

Eksamen IDR4000

Kandidatnummer 501

2024-11-22

Table of contents

Forord	4
1 Reliabilitet	5
1.1 Introduksjon	5
1.2 Metode	5
1.2.1 Gjennomføring av $\dot{V}O_{2max}$ -test	5
1.2.2 Standardisering	6
1.2.3 Intern validering	6
1.3 Resultater	6
1.4 Diskusjon	8
1.5 Konklusjon	9
2 Regresjonsmodeller	10
2.1 Introduksjon Regresjonsmodeller	10
2.2 Laktatterskel	10
2.2.1 Introduksjon	10
2.2.2 Metode	11
2.2.3 Resultat	11
2.2.4 Diskusjon	11
2.2.5 Konklusjon	11
2.3 Estimering av størrelse på DNA-fragmenter	13
2.3.1 Introduksjon	13
2.3.2 Metode	13
2.3.3 Resultat	14
2.3.4 Diskusjon	15
2.3.5 Konklusjon	16
2.4 Tolke en regresjonstabell	16
2.4.1 Introduksjon	16
2.4.2 Metode	16
2.4.3 Resultat	17
2.4.4 Diskusjon	17
2.4.5 Konklusjon	18
3 Statistisk inferens	19
3.1 Introduksjon	19

3.2	<i>T</i> -verdi, <i>P</i> -verdi, Standard Error (SE) og Estimat	19
3.3	To-sidet test (two-tailed test)	20
3.4	Standardavvik og standardfeilen (SE)	20
3.5	Effekten av utvalgsstørrelse for statistisk styrke	21
3.6	Koder for simuleringer	23
4	Studiedesign	25
4.1	Introduksjon	25
4.2	Metode	25
4.3	Resultater	26
4.4	Diskusjon	26
4.5	Konklusjon	27
5	Repeterte målinger	29
5.1	Introduksjon	29
5.2	Metode	29
5.2.1	Utvalg og studieoversikt	29
5.2.2	Spesifikke beskrivelser av testmetode	30
5.2.3	Styrketreningsprotokoll	30
5.2.4	Muskulær Styrkevurdering	31
5.2.5	Måling av Kroppssammensetning og Mager Muskelmasse	31
5.2.6	Dataanalyse og Statistikk	31
5.3	Resultater	31
5.3.1	Treningsvolums effekt på mager muskelmasse i underekstremitetene	31
5.3.2	Treningsvolums effekt på kraftutvikling i underekstremitetene	32
5.4	Diskusjon	32
5.5	Konklusjon	34
6	Vitenskapsfilosofi	35
6.1	Falsifikasjonisme	35
6.2	Bayesianisme	36
7	RNA ekstraksjon og analyse	38
7.1	Introduksjon	38
7.2	Metode	38
7.2.1	Overordnet metode	38
7.2.2	Detaljert fremgangsmåte for qPCR	39
7.3	Resultater	41
7.4	Diskusjon	41
7.5	Konklusjon	42
	Referanser	43

Forord

https://github.com/olekjolner/eksamen_idr4000_ole

1 Reliabilitet

1.1 Introduksjon

Reliabilitet er et viktig begrep innen treningsfysiologi og medisin, da det omhandler påliteligheten av målinger og tester. I forskning og klinisk praksis er det avgjørende at instrumenter og prosedyrer gir stabile resultater over tid, under samme forhold og for ulike individer. Hopkins (2000) definerer reliabilitet som graden av overensstemmelse mellom gjentatte målinger, og understreker at høy reliabilitet er nødvendig for å trekke pålitelige konklusjoner om effekten av trening eller behandling (Hopkins, 2000). Innen treningsfysiologi kan lav reliabilitet i målingene føre til unøyaktige vurderinger av treningsprogresjon, noe som kan resultere i at treningsprogrammet ikke er tilpasset det faktiske ferdighetsnivået en innehar på gitt tidspunkt. Dette kan føre til manglende fremgang, ettersom treningen ikke er optimalt justert for den enkeltes kapasitet.

1.2 Metode

1.2.1 Gjennomføring av $\dot{V}O_{2max}$ -test

Testen som skal gjennomføres er en $\dot{V}O_{2max}$ -test, der testdeltakeren starter på en forhåndsbestemt wattverdi (motstand). Denne wattverdien økes gradvis hvert minutt, avhengig av den valgte protokollen, og testdeltakeren sykler til en ikke lenger klarer å fortsette. Testen avsluttes dersom deltakerens tråkkfrekvens synker under 60 rpm (rotasjoner per minutt). Under testen skal deltaker forbli sittende på sykkelen.

Testdeltakeren vil få informasjon om hvor lenge de har syklet underveis i testen. Derimot vil de ikke ha tilgang til annen informasjon som puls, VO_2 eller lignende data underveis. Testlederen vil informere om hvor lang tid det er igjen til neste wattøkning og hvilken watt testdeltakeren sykler på for øyeblikket.

Testlederen vil mot slutten av testen bidra til å pushe testdeltakeren for å få ut det siste av deres kapasitet. Hvordan denne “pushingen” gjennomføres, kan variere mellom testledere, men det er viktig at tilnærmingen til hver deltaker er så konsekvent som mulig. Borg-skalaen (6-20) vil benyttes for å vurdere utmattelsen hos testdeltakeren. Ved testslutt vil deltaker bli bedt om å gi et tall på skalaen, der 6 tilsvarer “ingen anstrengelse” og 20 “maksimal anstrengelse”.

Ett minutt etter testslutt tas en laktatprøve fra testdeltakeren for å estimere maksimal laktatverdi (La_{max}). Etter at prøven er tatt, noterer testlederen ned hvor lenge testdeltakeren har syklet, og henter deretter ut relevant data fra testen via en rapport fra Vyntus-systemet.

1.2.2 Standardisering

Testdeltakeren skal spise sitt siste store måltid senest 2 timer før teststart. Det er tillatt å innta en mindre karbohydratkilde (eks: banan, energibar, gel) senest 30 minutter før teststart. Koffeininntak skal gjøres som normalt og testene gjennomføres på samme tidspunkt på dagen så godt det lar seg gjøre. Deltakeren anbefales å forsøke å opprettholde sin vanlige døgnrytme gjennom testperioden. For å forberede kroppen på testen, er det viktig at deltakerne ikke gjennomfører hard trening på underekstremitetene dagen før testen.

Oppvarmingen før testen ble satt til 5 minutter, der deltaker syklet på økende intensitet i intervaller på 2-2-1 minutt. Eksempelvis 2 minutter på 150W, 2 minutter på 175W og 1 minutt på 220W.

Testprotokollen er den samme for alle tester. Et eksempel på testprotokoll var start på 200W med en økning på 20W hvert minutt til utmattelse. Individuelle tilpasning mtp belastning ble gjort for hver deltaker, både ved oppvarming og testprotokoll.

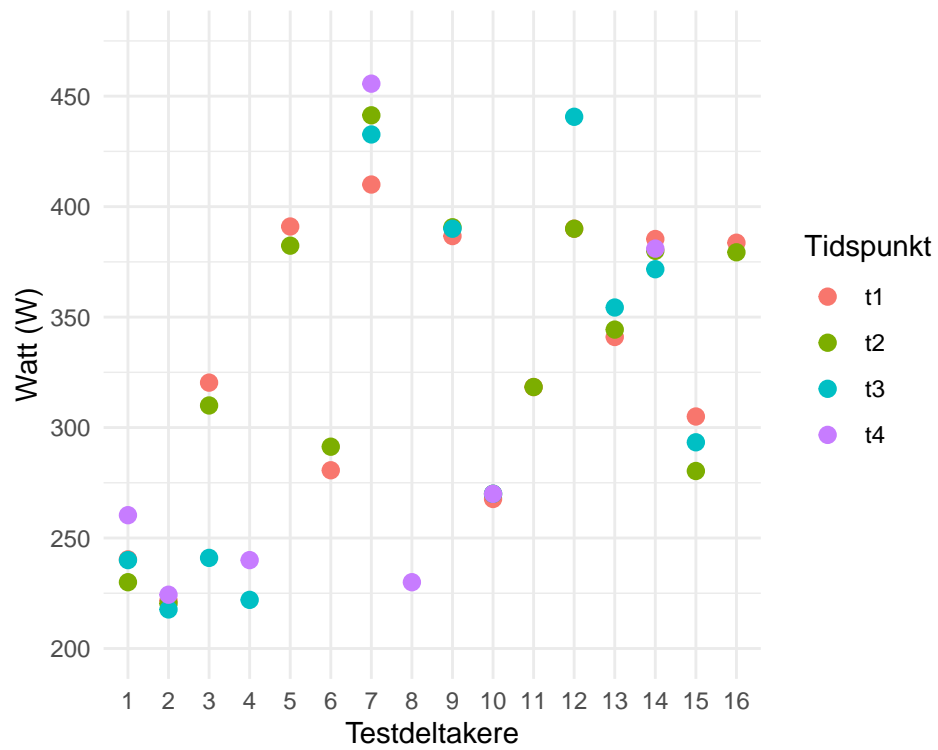
1.2.3 Intern validering

Både tekniske og biologiske faktorer kan potensielt påvirke resultatene og true den interne validiteten av testen (Halperin *et al.*, 2015). Ved å beskrive og etablere testprosedyrer, samt standardisere testen, ønsker vi å skape rammer som sikrer at vi kontrollerer for mulige konfunderende faktorer som kan påvirke testprestasjon. Dette vil bidra til å redusere variasjoner og sikre at resultatene reflekterer faktiske prestasjoner, snarere enn påvirkninger fra ytre faktorer.

Selv om vi har jobbet for å sikre god intern validitet, er det faktorer som vil kunne påvirke resultatene grunnet individuelle forskjeller i måten å opptre som testleder, humør og mental status til testdeltager (Halperin *et al.*, 2015).

1.3 Resultater

I denne oppgaven presenteres data for $watt_{max}$. I Figure 1.1 observeres $watt_{max}$ for samtlige deltakere ved de ulike testtidspunktene. Det er tydelig at enkelte deltakere hadde betydelige variasjoner i sluttverdiene for $watt_{max}$ mellom testene.



Figur 1.1: Watt-verdien testdeltaker holdt de siste 30 sek av VO₂max-testen ved de ulike tidspunktene

I Table 1.1 er det beregnet gjennomsnitt og standardavvik ved hvert testtidspunkt. En kan se at den gjennomsnittlige $watt_{max}$ -verdien synker fra t1 til t4, og at avviket i testene er større ved t3 og t4, sammenliknet med de to første testene.

Tabell 1.1: Gj.snitt og st.avvik for $watt_{max}$ ved de ulike testtidspunktene

Tidspunkt	$Watt_{max}$
t1	332 (62)
t2	331 (66.6)
t3	316 (84.7)
t4	294 (88.8)

Til slutt observerte en $watt_{max}$ -verdier ved t1 og t2 og beregnet gjennomsnitt, standardavvik, den typiske feilen (te) og variasjonskoeffisienten (cv%) (Table 1.2).

Tabell 1.2: Summerende statistikk

Gj.snitt	St.avvik	Typisk feil	Variasjonskoeffisient
331.07	12.59	8.90	2.69

Verdier gjelder for test 1 og 2

1.4 Diskusjon

Den gjennomsnittlige watt-verdien som testdeltakerne holdt de siste 30 sekundene av $\dot{V}O_{2max}$ -testen ved test 1 og 2 var 331.1 (± 12.6) og den typiske feilen ble målt til 8.9. Ifølge Hopkins (2000) er den typisk feilen, den gjennomsnittlige variasjonen ved en test. Den typiske feilen gir en indikasjon på den forventede variasjonen ved en retest (Hopkins, 2000). I dette tilfellet kan en forvente en variasjon på 2.69 % ved en eventuell retest. Dette antyder at det er en viss grad av variasjon som kan forventes i målingene, men at resultatene er tilstrekkelig stabile til å trekke pålitelige konklusjoner om deltakerens kapasitet.

Årsaken til at test 1 og test 2 ble sammenlignet, skyldes et betydelig redusert antall deltakere som valgte å gjennomføre test 3 og 4. På grunn av dette ble t1 og t2 ansett som de mest pålitelige og relevante testene å sammenligne i denne studien, da de ga det største og mest representative datagrunnlaget.

Andre potensielle konfunderende faktorer som kan ha påvirket resultatene fra t1 til t4 er deltakerens mentale tilstand, søvnkvalitet, kosthold, samt testlederens tilnærming under gjennomføringen av testen. Disse faktorene kan variere fra test til test og påvirke deltakerens prestasjon, noe som kan forklare deler av variasjonen i testresultatene. For eksempel kan en

deltaker som har sovet dårlig og er i dårlig fysisk form, prestere dårligere sammenlignet med en deltaker som er godt uthvilt og mentalt forberedt. Individuelle forskjeller i næringsinntak før testen kan også spille en rolle i testens reliabilitet, da lavt energinivå kan hemme prestasjon under testen.

