140-00001. PMID: 11735680.
- Carl Hempel. *Philosophy of Natural Science*. Princeton: Prentice Hall. 1966.
- W. G. Hopkins. Measures of reliability in sports medicine and science. *Sports Medicine (Auckland, N.Z.)*, 30(1):1–15, 07 2000. doi: 10.2165/00007256-200030010-00001. PMID: 10907753.



Stephen B. Hulley. *Designing clinical research*. Lippincott Williams Wilkins, 4th ed. edition, 2013. Book Title: Designing clinical research.

Hartmut Humberg, Hartmut Baars, Jan Schröder, Rüdiger Reer, and Klaus-Michael Braumann. 1-set vs. 3-set resistance training: a crossover study. *Journal of Strength and Conditioning Research*, 21(2):578–582, 05 2007a. doi: 10.1519/R-21596.1. PMID: 17530985.

Hartmut Humberg, Hartmut Baars, Jan Schröder, Rüdiger Reer, and Klaus-Michael Braumann. 1-set vs. 3-set resistance training: a crossover study. *Journal of Strength and Conditioning Research*, 21(2):578–582, 05 2007b. doi: 10.1519/R-21596.1. PMID: 17530985.

William J. Kraemer, Nicholas A. Ratamess, and Duncan N. French. Resistance training for health and performance. *Current Sports Medicine Reports*, 1(3):165–171, 06 2002. URL [https://journals.lww.com/acsm-csmr/Abstract/2002/06000/Resistance\\_Training\\_for\\_Health\\_and\\_Performance.7.aspx](https://journals.lww.com/acsm-csmr/Abstract/2002/06000/Resistance_Training_for_Health_and_Performance.7.aspx).

James W. Krieger. Single vs. multiple sets of resistance exercise for muscle hypertrophy: A meta-analysis. *The Journal of Strength Conditioning Research*, 24(4):1150–1159, 04 2010. doi: 10.1519/JSC.0b013e3181d4d436. URL [https://journals.lww.com/nsca-jscr/Fulltext/2010/04000/Single\\_vs\\_Multiple\\_Sets\\_of\\_Resistance\\_Exercise.36.aspx](https://journals.lww.com/nsca-jscr/Fulltext/2010/04000/Single_vs_Multiple_Sets_of_Resistance_Exercise.36.aspx).

Samir Okasha. *Philosophy of Science: A Very Short Introduction (2nd ed.)*. 2016.

C. M. Reid, R. A. Yeater, and I. H. Ullrich. Weight training and strength, cardiorespiratory functioning and body composition of men. *British Journal of Sports Medicine*, 21(1):40–44, 03 1987. doi: 10.1136/bjsm.21.1.40. URL <https://bjsm.bmj.com/content/21/1/40>. Publisher: British Association of Sport and Exercise Medicine Section: Research Article PMID: 3580729.

Delphine Thevenet, Magaly Tardieu, Hassane Zouhal, Christophe Jacob, Ben Abderraouf Abderrahman, and Jacques Prioux. Influence of exercise intensity on time spent at high percentage of maximal oxygen uptake during an intermittent session in young endurance-trained athletes. *European Journal of Applied Physiology*, 102(1):19–26, 11 2007. doi: 10.1007/s00421-007-0540-6. URL <http://link.springer.com/10.1007/s00421-007-0540-6>.

Delphine Thevenet, Magaly Tardieu, Hassane Zouhal, Christophe Jacob, Ben Abderraouf Abderrahman, and Jacques Prioux. Influence of exercise intensity on time spent at high percentage of maximal oxygen uptake during an intermittent session in young endurance-trained athletes. *European Journal of Applied Physiology*, 102(1):19–26, 11 2007. doi: 10.1007/s00421-007-0540-6. URL <http://link.springer.com/10.1007/s00421-007-0540-6>.

Benjamin R Wakefield and Mark Glaister. Influence of work-interval intensity and duration on time spent at a high percentage of

journal/jsr/04.03/00124278-200912000-00017/ov0312\_\_5/v/2021-02-09t093646z/r/image-png o2max during intermittent supramaximal exercise. *Journal of Strength and Conditioning Research*, 23(9): 2548–2554, 12 2009. doi: 10.1519/JSC.0b013e3181bc19b1. URL <https://journals.lww.com/00124278-200912000-00017>.