En typisk feil på 8,9 watt indikerer at standardiseringen av testene var effektiv i å sikre stabile og konsistente resultater. Den systematiske tilnærmingen bidro sannsynligvis til å minimere påvirkningen fra eksterne faktorer, og dermed redusere variasjonen som kunne ha oppstått under testene.

Resultatene fra disse testene gir trolig et pålitelig estimat av deltakerens maksimale kapasitet under de gitte testforholdene, og den relativt lave feilmarginen tyder på at endringer i watt_{max} kan være et nyttig mål for treningsfremgang. Det er dog viktig å huske på at betydelige endringer i resultatene bør undersøkes nærmere, ettersom de kan skyldes reelle forbedringer eller eksterne faktorer som stress eller helsetilstand på testdagen.

1.5 Konklusjon

En kan konkludere med at resultatene fra test 1 og 2 gir et pålitelig estimat av deltakerens maksimale kapasitet, da en typisk feil på 8,9 watt indikerer stabile og konsistente målinger. Standardiseringen av testene har trolig redusert påvirkningen fra eksterne faktorer og den observerte variasjonen i watt_{max} er i dette tilfellet tilstrekkelig lav til at man kan trekke pålitelige konklusjoner om deltakerens fysiske kapasitet. Betydelige endringer i prestasjon bør dog vurderes grundig, da de kan skyldes både reelle fysiske forbedringer og eksterne faktorer som stress, søvnkvalitet og kosthold på testdagen.

2 Regresjonsmodeller

2.1 Introduksjon Regresjonsmodeller

En regresjonsmodell er ifølge Pisica et al. (2022) en modell som kvantifiserer forholdet mellom en eller flere uavhengige variabler og en avhengig variabel. Innen medisin er regresjon den analysemetoden som er hyppigst anvendt, og det finnes flere forskjellige regresjonsmodeller. De vanligste er lineær regresjon, polynominal regresjon og logistisk regresjon. Valget av regresjonsmodell avhenger av hvilke data som er tilgjengelige (Pisică *et al.*, 2022).

En lineær regresjonsmodell er en modell der en kan estimere verdien av en avhengig variabel basert på verdien av andre kjente uavhengige variabler (Pisică *et al.*, 2022). I en slik modell benyttes en rett linje for å lage en modell som beskriver dataen. Følgende funksjon benyttes for å skape det lineære plottet:

$$y_i = b_0 + b_1 x_i + e_i$$

der y_i er den avhengige variabelen, x_i er den uavhengige variabelen, b_0 er skjæringspunktet til grafen og b_1 stigningstallet til grafen. e_i representerer feilen (residualen) for observasjonen i .

I denne rapporten vil ulike tema bli presentert. Det være seg laktatterskel, estimering av størrelse på DNA-fragment og tolkning av regresjonsmodeller. De ulike temaene presenteres med egne delkapitler. Sammenhengen mellom oppgavene er at de omhandler bruken av regresjon og hvordan vi kan tolke denne typen dataanalyse.

2.2 Laktatterskel

2.2.1 Introduksjon

Ifølge Antonutto og Di Prampero (1995) refererer laktatterskel til den grensen der kroppen begynner å produsere mer laktat enn den evner å omdanne. Laktatterskel brukes ofte som et mål på intensitet, samtidig sier det noe om treningsstatus og evne til å prestere (Antonutto & Di Prampero, 1995).

I denne delen undersøkes forholdet mellom laktatnivå og watt, samt tilpasningen av regresjonsmodeller for å beskrive dette forholdet.

2.2.2 Metode

I denne analysen ble det hentet data fra *cyclingstudy*-datasettet som en finner i pakken *exsci-data* i R. Fra datasettet ble deltaker 10 benyttet som eksempel. Datasettet ble omstrukturert til et mer passende format (tidy data) for deretter å lage et hensiktsmessig plot for å illustrere relasjonen mellom laktat og watt. Det ble utført ulike regresjonsmodeller for å fremstille forholdet mellom laktat og watt (se Figure 2.1). Laktatverdiene og wattverdiene ble plottet som gule punkter i grafen. Deretter ble en stiplet svart linje tegnet for å forbinde punktene. En ønsket å se på watt-verdier ved henholdsvis 2 og 4 mmol x L-1 laktat. Derfor ble de lagt til hjelpende horisontale (rød og lilla) og vertikale (stiplede svarte) linjer for å synliggjøre skjæringspunkter på grafen.

For å analysere sammenhengen mellom laktat og watt ble det utført regresjonsanalyser ved hjelp av tre forskjellige modeller: først en lineær modell (blå linje), deretter en andregradsmodell (grønn linje) og til slutt en tredjegradsmodell (rosa linje). Formålet med analysen var å vurdere hvilken av de tre modellene som best predikerer forholdet mellom laktat og watt.

2.2.3 Resultat

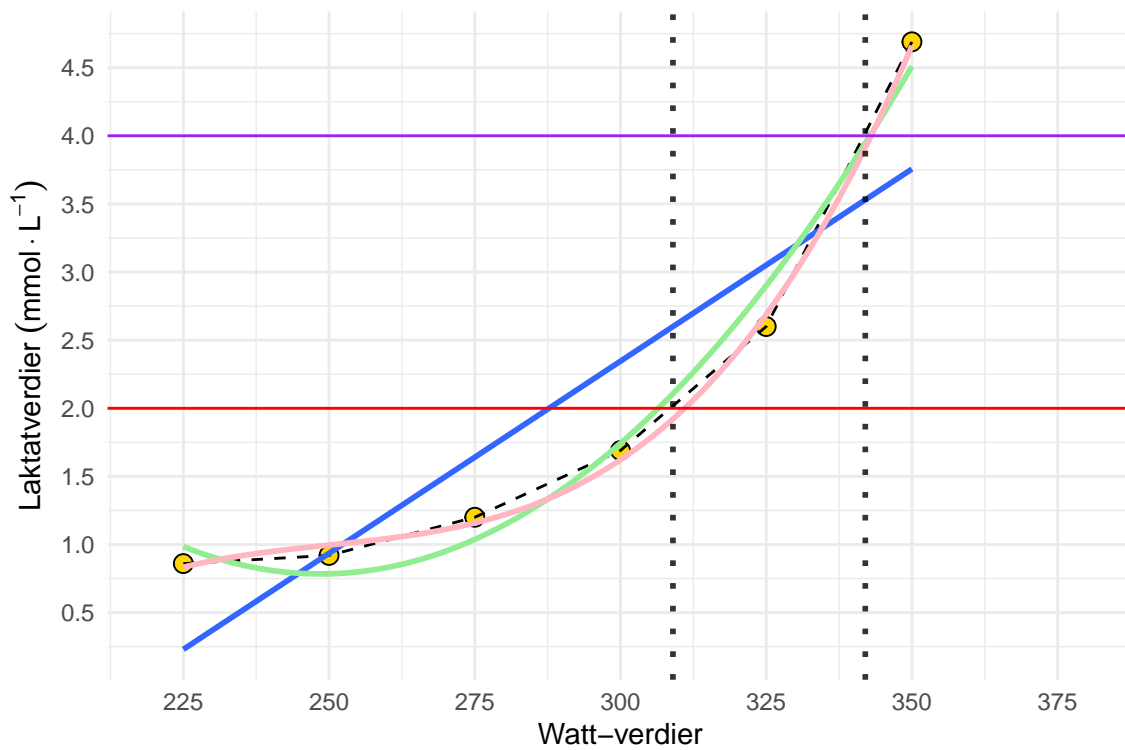
2.2.4 Diskusjon

Den lineære regresjonsmodellen estimerer watt-verdier for henholdsvis 2 og 4 mmol x L-1 intensitet til å være 288 W og 359 W. De reelle watt-verdiene for henholdsvis 2 og 4 mmol x L-1 er tilnærmet 310 og 340 W. Dette gir et avvik på 22 W og 19 W, som indikerer at den lineære modellen ikke er spesielt anvendbar i dette tilfellet. Avviket kan tyde på at forholdet mellom laktat og watt ikke er lineært, og at en lineær modell ikke fanger opp den reelle kurven i dataene.

Som vist i Figure 2.1, gir andregrads- og tredjegradsmodellen en betydelig bedre tilpasning til dataene. Begge disse modellene viser seg å gi betydelig bedre estimat på de faktiske målingene. Dette tyder på at en mer kompleks modell gir mer nøyaktige estimater av watt-verdier ved forskjellige laktatnivåer.

2.2.5 Konklusjon

Derfor kan man konkludere med at den lineære modellen har begrensninger i denne sammenhengen, og at andregrads- og tredjegradsmodellen er mer presis for å beskrive forholdet mellom laktat og watt.



Figur 2.1: Relasjon mellom laktatverdier og watt. Gule punkter = laktat/watt, blå linje = lineær regresjonsmodell, grønn linje = andregrads-modell, rosa = tredjegrads-modell, rød linje = 2 mmol/L, lilla linje = 4 mmol/L, stiplet svarte linjer = skjæringspunkter mot rød og lilla linje

2.3 Estimering av størrelse på DNA-fragmenter

2.3.1 Introduksjon

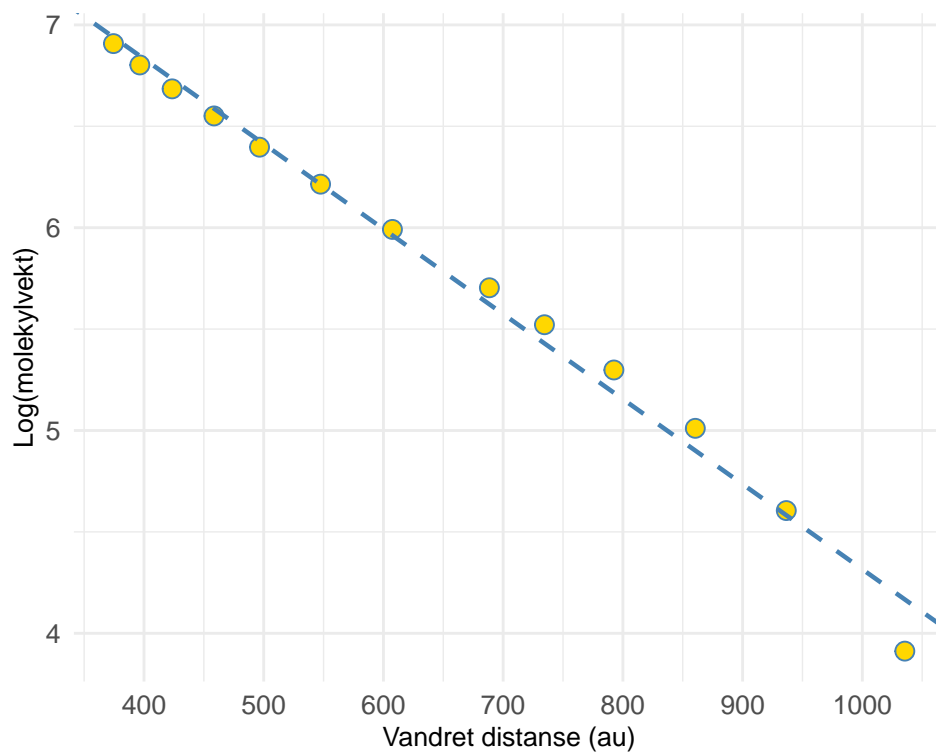
Beregning av størrelse på DNA-fragmenter er en viktig prosess i lab for å kunne identifisere og analysere DNA-prøver. Gelelektroforese er en av metodene som er hyppigst anvendt for dette formålet. DNA-prøvene tilsettes da en agarosegel og separeres etter størrelse ved hjelp av elektrisitet. Store fragment beveger seg langsommere enn små fragment og vil få en spredning på gelen. Ved å benytte en kjent referansemarkører, kan man estimere størrelsen på DNA-fragmentene. (Schadock *et al.*, 2015)

2.3.2 Metode

Først ble det tatt helblod fra en forsøksperson, og DNA ble ekstrahert ved hjelp av en standard protokoll for blod-DNA-isolering. Ekstraheringen involverte flere trinn, blant dem tilsetning av bufferløsninger og enzymatiske behandlinger for å bryte ned cellemembraner og frigjøre DNA-et. Det ble videre kjørt en elektroforese for å kunne separere DNA-fragmentene og bestemme lengden på DNA-fragmentene i basepar. Når elektroforesen var ferdig, ble det tatt et bilde av agarosegel 2%.

Videre har bildet fra elektroforesen blitt analysert ved hjelp av ImageJ/Fiji og videre data-analyser ble gjort i R. Logaritmen av molekylvekten ble plottet mot vandret distanse på gel for å estimere størrelse av DNA-fragment med høyere presisjon.

2.3.3 Resultat



Figur 2.2

Figure 2.2 viser sammenhengen mellom logaritmen av molekylvekten og distanse vandret på gel. Predikert molekylvekt er oppgitt i basepar (bp).

Tabell 2.1: Prediktert molekylvekt (bp) for gitte distanser (au)

Ukjente distanser (au)	Prediktert molekylvekt (bp)
18	4,596
384	994
408	895
436	796
470	690
508	588
560	475
600	400
618	371
696	267
742	221
798	174
862	133
936	98
994	77
1,208	31

Av Table 2.1 ses estimerte verdier for molekylvekten (bp) etter gitt distanse de har vandret nedover gelen. Testresultatene indikerer at for hver enhet man vandrer nedover gelen vil DNA-fragmentets vekt minke med -0.00419. Med modellen hvor molekylvekten predikerers , estimeres en minkning i molekylvekt på -0.00126.

2.3.4 Diskusjon

I oppgaven er det benyttet en kalibreringskurve for å estimere størrelsen av ukjente DNA-fragmenter. Det ble først benyttet en kjent DNA-stige for å etablere en modell mellom logaritmen av molekylvekten og distansen vandret på gel. Modellen ble så benyttet til å estimere de størrelsen på de ukjente DNA-fragmentene. Resultatene viser at mindre DNA-fragmenter vandrer lenger distanse på gelen, sammenliknet med større fragmenter, altså slik som forventet.

Prediksjonenes presisjon avhenger av flere faktorer. Blant dem kvaliteten på agarosegelen og målenøyaktighet av vandret distanse. Gelen ble laget manuelt, noe som kan føre til ujevnheter i konsistens og skape ulike forhold for molekylene som vandrer i gelen.

Selv om bildeprogrammet Fiji er et mye brukt verktøy for bildeanalyse, kan det være flere faktorer som kan påvirke nøyaktigheten av målingene. En mulig feilkilde kan være relatert til oppløsningen på bildene som ble tatt av agarosegelen. Redusert bildeoppløsning kan ha ført til unøyaktige målinger av distansen som DNA-fragmentene vandret under elektroforesen.

For å minimere de nevnte feilkildene, bør man vurdere å standardisere både gel-forberedelse og bildetakning ved fremtidige forsøk.

2.3.5 Konklusjon

2.4 Tolke en regresjonstabell

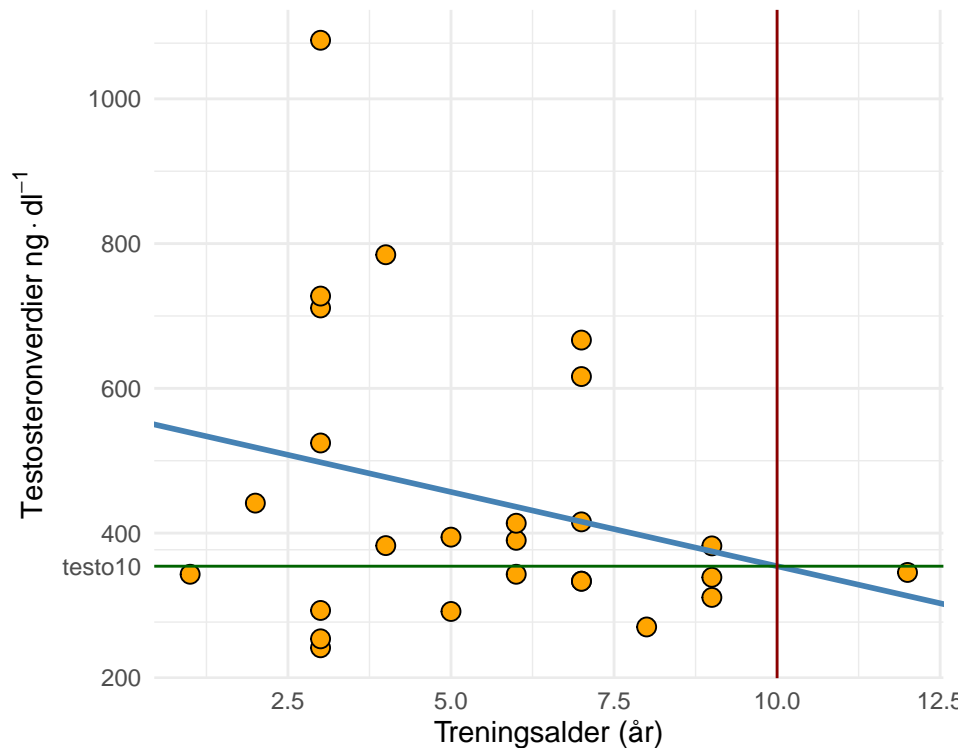
2.4.1 Introduksjon

Haun et al. (2019) ønsket å undersøke hvorfor mennesker reponderer forskjellig på styrketrening. Ulike biomarkører ble studert for å forsøke å finne ut av hva som er årsak til den ulike responsen mellom personer som responderer godt på styrketrening, kontra de som ikke responderer så godt (Haun *et al.*, 2019).

2.4.2 Metode

Denne studien undersøker biomarkører for muskelhypertrofi for lav (LOW) og høy (HIGH) respondere på motstandstrening. 30 mannlige universitetsstudenter gjennomførte et 6 ukers høyvolum treningsprogram, og ble delt inn i LOW og HIGH grupper basert på endringer i muskelstørrelse og sammensetning. Det ble tatt blodprøver der testosteronenivå i blodet var en av biomarkørene av interesse.

2.4.3 Resultat



Figur 2.3: Figur 3: Sammenheng mellom treningsalder og testosteronverdier i blodet

Den lineære modellen forteller i dette tilfellet at for hvert år man trener, så vil nivå av testosterone synke med 20.51 ng*dl-1.

2.4.4 Diskusjon

Fra datasettet hypertrophy valgte vi å se på sammenhengen mellom testosteronkonsentrasjon i blodet ($\text{ng} \times \text{dl}^{-1}$) og treningsalder (antall år med trening). Den lineære modellen forteller at testosteronkonsentrasjonen i blodet synker med 20.51 $\text{ng} \times \text{dl}^{-1}$ for hvert treningsår.

Analysen av dataene viser en t -verdi på -1.39 og en p -verdi lik 0.178, noe som indikerer at det ikke er statistisk signifikant bevis for en sammenheng mellom treningsalder og nivået av testosteron i blodet. Den negative t -verdien indikerer at testosteronnivå synker i tråd med økende treningsalder. Det kan diskuteres hvorvidt den negative nedgangen i testosteronverdier faktisk skyldes treningsalder, eller en annen konfunder som ikke er tatt hensyn til i dette tilfellet, nemlig økende biologisk alder.

Siden p -verdien er høyere enn det tradisjonelle signifikansnivået på 0.05, kan vi ikke avvise nullhypotesen, som antyder at det ikke er noen betydelig effekt eller sammenheng mellom de to variablene i dette datasettet. Dette betyr at variasjonen i testosteronnivåer ikke ser ut til å være relatert til hvor lenge individene har trent.

2.4.5 Konklusjon

3 Statistisk inferens

3.1 Introduksjon

Ifølge Cremers et al. (2017) er statistisk inferens en type statistikk som omhandler metoder for å trekke konklusjoner om en populasjon basert på et utvalg. Målet er å generalisere funnene fra utvalget og bruke det til å si noe om hele populasjonen (Cremers *et al.*, 2017).

En viktig del av statistisk inferens er å vurdere usikkerheten i estimatene, typisk ved hjelp av konfidensintervaller og p-verdier (Cremers *et al.*, 2017). Videre skriver Cremers et al. (2017) at effektstørrelser gir en kvantitativ vurdering av hvor betydelig en observert effekt er, og at effekten ikke er avhengig av utvalgsstørrelse. Dette er nyttig i forskning, da det hjelper med å forstå de praktiske betydningene av funnene man har gjort gjennom sammenlikning av grupper med forskjellige utvalgsstørrelser. Ved å kombinere statistisk inferens med effektstørrelser kan man bedre vurdere resultatene sine og sette det i en bred kontekst (Cremers *et al.*, 2017).

I første del av oppgaven defineres statistiske begrep og uttrykk. Deretter simuleres to utvalg på henholdsvis 8 stk og 40 stk. Det er ønskelig å undersøke hvordan utvalgsstørrelse påvirker estimater og resultater av statistiske analyser. Det blir simulert 1000 studier med de to utvalgene og gjort statistiske bergeneringer basert på dette. Det er vedlagt kode for gjennomføring av simuleringer bakerst i kapittelet.

3.2 *T*-verdi, *P*-verdi, Standard Error (SE) og Estimat

T-verdien er et mål som brukes i hypotesetesting for å vurdere om det er en signifikant forskjell mellom gjennomsnittet av et utvalg og et kjent eller hypotetisk populasjonsgjennomsnitt. *T*-verdiens størrelse er en indikasjon på hvor stor forskjell det er mellom utvalget og populasjonen. Jo større *t*-verdi, jo større forskjell mellom gruppene og desto mer signifikant forskjell.

$t = \frac{(\bar{x} - \mu_0)}{\frac{s}{\sqrt{n}}}$, der \bar{x} = gjennomsnittet av utvalget, μ_0 = gjennomsnittet av populasjonen, s = standardavviket, n = størrelsen av utvalget

P-verdien sier noe om usikkerheten i trekningen av utvalget. Kanskje trakk man i utvalg 1 bare ekstreme verdier, eller kun middelveidier. Utvalget vil neppe bli en perfekt representasjon av populasjonen uansett hvordan man trekker ett utvalg. Denne usikkerheten man må ta høyde for i studier og dette kan *p*-verdien hjelpe oss å si noe om. Dersom en tar høyde for at nullhypotesen (H_0) stemmer, hvor mange verdier kan en forvente er mer ekstreme enn våre

observasjoner. P -veriden måler altså observasjonene våre opp mot den spesifikke nullhypotesen. Lave p -verdier indikerer at det er liten sannsynlighet for at den observerte forskjellen skyldes tilfeldigheter, og gir støtte for å forkaste nullhypotesen. Høye p -verdier tyder derimot på at den observerte forskjellen kan være et resultat av tilfeldigheter, og gir dermed ikke tilstrekkelig grunnlag for å forkaste nullhypotesen og konkludere med en reell forskjell mellom gruppene.

Standardfeil (SE) er en måte å angi feilmarginen av et estimat eller måling. Sagt på en annen måte, beskriver SE variasjonen ved repeterte gjennomsnitt for et utvalg knyttet til en populasjon. Man benytter ofte sentralgrenseteoremet i denne sammenhengen, som sier at hvis en gjør en undersøkelse mange ganger, vil resultatene samle seg omkring den sanne verdien. SE kan beregnes manuelt ved å dele standardavviket på kvadratroten av antall observasjoner. Dette gir et mål på hvor nær det sanne gjennomsnittet vi kan forvente å komme. I dette tilfellet ser vi at SE for utvalg 1 er betydelig større enn for utvalg 2, noe som skyldes at utvalg 1 er mindre. En kan altså forvente å få verdier lenger unna den sanne gjennomsnittsverdien ved et mindre utvalg. Jo større utvalg, jo mindre feilmargin.

Estimatet er en verdi som blir kalkulert basert på et utvalg og blir brukt til å estimere noe om en populasjon. Det kan eksempelvis være et estimat på det sanne gjennomsnittet i en befolkning.

3.3 To-sidet test (two-tailed test)

En two-tailed test tillater at man ser på mulighet for endring i begge retninger, altså positiv og negativ retning. Når man undersøker ulike behandlinger, for eksempel for blodtrykk, vil man være interessert i å vurdere både positive og negative endringer mellom gruppene. En ser etter et avvik i begge retninger av fordelingen og ser om avviket er stort nok til å kunne forkaste nullhypotesen. Høye t -verdier referer ofte til lave p -verdier, som vil si at det er lite sannsynlig at den observerte forskjellen skyldes tilfeldigheter, gitt at nullhypotesen er sann. Ettersom en ser på forskjeller i begge retninger av fordelingen, må en doble p -verdien ved bruk av two-tailed test.

3.4 Standardavvik og standardfeilen (SE)

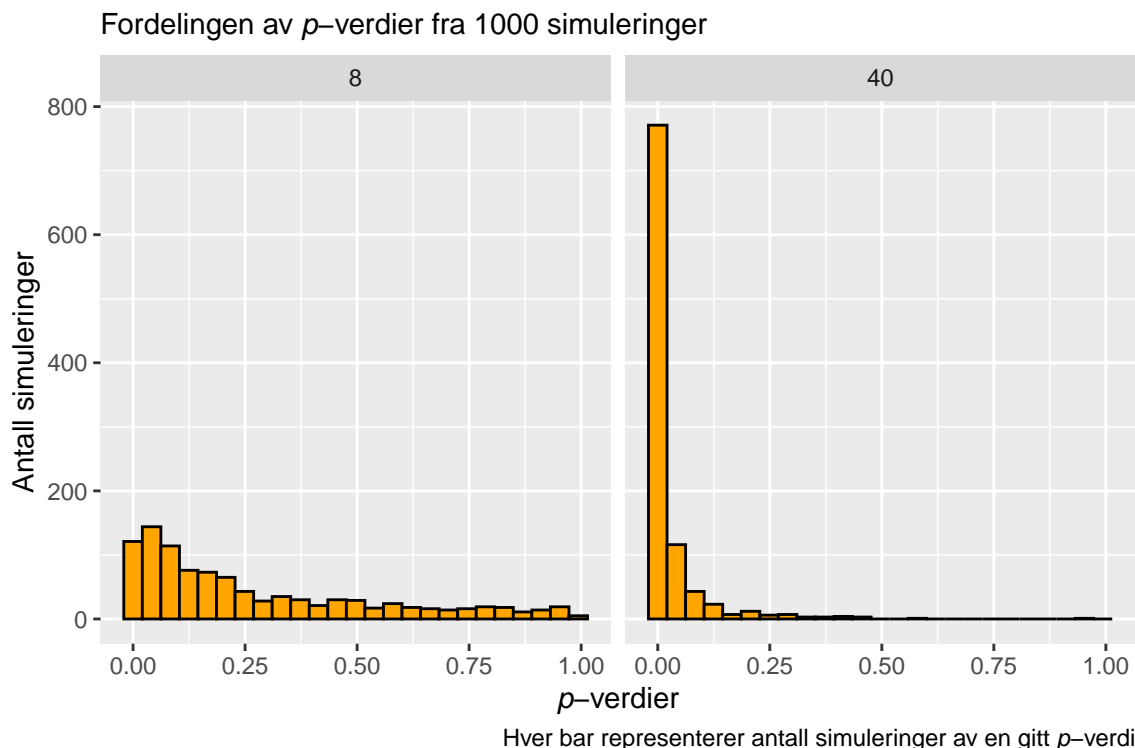
Hva er sammenhengden mellom standardavvik og standardfeil? For å illustrere sammenhengen beregnes standardavviket for estimat-variabelen og gjennomsnittet av SE-variabelen

Standard error: $SE = \frac{\sigma}{\sqrt{n}}$, der σ = standardavviket for utvalget og n = antall i utvalget.

For utvalget med $n = 8$ er standardavviket for estimatet 1.071 og gjennomsnittet av standardfeilen 1.021, altså tilnærmet likt. Det samme gjelder for utvalget med $n = 40$ der standardavviket for snittet av utvalget er 0.484 og gjennomsnittet av standardfeilen er 0.47. Nesten tilnærmet likt som i utvalg 1 ($n = 8$), bare med mindre avvik. Hvordan kan man så definere

standardfeilen, sett i lys av disse variablene? På generell basis kan man si at SE for et utvalg er et estimat på standardavviket til den teoretiske fordelingen av gjennomsnitt. Observerer SE synker når “n” (utvalget) øker.

3.5 Effekten av utvalgsstørrelse for statistisk styrke

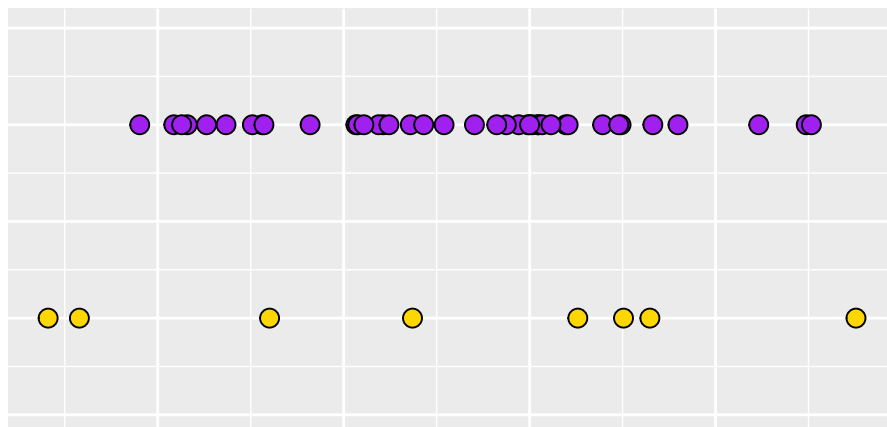


Figur 3.1: Fordelingen av p -verdier fra 1000 simuleringer

Som en nå har observert, ligger forskjellen mellom utvalg 1 og 2 altså først og fremst i størrelsen på utvalgene. Med et større utvalg ser en typisk lavere standardfeil (SE) og p -verdier, samt høyere t -verdier. I tillegg minimerer det risikoen for å gjøre type II-feil, altså at man forkaster sin hypotese når nullhypotesen faktisk er usann. Dette skyldes at større utvalg gir mindre variabilitet og dermed lavere usikkerhet sammenliknet med mindre utvalg. Utvalgsstørrelse påvirker dog ikke risikoen for å gjøre type I-feil, da dette bestemmes av signifikansnivået som settes (typisk $\alpha < 0.05$).

I Figure 3.1 observeres en jevnere fordeling av simulerte p -verdier i utvalget $n = 8$, sammenliknet med $n = 40$. Dette er et resultat av at den tilfeldige variasjonen blir større i små utvalg (Figure 3.2). Større utvalg er mer sensitive for små effekter. Derfor får man ofte lavere p -verdier med større utvalg og det blir enklere å forkaste nullhypotesen.

Illustrasjon av spredning innad i utvalgene



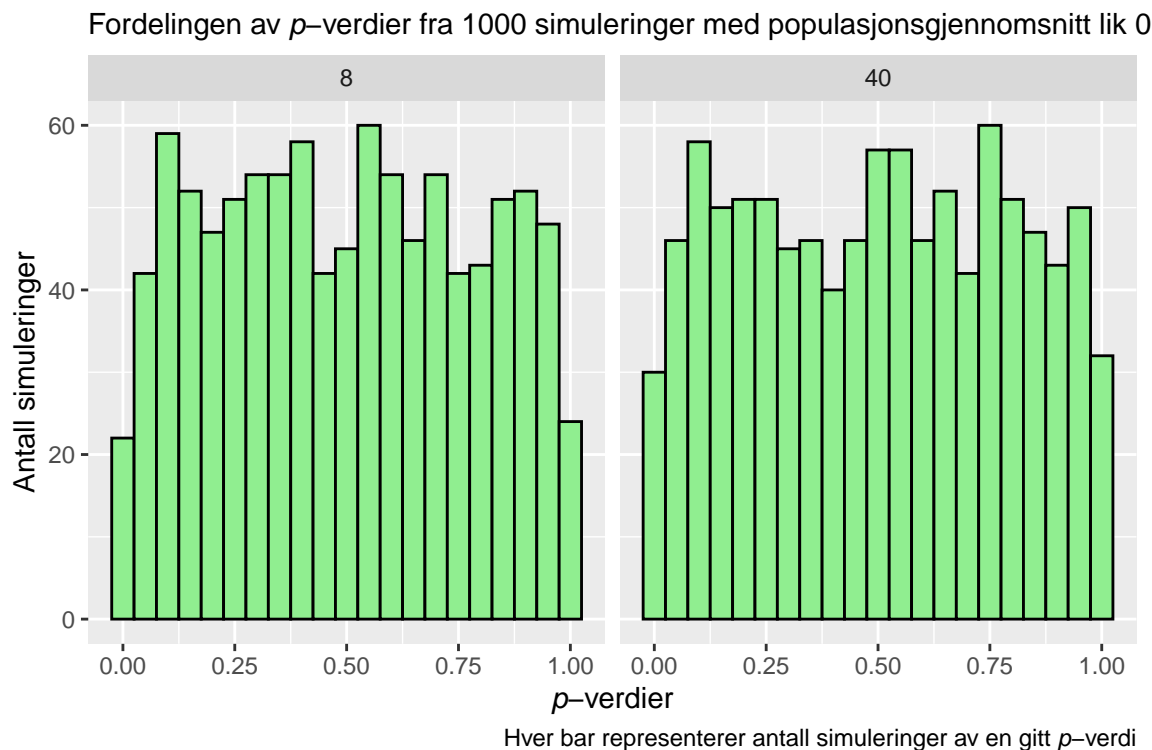
Utvalg 1 (gule punkter) og utvalg 2 (lilla punkter) fra populasjonen

Figur 3.2: Spredning i verdier for $n = 8$ og $n = 40$

Ved å plote p -verdiene for 1000 simuleringer fremheves også tendensen til falsk-positive (type I-feil) simuleringer ved $p = 0.05$. En falsk positiv oppstår når en test indikerer en signifikant effekt, slik at man forkaster nullhypotesen (H_0), når det egentlig ikke er en effekt der (H_0 er sann). Jo større utvalg man har, desto større risiko er det for å gjøre type I-feil. Når en setter signifikansnivået til $p < 0.05$, kan en observere antallet signifikante resultater i hvert utvalg. For utvalget med 40 observasjoner finner en 865 signifikante simuleringer, mens for utvalget med 8 observasjoner finner en 227 signifikante simuleringer. Her observeres langt flere signifikante resultater for utvalg med $n = 40$, som er i tråd med det vi forventet. Sannsynligvis vil flere av de 865 testene være type I-feil. På en annen side gir også større utvalg større statistisk styrke.

Statistisk styrke (power) refererer til sannsynligheten for å korrekt avvise nullhypotesen, og ved å sammenligne antall observasjoner under signifikansnivået vil en få en indikasjon på styrken i testene. Den statistiske styrken for utvalg $n = 8$ blir beregnet til 23 %, mens for $n = 40$ blir den beregnet til 87 %. Effektstørrelsen er i dette tilfellet satt til 0.5. Ifølge Cohens'd er den forventede effekten liten ved 0.2, moderat ved 0.5 og stor ved 0.8 (Lakens, 2013).

Den beregnede statistiske styrken for $n = 8$ på 23 % indikerer lav sannsynlighet for å oppdage en reell effekt, dersom den er tilstede. Det vil altså være større sjans for å gjøre en type II-feil med det lille utvalget. For $n = 40$ var styrken 87 %, som betyr at sjansen for å oppdage en reell effekt er stor dersom den eksisterer. Dette basert på den lave usikkerheten i estimatene som følger med et større utvalg.



Figur 3.3: Antall simuleringer av en gitt p -verdi med et befolkningsgjennomsnitt på 0

Gjør så nye 1000 simuleringer med samme utvalgsstørrelser, men med et gjennomsnitt i befolkningen på 0. I dette tilfellet vil altså nullhypotesen være sann. Ved den nye simuleringen vil en komme over 41 falsk-positive tester for utvalget med 8, og 61 for utvalget med 40. Sammenlikner man Figure 3.1 og Figure 3.3, ser en at fordelingen av p -verdier er jevnere ved et gjennomsnitt i befolkningen på 0. Dette kommer av at de estimerte gjennomsnittsverdiene for begge utvalg vil ligge nært 0 og man vil sjeldent få verdier som motbeviser nullhypotesen.

I denne simuleringen finner en flere falsk positive tester med det store utvalget, sammenliknet med den lille utvalget. Dette er kanskje motsatt av hva en kan forvente og det kan være flere grunner til det. En av grunnene kan være at et større utvalg gjør tester mer sensitive for selv små effekter, altså at man finner flere tilfeldige signifikante resultater. Det blir altså lettere å få p -verdier under signifikansnivå og dermed kan det oppstå flere type I-feil. Dette gjelder også der gjennomsnittet er forskjellig fra 0.

3.6 Koder for simuleringer

```

set.seed(1)

#Første simulering med gjennomsnitt 1.5 og sd 3
population <- rnorm(1000000, mean = 1.5, sd = 3)

#lager to ulike utvalg der utvalg 1 er n=8 og utvalg er 2 n=40
samp1 <- data.frame(y = sample(population, 8, replace = FALSE))
samp2 <- data.frame(y = sample(population, 40, replace = FALSE))

# Andre simulering med gjennomsnitt 0 og sd 3. Hva skjer når nullhypotesen er sann.
population <- rnorm(1000000, mean = 0, sd = 3)

samp1 <- data.frame(y = sample(population, 8, replace = FALSE))
samp2 <- data.frame(y = sample(population, 40, replace = FALSE))

#Det gjøres 1000 simuleringer ved både første og andre simulering
#Har med denne chunken primært for å illustrere forskjellen i gjennomsnitt ved de to ulike s
#NB: Koden i chunken er ufullstendig da det mangler funksjon for simulering.

```


4 Studiedesign

4.1 Introduksjon

Studiedesign er en viktig komponent i forskning da det vil være avgjørende for hvordan data samles inn, tolkes og analyseres. Ifølge Zabor et al. (2020) vil et godt studiedesign ofte resultere i mer pålitelige resultater og større grad av reproduserbarhet. Studier som ser seg mer anvendbare i praksis er å foretrekke (Zabor *et al.*, 2020).

Struktur er en viktig komponent når man planlegger en studie. QALMRI-metoden kan bidra til å sikre en ryddig og god struktur. QALMRI adresserer spørsmål som en ønsker svar på, for så å oppgi hvilke tilnærminger som skal benyttes for å besvare spørsmålet. Videre ser man på sammenhengen mellom metodene som benyttes i studien og koblingen til spørsmålet, eller hypotesten, man har stilt i oppgaven. Til slutt presenteres resultatene og man drar konklusjoner til den aktuelle populasjonen.

I denne oppgaven studeres fem originalartikler som tar for seg effekten av blokkperiodisering på $\dot{V}O_{2max}$ (Breil *et al.*, 2010; Rønnestad *et al.*, 2014b, 2014a, 2016; Rønnestad *et al.*, 2019). I oppgaven blir det lagt mest vekt på studiedesignene og de statistiske analysemetodene fra de utvalgte studiene, samt en vurdering av styrker og svakheter knyttet til disse.

4.2 Metode

Alle studiene i oppgaven er randomiserte kontrollerte forsøk (RCT) der man var interessert i å undersøke effekten av periodisering av høyintensiv trening. Utvalget ble delt i to grupper, en blokkperiodiseringsgruppe (BP) og en kontrollgruppe (CT). I studiene trente utvalget i BP og CT det samme volumet totalt over intervensjonsperioden. Forskjellen var at BP periodiserte sine høyintensive økter i bolker/mikrosykluser (periodiserignen varierte mellom studiene). Det ble gjennomført målinger både pre og post treningsintervensjonen og intervensjonene hadde varighet fra 4 til 12 uker.

I studiene lå utvalget et sted mellom 16 og 22 personer. I noen studier så en på både kvinner og menn, i andre studier kun menn. Gjengangeren var at utvalget ble definert som godt trente utøvere, eller at de hadde drevet kompetetivt i sin idrett de siste årene.

Det ble gjennomført ulike statistiske analyser i de ulike forsøkene. Noen benyttet ANOVA, mens andre benyttet *t*-tester, eller en kombinasjon av begge. Flere av studiene benyttet

Cohen's d for å beregne effektstørrelse (ES) da dette er en gunstig metode når man vil vurdere effekten av en intervensjon og sammenlikne to grupper (Hopkins *et al.*, 2009). Med Cohen's d får man standardisert forskjellen mellom gruppene og man kan sammenlikne effekter på tvers av studier hvor det benyttes ulikt design eller ulike måleenheter. Med bruk av Cohen's d er det viktig at utvalget er representativt for populasjonen man ønsker å si noe om.

I studiene ble resultater presentert som gjennomsnittsökning i % \pm standardavvik. Resultater ble ansett som statistisk signifikante dersom p -verdier ≤ 0.05 .

4.3 Resultater

Etter å ha sammenliknet kontrollgruppen med blokkperiodiseringsgruppen post intervensjon, kom samtlige studier frem til at blokkperiodisering gav signifikant økning i prestasjonsparametere. Endringene kan ses i Table 4.1.

P -verdier < 0.05 i samtlige studier. Den gjennomsnittlige effekten ble målt til å ligge en plass mellom 0.76 og 1.34 i 4/5 studier. En av studiene gjorde i tillegg beregninger på kun de mannlige utøverne i sin studie og fant mer signifikante resultater. Denne studien presenterte ingen effektstørrelser.

Tabell 4.1: Endringer i BP-gruppen

Studie	Endring VO2max	p -verdi	Effektstørrelse
1	$5.1 \pm 3.3\%$	$=0.05$	1.04
2	$5.1 \pm \text{na}$	<0.01	NA
3	$2.0 \pm 2.5\%$	<0.05	0.76
4	$4.6 \pm 3.7\%$	<0.05	1.34
5	$8.8 \pm 5.9\%$	<0.05	1.08

Endring VO2max er gjennomsnitt og standardavvik

4.4 Diskusjon

Ifølge Zabor et al. (2020) er styrken ved å benytte RCT som studiedesign at det minimerer bias og skaper godt sammenlikningsgrunnlag blant liknende intervensjoner. Bias innen forskning referer til faktorer ved studien som trekker resultat i en gitt retning. Ved å benytte RCT blir resultatene mer pålitelige. Ved å benytte objektive data i sine resultater skapes en anvendbarhet i form av at studiene i større grad er sammenliknbare på tvers av studier og intervensjoner (Zabor *et al.*, 2020). KILDE TIL BIAS

I lys av avsnittet over kan en diskutere begrepet konfundring, som refererer til en tredje variabel som påvirker både den avhengige og uavhengige variabelen i studien. I denne studien kan vår avhengige variabel være prestasjonsparametere og den uavhengige variabelen være treningsintervensjonen. Den tredje variabelen (konfunderen) kan i dette tilfellet være søvnkvalitet. Dersom man ikke tar høyde for den tredje variabelen, kan man trekke feilaktige konklusjoner om effekten av treningsintervensjonen. KILDE TIL KONFUNDER

En svakhet ved studiene er de små utvalgene. Dette øker sjansen for å gjøre type II-feil (godtakelse av nullhypotesen, tross at den er falsk). Gjennomsnitt i små utvalg påvirkes i relativt stor grad av verdier som ligger langt unna gjennomsnittet (outliers) (Mishra *et al.*, 2019). For å objektivt fjerne outliers fra datasettet kan en bestemme seg for grenseverdier man ønsker at sine verdier skal ligge innenfor (Hopkins *et al.*, 2009).

Det kan tenkes at det er gunstig for fremtidige studier å ha større grad av mangfold, dvs. menn og kvinner, ulike alder og treningsbakgrunn. Dette for å kunne si noe om en bredere populasjon. Utvalget bør dog stå i stil til populasjonen man ønsker å si noe om. Ønsker man eksempelvis å forskrive gode treningsmetoder for toppidrettsutøvere innen sykkelporten, bør utvalget bestå av godt trente syklister. Da utvalgene i disse studiene bestod av godt trente utøvere, kan det antyde at blokkperiodisering av høyintensiv trening kan være gunstig for utøvere som allerede er på et høyt nivå.

Tid og varighet av intervensjoner er ofte en svakhet i studier. Disse studiene varte i hhv. 5-, 4-, 5-, 6- og 12-uker, noe som gjør det vanskelig å si noe om de langsiktige effektene av blokkperiodisering vs tradisjonell trening. Funnene i studiene kan styrkes dersom man studiene repeteres og/eller følges opp over lengre perioder.

Målet med studiene var å se hvorvidt mikrosykluser med høyintensiv trening øker prestasjonsparameteren $\dot{V}O_{2max}$ mer enn tradisjonell periodisering av trening. For å undersøke potensielle endringer mellom BP og CT i studiene ble det som nevnt benyttet ulike statistiske tester. Grunnet små utvalg og en forventet liten endring i prestasjonsparametere blant de godt trente utøverne, ble det benyttet t-tester da det viser seg hensiktsmessig ved små utvalg og benyttes for å sammenlikne gjennomsnittet mellom to grupper (Mishra *et al.*, 2019).

Det at én av studiene ikke presenterte effektstørrelse kan man stille spørsmål ved. En av grunnene kan være at små utvalg kan skape store variasjoner i resultater og skape lite pålitelige effekter. Et annet punkt å diskutere er hvorvidt resultatene er av klinisk relevans. Dersom forskerne mener at en effektsørrelsen kan gi et feilaktig inntrykk av resultatenes styrke, kan de la være å presentere effekten.

4.5 Konklusjon

Blokkperiodisering viser å gi større økning i $\dot{V}O_{2max}$ i samtlige studier, sammenliknet med tradisjonell trening. Langrennsutøvere, ishockeyspillere, alpinister og syklister har alle hatt økning i $\dot{V}O_{2max}$ ved å periodisere hard trening. Godt trente utøvere, som potensielt har

stagnert med tradisjonell trening, kan ha god effekt av blokkperiodisering. Funnene vil særlig være relevant for godt trente utøvere, da dette var utvalget som ble sett på i samtlige studier. Dette er dog et snevert utvalg, som trolig har egenskaper som skiller seg fra resten av befolkningen. Derfor vil det være utfordrende å generalisere funn og anbefalinger opp mot den generelle befolkningen. Repetisjon av studiene der utvalget innehar større grad av spredning når det kommer til treningsbakgrunn, kjønn og alder kan derav være av interesse.

Alle studiene er randomiserte kontrollerte studier, noe som gir tyngde til grad av validitet og grunnlag for direkte sammenlikning av ulike treningsmetoder. Det parallelle gruppedesignet som ble benyttet isolerer effekten av de ulike treningsintervensjonene.

ANOVA (analysis of variance) og t -tester ble benyttet slik at man kan sammenlikne flere grupper over ulike tidpunkter i intervensjonen. På den måten kan man få innblikk i hvordan ulik form for trening påvirker prestasjonparametere over tid. Valget av testene bidrar å styrke funnene om at blokkperiodisering gir økning i $\dot{V}O_{2max}$, sammenliknet med tradisjonell trening.

5 Repeterte målinger

5.1 Introduksjon

I dagens samfunn er det i stor grad enighet om at styrketrening er viktig for god helse. Befolkningen trender mot en mindre aktiv livsstil, samtidig som man blir eldre og eldre. Etter fylte 50 år, viser studier at man taper et sted mellom 5 og 10 % muskelmasse årlig, som tilsvarer ~ 0.4 kg (Westcott, 2012). Tilstander som aldersrelatert muskeltap, diabetes type II, hjertesykdom, redusert beinmasse og økende grad av fettmasse ses hyppigere og hyppigere. Ved sykdom ser en høy grad av inaktivitet som igjen leder til tap av muskelmasse, redusert metabolsk respons, økt fettmasse og andre liknende tilstander. Aerob aktivitet er for tiden veldig populært blir i stor grad anbefalt for å motvirke utvikling av disse tilstandene, men hva gjelder så styrketrening? (Westcott, 2012)

Styrketrening kan defineres som trening med mål om å bli sterkere, forebygge skader, fremme prestasjon eller trening med mål om å øke muskelmasse (Stone *et al.*, 2007). Noen definerer volum som den viktigste faktoren for å se endringen på cellenivå, mens andre mener at høyintensiv trening kan gi samme effekten (Santana *et al.*, 2021). Radaelli *et al.* (2024) ser at lavintensiv trening også kan ha god effekt for å øke muskelmasse og styrke i underekstremiteter blant eldre individer (≥ 60 år) (Radaelli *et al.*, 2024). Antall sett, antall repetisjoner og motstanden det trenes på benyttes for definere *volumet* av trening, mens *intensiteten* på treningen ofte defineres som en prosentandel av én maksimal repetisjon (1RM). For mange er tidsklemma en problem i hverdagen og flere er trolig interessert i å vite hvor lite trening som skal til for å oppnå ønsket stimuli av treningen man gjennomfører. (Santana *et al.*, 2021)

I denne studien er det ønskelig å se nærmere på hvorvidt volumet av styrketrening er av betydning for økning av muskelmasse og muskelstyrke.

5.2 Metode

5.2.1 Utvalg og studieoversikt

I studien ble det rekruttert 41 individer, menn ($n = 19$) og kvinner ($n = 22$). For å bli inkludert i studien måtte man kunne få lokal anestesi, ikke ha trent mer enn én styrkeøkt i uken siste året, ikke ha nedsatt eller påvirket muskelstyrke grunnet tidligere eller nåværende skade og ikke innta medisin som kan påvirke treningsadaptasjon. Etter gjennomført treningsintervensjon

ble 7 deltakere ekskludert fra dataanalysen da de ikke oppfylte kravene om gjennomføring av intervensjon. Table 5.1 viser en oversikt over antall kvinner og menn som er inkludert i analysering av data og deres respektive gjennomsnittlige karakteristikk. Deltakere som ble inkludert i analysen hadde alle tidligere erfaring med idrett.

Tabell 5.1: Deltakerkarakteristikk

Kjønn	Alder (År)	Høyde (cm)	Vekt (kg)
Kvinne n = 18	22 (1.3)	167.7 (6.9)	64.4 (10.4)
Mann n = 16	23.6 (4.1)	182.9 (5.9)	75.8 (10.7)

Alle verdier er gjennomsnitt og standardavvik.

5.2.2 Spesifikke beskrivelser av testmetode

Intervensjonen varte i 12 uker og bestod av helkroppss styrketrening. Beinøvelser ble gjennomført unilateralt. De ble tilfeldig trukket hvilket bein som skulle gjennomføre ett sett (single set) og hvilket som skulle gjøre tre sett (multiple sets).

5.2.3 Styrketreningsprotokoll

Treningsøktene startet med en standardisert oppvarming som besto av (i) ergometer-sykling i 5 minutter (RPE 12–14), (ii) øvelser med kroppsvekt med 10 repetisjoner av push-ups, sit-ups, ryggekestensjon og knebøy, og til slutt (iii) et sett med 10 repetisjoner på ~ 50 % av 1RM for hver styrkeøvelse i intervensjonen.

Beinøvelsene ble gjennomført i følgende rekkefølge: ettbeins beinpress, leg curl og kneekstensjon. Single settene ble gjennomført mellom andre og tredje sett for multiple settene. Treningsintensiteten økte gjennom intervensjonen. Det ble trent 10RM første to ukene, 8RM de neste tre ukene og 7RM de siste syv ukene. For å gjøre treningen mer gjennomførbar for deltakerne, ble noen av øktene gjennomført uten tilsyn. Deltakerne ble da bedt om å føre detaljert logg. Fra den niende treningsøkten, ble én av de tre ukentlige øktene gjennomført på 90 % av intensitet, mens tiltenkt volum skulle opprettholdes. Det skulle være minimum 48 timer mellom hver fullintensitetsøkt og minimum 24 timer mellom submaksimale økter (90 %). For å sikre restitusjon, ble hver deltaker tildelt en drikk som inneholdt $0.15 \frac{g}{kg}$ protein, $11.2 \frac{g}{kg}$ karbohydrat og $0.5 \frac{g}{kg}$ fett.

5.2.4 Muskulær Styrkevurdering

Muskelstyrke ble vurdert ved hjelp av en dynamometer for å måle isokinetisk og isometrisk styrke. Deltakerne ble kjent med testprosedyrer og fikk flere forsøk på hvert nivå for å sikre nøyaktighet. En repetisjon maksimum (1RM) ble testet for unilateral beinpress og kneekstension, der testene ble forberedt med spesifikke oppvarmingssett.

5.2.5 Måling av Kroppssammensetning og Mager Muskelmasse

Kroppssammensetning ble vurdert med dual-energy X-ray absorptiometri (DXA). Deltakerne ble instruert om å faste og unngå intensiv trening før målingene.

5.2.6 Dataanalyse og Statistikk

Utvalgsstørrelsesberegninger indikerte at 40 deltakere var tilstrekkelig for å oppdage meningsfulle forskjeller. Lineære blandede modeller (LMM) ble benyttet for å analysere endringer i muskelhypertrofi og styrke. Alle analyser ble utført i R, med signifikansnivå satt til $\alpha = 0,05$.

De statistiske analysene hadde som mål å avdekke effekten av forskjellige motstandstreningsvolumer på muskelstyrke og hypertrofi.

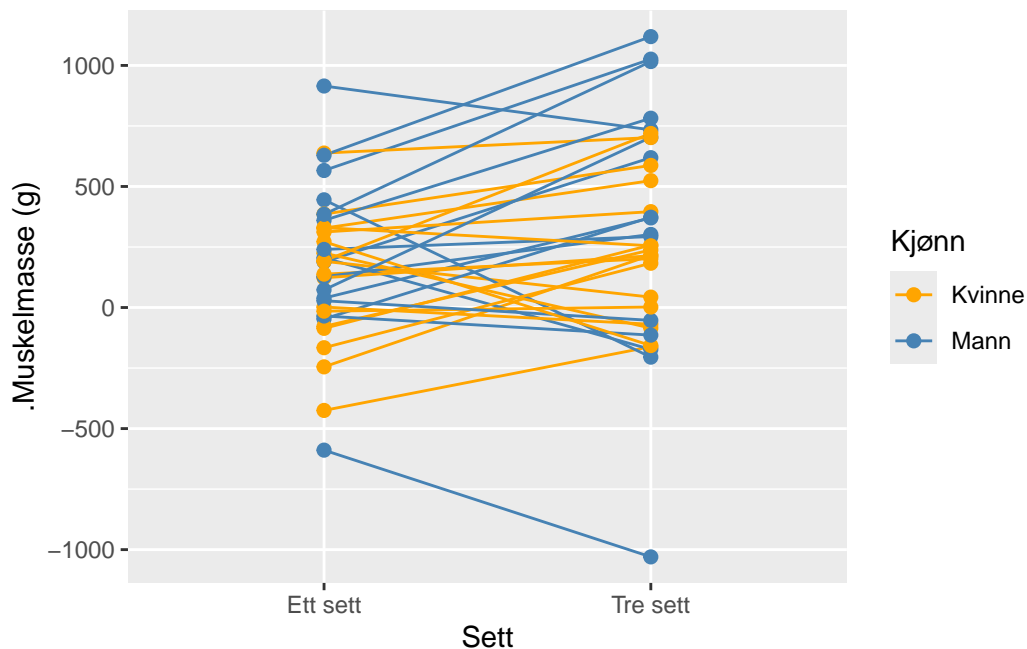
5.3 Resultater

5.3.1 Treningsvolums effekt på mager muskelmasse i underekstremitetene

Etter den 12 uker lange styrketreningsintervensjonen er den prosentvise økningen i mager muskelmasse for benet som trente tre sett 3.25% og for det som trente ett sett 1.9%. Den gjennomsnittlige forskjellen i mager muskelmasse i beina mellom de to treningsvolumene (ett sett vs. tre sett) var 122.8 g (95% CI: [8.6, 237.0], p -value = 0.036, $t_{33} = 2.19$).

Benytter man en blandet modell som tar man høyde for hver deltaker sine pre-verdier, og ser på hvordan antall sett påvirker mager muskelmasse etter treningsintervensjonen, blir den gjennomsnittlige endringen i mager muskelmasse mellom ett og tre sett 122.7 g (95% CI: [10.9, 234.5], p -value = 0.037, $t_{32.8} = 2.18$).

Figure 5.1 viser endringen i mager muskelmasse målt i gram for de ulike deltakerenes bein. Linjen som er trukket mellom punkt viser hver deltakers resultat og gir et inntrykk av forskjellen mellom effekten av treningsvolum på muskelmasse. Tre sett virker av figuren å ha større effekt for å øke muskelmasse, sammenliknet med lavere treningsvolum.



Figur 5.1: Endring i mager muskelmasse fra pre- til post-dexamåling (målt i gram)

5.3.2 Treningsvolums effekt på kraftutvikling i underekstremitetene

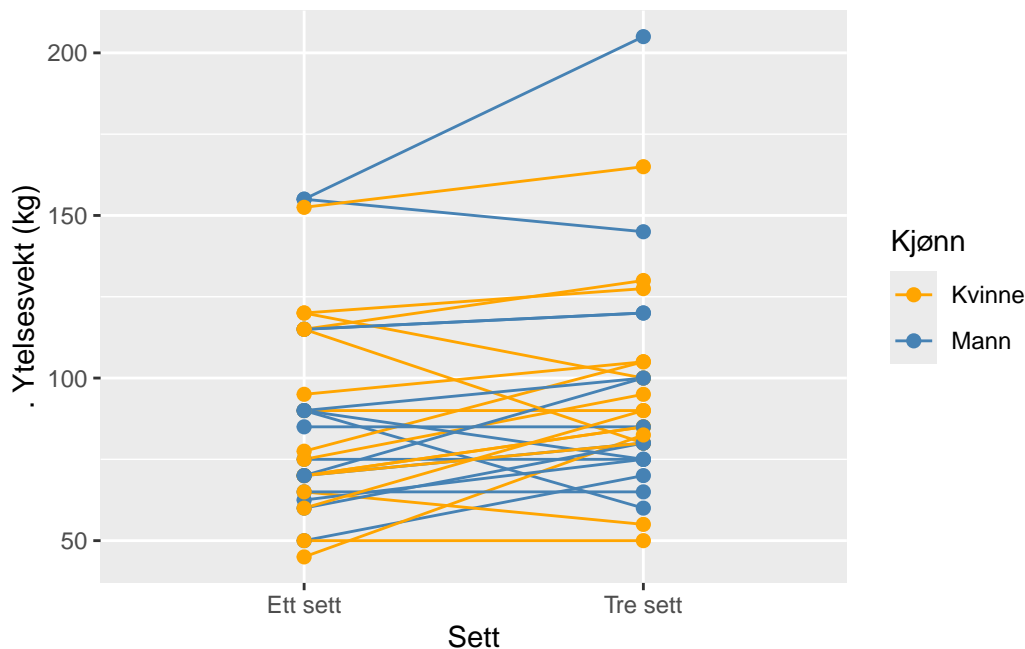
Den gjennomsnittlige prosentvise økningen i ytelsesvekt ved 1RM (målt i kg) for henholdsvis tre og ett sett var 36.89 og 34.15. Den gjennomsnittlige forskjellen i ytelsesvekt mellom beina ved de ulike treningsvolumene (ett sett vs. tre sett) var 7.8 kg (95% CI: [1.1, 14.6], p -value = 0.025, $t_{30} = 2.36$).

Benytter man en blandet modell der man tar høyde for hver deltaker sine pre-verdier, og ser hvordan antall sett påvirker ytelsesvekt etter treningsintervensjonen, blir den gjennomsnittlige endringen i ytelsesvekt mellom ett og tre sett 8.4 kg (95% CI: [1.9, 15], p -value = 0.017, $t_{30.5} = 2.53$).

Figure 5.2 viser endringen i den maksimale ytelsesvekten deltakerne evnet å presse ved 1RM for henholdsvis de som trente ett og tre sett. Av figuren ser man en tendens til at høyere treningsvolum gir større evne til å yte mer på øvelsen beinpress.

5.4 Diskusjon

I tråd med tidligere forskning, som har undersøkt effekten av treningsvolum på muskelmasse og styrke (Stone *et al.*, 2007), (Santana *et al.*, 2021), viser denne studien at høyere volum



Intensiteten (70–90 % 1RM) brukt i denne studien er i tråd med anbefalingene for både muskelstyrke og hypertrofi (Santana *et al.*, 2021). Imidlertid kunne det vært interessant å undersøke om lavere intensitet på treningen kombinert med høyere volum kunne ha hatt en like god effekt på muskelmasseøkning for denne gruppen.

En langtidsoppfølging av deltakerne etter at treningsintervensjonen er avsluttet, kan gi mer informasjon om hvordan treningsvolum påvirker vedlikehold av muskelmasse og styrke i lengden. Hvordan deltakerne vedlikeholder gevinster over tid kan være viktig for å vurdere praktisk anvendbarhet av treningsvolumet som en strategi for å bevare helsen.

Ved å benytte en blandet modell (ANCOVA) i tillegg til t-test får man en modell som tar hensyn til individuell variasjon. Modellen tar høyde for at hver deltaker har ulike pre-verdier og at de responderer ulikt gjennom intervensjonen. Ved å gi hver deltaker sin egen startverdi, kan man mer presist vurdere og isolere effekten av treningsintervensjonen (ett sett vs tre sett).

5.5 Konklusjon

Treningsvolum og intensitet virker å være viktige parametre for å øke muskelmasse og muskelstyrke. Tre sett med styrkeøvelser synes å gi en større tilpasning enn kun ett sett med øvelser. De langsiktige effektene av denne treningsmodellen bør undersøkes nærmere, i tillegg til om de samme effektene av treningsvolum kan observeres i andre deler av befolkningen.

6 Vitenskapsfilosofi

6.1 Falsifikasjonisme

Falsifikasjon, eller det motsatte av verifikasjon, er en teori som etterstreber å motbevise hypoteser og teorier (Popper, 1963). Tradisjonelt sett bygger man opp en hypotese for å gjøre en kvalifisert gjentning eller antakelse om noe man ønsker å finne ut av, enten det er å avkrefte eller bekrefte noe. Men kan man noen gang bekrefte en teori? En kjent forsker og filosof som mente at man aldri kunne bekrefte en teori, var Karl Popper. Popper (1963) mente at teorier aldri kan bevises med absolutt sikkerhet og at man heller bør legge vekt på å avsløre usannheter. En teori kan dog midlertidig aksepteres dersom den overlever flere forsøk på falsifisering, men det vil alltid være en mulighet for at den på en senere tidspunkt kan motbevises (Popper, 1963).

Poppers motivasjon for å utarbeide teorien om falsifisering hadde trolig sammenheng med at han ikke trodde induksjonsproblemet kunne løses. Et eksempel på induktivt resonnement er: alle svanene jeg har sett er hvite, derfor er alle svaner som finnes hvite. Popper så trolig nytten av å forkaste hypoteser som ikke var gode nok, slik at nye og bedre hypoteser kunne konstrueres. En skal dog se at Popper sine ideer byr på flere utfordringer når det kommer til dagens teknologi, data og forskning.

Innen falsifikasjonisme er altså begrepet testbarhet et viktig faktum. En teori eller hypotese bør være formulert på en slik at måte den enkelt kan testes og motbevises. Ifølge Popper kreves det kun én observasjon som ikke stemmer overens med hypotesen for at vi kan forkaste den (Popper, 1963). Et eksempel på et slikt scenario kan være følgende: *Vann koker ved 100 grader celsius*. Hvis man er 1000 m.o.h. er det kjent at vann koker ved lavere temperaturer enn ved havnivå. Ifølge Popper kan man da forkaste hypotesen om at vann koker ved 100 grader celsius, da det ikke nødvendigvis alltid gjør det.

Som nevnt over er testbarheten til en hypotese et viktig punkt innen falsifikasjonisme og også et av problemene som oppstår med denne ideen. La oss bruke eksempelet over med kokende vann og omformulere hypotesen: *Vann har en tendens til å koke under bestemte forhold*. Denne formuleringen er vag og åpen for tolkning, men dog muligens enkel å falsifisere, eller? Hvilke forhold er det snakk om og hva menes med tendens til å koke? Mindre konkrete hypoteser er generelt sett mer utfordrende å motbevise. Popper mente derfor at lite konkrete hypoteser er mindre vitenskapelige. Han mente derfor at en bør foretrekke spesifikke teorier som enkelt lar seg sette på prøve. Jo lettere det er å falsifisere en teori, destod mer vitenskapelig er den (Popper, 1963).

Ved å være enda mer spesifikke i formuleringen av teorier og hypoteser kan man motarbeide dette problemet med vaghet. Tydelige kriterier og konkrete utsagn i sin hypotese vil føre til at hypotesen er enklere å utfordre. Igjen kan en se på et eksempel med kokende vann som er helt i andre enden av skalaen sammenliknet med det vage eksempelet i avsnittet over. *Rent vann som ikke er forurensset vil koke ved 100 grader celsius ved normalt atmosfærisk trykk (760 mmHg) når det ikke er bevegelse i vannet og det foregår i et lukket miljø.* Her har en spesifisert teorien i stor grad, noe Popper var begeistret for. Består denne hypotesen ved gjentatt testing, vil den fremstå som robust, men som tidligere nevnt, kun midlertidig akseptert, ifølge Popper (Glass, 2010).

Et annet problem med Poppers falsifiseringsteori er at den ikke tar høyde for teorier som utvikles over tid. En teori som først virker falsifiserbar, kan senere vise seg å være delvis riktig og revideres eller utvides i lys av nye data. Innen medisin er det vanlig at hypoteser modnes og utvikles før de kan testes på en endelig måte. Poppers strenge krav til falsifisering kan derfor hemme vitenskapelig utvikling ved å avvise teorier i et tidlig stadium.

6.2 Bayesianisme

Det finnes ulike modeller for å beregne sannsynligheten for noe. Noen modeller baserer seg på empirisk data, som at solen stod opp i går, og dagen før der, og dagen før der igjen, osv. Derfor vil sannsynligvis solen stå opp i morgen også (induktivt resonnement). I nyere tid er det populært å forkaste hypoteser, på samme måte som Popper og falsifisering. Ifølge Joyce (2004) skiller Bayesianismen seg fra slike sluttregninger ved at sannsynligheten beregnes via matematiske formler. Bayesianere benytter data for å forklare hvorvidt en hypotese er mer eller mindre sannsynlig. Troen på en hypotese kan altså oppdateres med ny data. Ergo benytter Bayesianere en modell som tar høyde for ny og kontinuerlig data (Joyce, 2004).

Bayesianismen bygger på Bayes teorem som ser slik ut:

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)}$$

der $P(H|D)$ er den posteriore sannsynligheten, $P(D|H)$ er sannsynligheten for å observere dataen gitt hypotesen (likelihood), $P(D)$ er den totale sannsynligheten for dataen og $P(H)$ er troen på hypotesen før dataen er observert (prior sannsynlighet). Posterior sannsynlighet refererer til sannsynligheten for en gitt hypotese eller teori etter at man har tatt hensyn til informasjon eller data man har tilgjengelig. Enkelt forklart forteller posterior sannsynlighet hvor sann en hypotese er etter å ha vurdert ny og gammel informasjon på området.

Benytter videre et eksempel for å illustrere hvordan Bayes teorem kan tolkes. Spørsmålet en lurer på er hvor stor sannsynlighet det er for at en pasient har en gitt sykdom. Basert på forholdene settes sannsynligheten for at pasientene har sykdommen til 2 % ($P(H) = 0.02$). En test benyttes for å undersøke om pasienten har sykdommen. Gitt at pasienten har sykdommen, vil testen slå ut positivt 95 % av gangene ($P(D|H) = 0.95$). Motsatt vil testen slå ut negativt i 90 % av tilfellene når pasienten ikke er smittet av sykdommen ($P(\neg D|\neg H) = 0.9$).

Ytterligere beregninger gjøres med Bayes teorem og finner en sannsynlighet tilnærmet 16 % for at pasienten faktisk har sykdommen gitt positiv test.

Det er altså kun ≈ 16 % sannsynlighet for at pasienten er smittet når testen slår ut positivt. Grunnet til dette er den i utgangspunktet lave sannsynligheten for å være smittet av den gitte sykdommen. Prior sannsynlighet bestemmer altså i stor grad den posteriore sannsynligheten $P(H|D)$.

Bayesianer benytter altså matematiske formler der man bruker ulike sannsynligheter man beregner på forhånd basert på forholdene som ligger til grunn. På denne måten kan Bayesianere objektivt analysere og sannsynliggjøre hendelser. Eller er disse beregningene så objektive?

Et problem som ofte kommenteres ved Bayesianismen er den subjektive oppfatningen av prior sannsynligheten ($P(H)$) og hvilken verdi denne settes til. En forskers subjektive oppfatning vil kunne ha stor påvirkning på posterior sannsynlighet. Dette ses på som en svakhet innen Bayesianismen (Redmayne, 2019).

Det er ønskelig å minimere grad av subjektivitet og usikkerhet knyttet til valg av priors. Et alternativ kan være å la noen utenforstående, med høy grad av kunnskap på området bestemme prior-verdien. På samme måte kan en se til eksisterende forskning på området og benytte samme priors som er benyttet ved tidligere forskning på området. En kan også benytte flere ulike priors og se hvordan dette vil påvirke utfallet. Deretter kan man vurdere hvilken prior-verdi som passer best til sin aktuelle modell. Ulike modeller vil ha hvert sitt behov når det kommer til hvilken prior som passer best.

Et annet problem med Bayesianisme er at metoden kan være sensitiv for små justeringer i dataene, spesielt når man har lite informasjon. Dersom man starter med en usikker prior, kan selv små justeringer i de nye dataene føre til store forandringer i den posterior sannsynligheten. Dette kan føre til overdreven usikkerhet i estimatene, eller i verste fall føre til at man trekker konklusjoner som ikke stemmer overens med virkeligheten.

7 RNA ekstraksjon og analyse

7.1 Introduksjon

I forskning finnes flere metoder for å studere genuttrykk, og en av de mest brukte teknikkene er kvantitativ fluoresens-basert sanntids polymerase kjedereaksjon (qPCR). Ifølge Kuang et al. (2018) gir denne metoden en presis sanntidsmåling av genuttrykk og er særlig verdifull innen treningsfysiologi for å undersøke hvordan trening påvirker uttrykket av spesifikke gener. qPCR gjør det mulig å kvantifisere nivåene av et målgen, for eksempel i muskelvev, og studere hvordan disse nivåene kan endres som respons på fysiologiske stimuli (Kuang *et al.*, 2018).

Videre beskriver Kuang et al. (2018) hvordan qPCR kan brukes til å undersøke treningsinduserte endringer i genuttrykk, inkludert gener relatert til muskelfibertyper. Dette gir verdifull innsikt i de molekylære mekanismene bak muskeltilpasning og hvordan gener responderer på forskjellige faktorer ved trening, noe som bidrar til en dypere forståelse av treningsinduserte tilpasninger i kroppen (Kuang *et al.*, 2018).

7.2 Metode

7.2.1 Overordnet metode

Kuang et al. (2018) skriver så om metoden for å studere genuttrykk. Første steget er at RNA-et må bli cDNA (komplementært DNA). Denne prosessen kalles reversert transkripsjon. Videre vil cDNA bli kopiert slik at man får milliarder av kopier gjennom PCR prosessen. cDNA-et blir først utsatt for høy temperatur slik at den doble DNA tråden blir splittet til en enkeltråd (denaturering). Det andre steget er at temperaturen så senkes. Da vil cDNA-primere kunne binde seg (annealing). I tredje og siste steget øker temperaturen slik at primere binder seg og vi får en ny dobbelt tråd (elongering) (Kuang *et al.*, 2018).

Videre beskriver Kuang et al. (2018) hvordan PCR-syklusen gjentas flere ganger, og mengden cDNA øker eksponentielt for hver syklus. En vanlig metode for å overvåke denne prosessen i sanntid er SYBR Green-metoden. I denne metoden brukes et fluorescerende fargestoff (SYBR Green), som binder seg til den doble cDNA-tråden. Ved å bruke fluorescens kan man følge cDNA-amplifikasjonen i sanntid, ettersom det tas et bilde av fluorscensen etter hver syklus (Kuang *et al.*, 2018).

Fluorescensen øker eksponentielt med hver syklus, ettersom mer cDNA produseres. Målet med metoden er å identifisere syklisk terskel (CT)-verdien, som er den syklusen hvor fluorescensen når en forhåndsbestemt terskelverdi. CT-verdien gir et mål for hvor raskt fluorescensen når denne terskelen, og dermed hvor mye cDNA som er til stede i prøven (Kuang *et al.*, 2018).

Jo lavere CT-verdi, desto høyere nivå av det spesifikke målgenet er til stede i prøven. Derfor kan CT-verdien brukes til å kvantifisere mengden av det målte genet i forhold til et referansegene eller kontroll (Livak & Schmittgen, 2001).

7.2.2 Detaljert fremgangsmåte for qPCR

Ved start på forsøket ble ferdig cDNA utdelt fra et tidligere gjennomført styrkeprosjekt av labansvarlig. For å kunne kjøre en qPCR ble det brukt cDNA og en Master mix. Denne Master mixen bestod av 50 µl SYBR-green, 20 µl H₂O og 10 µl primer mix (myhc I, myhc IIx eller myhc IIa). Det ble også laget en Master mix som kontroll. Denne mestod av 50 µl b2m primer mix, 100 µl H₂O og 250 µl SYBR-green.

Tabell 7.1: Fortynningsrekkene som ble laget i forsøket

Fortynning	Prøve	H2O
1	30	0
1/10	2	18
1/100	2	18
1/1000	2	18
1/2	10	10
1/20	2	18
1/200	2	18

Forsøket innebar også å lage to fortynningsrekker (se Table 7.1). Det ble brukt cmyc primer som ble fortynnet med H₂O. I 1/1 prøven var det 30 µl ved start. 1/1 prøven ble benyttet i fortynningen av de andre prøvene. Der var til slutt 20 µl i alle fortynningene.

Tabell 7.2: Tabelloversikt over fortynningsrekkene i triplikat

	5	6	7	8	9	10	11
A	cmyc1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b
B	cmyc1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b
C	cmyc1	cmyc 2a	cmyc 3a	cmyc 4a	cmyc 2b	cmyc 3b	cmyc 4b

Slik ble triplikatene plassert i brønnene

Tabell 7.3: Tabelloversikt over genuttrykk

	1	2
A	myhc I	myhc I
B	myhc I	myhc I
C	myhc I	myhc I
D	myhc IIa	myhc IIa
E	myhc IIa	myhc IIa
F	myhc IIa	myhc IIa
G	myhc IIx	myhc IIx
H	myhc IIx	myhc IIx
I	myhc IIx	myhc IIx
J	b2m	b2m
K	b2m	b2m
L	b2m	b2m

Kolonne 1 = prøver ved uke 0, kolonne 2 = prøver ved uke 12

Videre ble prøvene pipetert over i brønner etter pippetingskartet, se Table 7.3. Brønnene ble fylt med 8 μ l primer-spesifikk prøve samt 2 μ l cDNA-løsning eller kontrolløsning (Brønner: Rad A-L, kolonne 1 og 2). Deretter ble fortynningsrekkene pipetert over i brønner (Rad A-C, kolonne 5-11) (Table 7.2). Dette ble utført i triplikat for alle prøvene.

Videre ble qPCR-prøven kjørt i sanntids PCR (Applied Biosystems 7500 Fast Real-Time PCR Systems, Life Technologies AS) ved bruk av Quant Studio 5. Prosessen bestod av tre ulike steg. Første steget var et «Hold stage», hvor temperaturen økes med 1,99 °C/s opp til 50 °C. Temperaturen lå deretter konstant på 50 °C i 2 minutter før den videre ble økt med 1,99 °C/s opp til 95 °C, og forble på 95 °C i 2 minutter.

Det neste steget var selve PCR-prosessen, kalt «PCR stage», som besto av 40 sykluser. Én syklus inkluderte 1 sekund på 95 °C, deretter senkes temperaturen med 1,77 °C/s til 60 °C, hvor temperaturen ble holdt konstant i 30 sekunder. Etter hver syklus ble det tatt et bilde av fluoresensen.

Siste steget, kalt «Melt stage», begynte med at temperaturen økte med 1,99 °C/s opp til 95 °C. Temperaturen ble holdt konstant i 15 sekunder, før den gradvis ble senket med 1,77 °C/s til den nådde 60 °C, hvor temperaturen ble holdt konstant i 1 minutt. Temperaturen ble deretter økt med 0,15 °C/s opp til 95 °C, og holdt konstant i 15 sekunder.

Når PCR-prosessen var ferdig, ble CT-verdiene hentet ut. Verdiene ble deretter databehandlet og analysert ved hjelp av Excel, Microsoft Office og primer effektiviteten ble beregnet.

7.3 Resultater

Tabell 7.4: Ct-verdier

Prøve	Genuttrykk	Ct1	Ct2	Ct3	Ct gj.snitt
Kontroll	b2m	23.43	24.07	23.32	23.61
U0	myhc I	18.30	19.22	19.76	19.10
U12	myhc I	18.43	19.08	18.35	18.62
U0	myhc IIa	22.42	17.71	18.31	19.48
U12	myhc IIa	18.39	18.73	35.24	24.12
U0	myhc IIx	25.71	25.04	23.77	24.84
U12	myhc IIx	25.12	23.57	23.05	23.92

Ct gj.snitt = gjennomsnitt av Ct1, Ct2 og Ct3

Endringen i CT-verdier fra uke 0 til uke 12 viser en reduksjon i antall sykluser for myhc I, mens for myhc IIa og myhc IIx er det observert en økning i antall sykluser før Ct-verdien nås (Table 7.4).

Tabell 7.5: Prosentvis fordeling av genuttrykk

Tidspunkt	myhc I	myhc IIa	myhc IIx
Uke 0	56	43	1
Uke 12	95	2	2

Andel genuttrykk (%) ved uke 0 og uke 12

Som vist i Table 7.5, økte andelen myhc I markant fra før til etter treningsintervensjonen. Andelen myhc IIx økte også noe, mens andelen myhc IIa viste en betydelig reduksjon.

Basert på de gjennomsnittlige Ct-verdiene og logaritmen av fortynningene, ble primerens effektivitet beregnet til 153 %.

7.4 Diskusjon

Målet med forsøket var å undersøke endringer i myosintungkjedene etter en styrketreningsintervensjon på 12 uker for en utrent forsøksperson ved hjelp av qPCR.

I forsøket ble det undersøkt hvor mange sykluser CT-verdiene for de ulike myosintungkjedene trengte for å nå sin sykliske terskelverdi (CT). Færre sykluser og lavere CT-verdier indikerer

større genuttrykk. Våre resultater viser en endring i antall sykluser som kreves for at myosintungkjedene skal nå sin CT-verdi. Den prosentvise endringen for tungkjedenes CT-verdier er betydelig for myhc I og IIa. Genuttrykket for myhc I økte betydelig, mens uttrykket for myhc IIa og myhc IIx viste begge en reduksjon, med en større nedgang for myhc IIa sammenlignet med myhc IIx.

I en tidligere studie av Ellefsen et al. (2014), hvor en styrketreningsintervensjon ble gjennomført på utrente individer over 12 uker, ble det observert en økning i myhc IIa, en reduksjon i myhc IIx, samt stabilitet i myhc I (Ellefsen *et al.*, 2014). I kontrast til dette viser våre resultater motstridende funn, med både reduksjon i myhc IIa og myhc IIx, samt en betydelig økning i myhc I.

Andre studier, som Terzis et al. (2008) og Andersen et al. (2000), viste også at utrente personer med overvekt av myhc IIx opplever en reduksjon i myhc IIx og en økning i myhc IIa ved trening, med minimal endring i myhc I (Andersen & Aagaard, 2000; Terzis *et al.*, 2008). Det er kjent at genuttrykk ikke kan endres fra myhc I til myhc IIa eller myhc IIx, noe som gjør det vanskelig å forklare de resultatene vi har fått fra analysen av myosintungkjeder. Dette reiser spørsmål om hva som kan ha skjedd under vår analyse og om det er spesifikke faktorer ved vårt eksperiment som kan ha bidratt til disse avvikene fra tidligere forskning.

En mulig kilde til usikkerhet kan være pippeteringsferdighetene og kvaliteten på primere som ble brukt i forsøket. Det er en risiko for at primere kan ha vært utgått på dato eller at feil primer ble valgt. Primerens effektivitet burde ligge mellom 90 % og 110 %, men i vårt tilfelle ble effektiviteten målt til 153 %. Dette kan tyde på at menneskelige feil kan ha påvirket resultatene.

Videre er det problematisk å trekke entydige konklusjoner basert på kun én prøve. I tillegg mangler vi forkunnskap om hvilken type treningsstimuli deltakerne har vært utsatt for, bortsett fra den informasjonen vi har fått fra labansvarlig.

7.5 Konklusjon

Basert på resultatene vi har fått i dette forsøket, kan vi ikke trekke noen konklusjoner om endringene i myosintungkjedene for denne forsøkspersonen. De observerte resultatene er ikke i samsvar med det som er rapportert i tidligere forskning, og derfor kan vi ikke vurdere disse funnene som representative eller pålitelige.

Referanser

- Andersen JL & Aagaard P (2000). [Myosin heavy chain IIX overshoot in human skeletal muscle](#). *Muscle & Nerve* **23**, 1095–1104.
- Antonutto G & Di Prampero PE (1995). The concept of lactate threshold. A short review. *J Sports Med Phys Fitness* **35**, 6–12.
- Breil FA, Weber SN, Koller S, Hoppeler H & Vogt M (2010). Block training periodization in alpine skiing: Effects of 11-day HIT on VO2max and performance. *Eur J Appl Physiol* **109**, 1077–1086.
- Cremers HR, Wager TD & Yarkoni T (2017). The relation between statistical power and inference in fMRI. *PLoS One* **12**, e0184923.
- Ellefsen S, Vikmoen O, Zacharoff E, Rauk I, Sletdaløkken G, Hammarström D, Strand TA, Whist JE, Hanestadhaugen M, Vegge G, Fagernes CE, Nygaard H, Hollan I & Rønnestad BR (2014). Reliable determination of training-induced alterations in muscle fiber composition in human skeletal muscle using quantitative polymerase chain reaction. *Scandinavian Journal of Medicine & Science in Sports*; DOI: [10.1111/sms.12185](#).
- Glass DJ (2010). A critique of the hypothesis, and a defense of the question, as a framework for experimentation. *Clin Chem* **56**, 1080–1085.
- Halperin I, Pyne DB & Martin DT (2015). Threats to internal validity in exercise science: A review of overlooked confounding variables. *Int J Sports Physiol Perform* **10**, 823–829.
- Haun CT, Vann CG, Mobley CB, Osburn SC, Mumford PW, Roberson PA, Romero MA, Fox CD, Parry HA, Kavazis AN, Moon JR, Young KC & Roberts MD (2019). Pre-training skeletal muscle fiber size and predominant fiber type best predict hypertrophic responses to 6 weeks of resistance training in previously trained young men. *Front Physiol* **10**, 297.
- Hopkins WG (2000). Measures of reliability in sports medicine and science. *Sports Med* **30**, 1–15.
- Hopkins WG, Marshall SW, Batterham AM & Hanin J (2009). Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc* **41**, 3–13.

- Joyce JM (2004). 132 Bayesianism. In *The Oxford Handbook of Rationality*. Oxford University Press. Available at: <https://doi.org/10.1093/oxfordhb/9780195145397.003.0008>.
- Kuang J, Yan X, Genders AJ, Granata C & Bishop DJ (2018). [An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research](#) ed. Kalendar R. *PLOS ONE* **13**, e0196438.
- Lakens D (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front Psychol* **4**, 863.
- Livak KJ & Schmittgen TD (2001). [Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- \$\Delta\Delta\$ CT Method](#). *Methods* **25**, 402–408.
- Mishra P, Singh U, Pandey CM, Mishra P & Pandey G (2019). Application of student’s t-test, analysis of variance, and covariance. *Ann Card Anaesth* **22**, 407–411.
- Pisică D, Dammers R, Boersma E & Volovici V (2022). Tenets of good practice in regression analysis. A brief tutorial. *World Neurosurg* **161**, 230–239.e6.
- Popper KR (1963). Science as falsification. *Conjectures and refutations* **1**, 33–39.
- Radaelli R, Rech A, Molinari T, Markarian AM, Petropoulou M, Granacher U, Hortobágyi T & Lopez P (2024). Effects of resistance training volume on physical function, lean body mass and lower-body muscle hypertrophy and strength in older adults: A systematic review and network meta-analysis of 151 randomised trials. *Sports Med*.
- Redmayne M (2019). Bayesianism and proof. *Science in court* 61–82.
- Rønnestad BR, Ellefsen S, Nygaard H, Zacharoff EE, Vikmoen O, Hansen J & Hallén J (2014a). Effects of 12 weeks of block periodization on performance and performance indices in well-trained cyclists. *Scand J Med Sci Sports* **24**, 327–335.
- Rønnestad BR, Hansen J & Ellefsen S (2014b). Block periodization of high-intensity aerobic intervals provides superior training effects in trained cyclists. *Scand J Med Sci Sports* **24**, 34–42.
- Rønnestad BR, Hansen J, Thyli V, Bakken TA & Sandbakk Ø (2016). 5-week block periodization increases aerobic power in elite cross-country skiers. *Scand J Med Sci Sports* **26**, 140–146.
- Rønnestad BR, Øfsteng SJ & Ellefsen S (2019). Block periodization of strength and endurance training is superior to traditional periodization in ice hockey players. *Scand J Med Sci Sports* **29**, 180–188.

- Santana DA de, Castro A & Cavaglieri CR (2021). Strength training volume to increase muscle mass responsiveness in older individuals: Weekly sets based approach. *Front Physiol* **12**, 759677.
- Schadock I, Schneider A, Silva ED, Buchweitz MRD, Correa MN, Pesquero JB, Paredes-Gamero EJ, Araujo RC & Barros CC (2015). Simple method to genotype the ACTN3 r577x polymorphism. *Genet Test Mol Biomarkers* **19**, 253–257.
- Stone MH, Stone M & Sands WA (2007). *Principles and practice of resistance training*. Human Kinetics.
- Terzis G, Georgiadis G, Stratakos G, Vogiatzis I, Kavouras S, Manta P, Mascher H & Blomstrand E (2008). [Resistance exercise-induced increase in muscle mass correlates with p70S6 kinase phosphorylation in human subjects](#). *European Journal of Applied Physiology* **102**, 145–152.
- Westcott WL (2012). Resistance training is medicine: Effects of strength training on health. *Curr Sports Med Rep* **11**, 209–216.
- Zabor EC, Kaizer AM & Hobbs BP (2020). Randomized controlled trials. *Chest* **158**, S79–S87.