

Mappeeksamen IDR4000

Kandidatnr 101

2022-02-24



# Contents

0.1	Forord . . . . .	3
<b>1</b>	<b>Reliabilitet</b>	<b>5</b>
1.1	Introduction . . . . .	5
1.2	Method . . . . .	5
1.3	Results and discussion . . . . .	6
<b>2</b>	<b>Lab-rapport</b>	<b>7</b>
2.1	Introduksjon . . . . .	7
2.2	Metode . . . . .	8
2.3	Resultater . . . . .	8
2.4	Diskusjon . . . . .	9
<b>3</b>	<b>Vitenskapetikk</b>	<b>11</b>
3.1	Falsifikasjonisme . . . . .	11
3.2	HD-metoden og abduksjon . . . . .	12
3.3	Replikasjonskrisen . . . . .	13

## 0.1 Forord

Dette er en mappeeksamen i Kvantitativ metode og statistikk (IDR4000). Denne eksamen inneholder fem deler, og alle delene med tekst og koding finnes i denne lagringsplassen:

<https://github.com/veli1793/IDR4000-bookdown-VLi>



# Chapter 1

## Reliabilitet

### 1.1 Introduction

When testing in exercise physiology you need tests that are reliable. Reliability refers to the ability to reproduce measurements from time to time. This is important to discover any possible measurement error which may make the observed value different from the true value. A reliable test should measure what is supposed to be measured. For example, a test is reliable if it produces the same results within the same conditions from time to time Hopkins [2000]. Reliable tests are important for example when studying whether a training intervention will improve maximal oxygen consumption or not [Peng et al., 2006]. If the test is not reliable it can not tell us if the improvements in oxygen consumption is true or due to measurement error. To be sure a test is reliable it will be useful to perform a reliability study. This is done by include enough participants and several trials in the same test with well done preparations and standardization. Measurement error can be detected in several ways. It is common to look at the individual variation from one trial to another. A good measure of reliability is the typical error of the percentage of the mean (CV). This means the average variation within individuals from between tests [Hopkins, 2000].

### 1.2 Method

Nine healthy sports science students participated in this reliability study. They performed a one repetition-maximum (1RM) strength test in leg press on two occasions. Trials were performed one week apart at the same time of the day. Activity the day before and energy intake at test days were individually standardized. 1RM leg press tests were performed with the Keiser Air 300 leg press machine. To establish reliability of the measure, participants had the same test

Table 1.1: Table 1. Descriptive statistics from repeated 1RM tests ( $n = 9$ ).

SD	TE	Mean (kg)	CV (%)
8.7	6.15	286.94	2.14

SD = Standard deviation, TE = Typical error, CV = Coefficient of variation

leader. Also the number and gender of observers in the room was the same for both test occasions [Halperin et al., 2015].

#### 1.2.0.1 Preperation

Before the 1RM test, the Keiser Air 300 machine was individually customized. Participants performed the test without shoes with legs on separate force plates and  $90^\circ$  in the knee joint. Seat length and leg position on the footplate were recorded for each individual to ensure similar conditions in both tests.

Three warm-up sets were performed before the 1RM test consisting of 8, 6 and 4 repetitions with increasing resistance and one minute rest in between. The resistance of the warm-up sets were recorded, to be repeated in post test.

#### 1.2.0.2 One repetition maximum test (1RM)

After warm-up sets, participants rested for one minute followed by a maximum of five trials to reach 1RM. In between each trial they rested for three minutes. The resistance gradually increased and was of each trial was recorded. The maximum weight the participants manage to perform with correct technique was noted as 1RM.

### 1.3 Results and discussion

The typical percentage error (CV) of the test was 2.14%. This means that the variation between pre and post test is 2.14% of the mean. Compared to Schroeder et al. [2007] our results of CV is smaller ( $< 6.3\%$ ), and the test seems to be acceptable and reliable. This is a small variation and the test seems to be reliable. It may be explained by elimination of measurement errors because of standardized preparation and test procedures. The small variation that was found in this reliability test may be explained by day-to-day variations which can be physiological and psychological factors, for example motivation and fatigue [Schroeder et al., 2007].

## Chapter 2

# Lab-rapport

### 2.1 Introduksjon

I dette laboratorieforsøket gjennomførte vi ekstraksjon og analyse av proteiner. Hensikten var å påvise to ulike proteiner, UBF fra cellekjernen og ribosomal S6 fra cytosol. Et protein er sammensatt av flere aminosyrer som til sammen danner flere kjeder, disse kjedene settes deretter sammen og danner sløyfer på kjeden av aminosyrer. Bindinger dannes mellom deler av kjeden som ligger langt fra hverandre og danner til slutt selve proteinet. Sammensetningen av aminosyrene i proteinet vil avgjøre hva slags oppgaver proteinet har. Proteiner kan fungere i nødvendige og kritiske oppgaver i kroppens struktur og funksjon, i tillegg til å være viktige i regulering av vev og organer. Proteiner kan for eksempel fungere som antistoffer for å beskytte kroppen mot virus og bakterier, være enzymer i fordøyelsen vår, fungere som transportproteiner som bringer signaler over cellemembraner [Alberts, 2015] eller være viktige for dannelsen av nye proteiner.

Proteiner som er viktige for å danne nye proteiner er blant annet UBF og S6. UBF er en transkripsjonsfaktor som finnes i cellekjernen. Denne spiller en viktig rolle i aktiveringen av transkripsjonen av ribosomalt RNA. UBF muliggjør binding av enzymet polymerase 1 for senere å kunne starte transkripsjon av ribosomalt RNA [Copenhaver et al., 1994]. S6 protein er et ribosomalt protein som er en del av underenheten 40S. Underenheten 40S er en del av ribosomet som finnes i Cytosol. S6 proteinet er involvert i translasjonen av nye proteiner [Puighermanal et al., 2017].

For å påvise proteiner både i cytosol og cellekjernen kreves ulike buffere med tanke på hvor i cellen de befinner seg. For å finne proteiner i cytosol brukes Hepes buffer, og Laemmle buffer for å finne proteiner i cellekjernen. Etter dette ble proteinene separert og analysert ved bruk av western blotting. Vi forventet å finne UBF innenfor 94-97 Kilodaton, og S6 på 32 Kda.

## 2.2 Metode

**Homogenisering** For å detektere proteiner benyttes ofte en metode som kalles western blotting. Dette er en metode som går ut på å påvise proteiner gjennom binding og gjenkjennelse av antistoffer til ett eller flere proteiner vi er ute etter [Bass et al., 2017]. I forkant av western blotting må prøvene homogeniseres. Dette innebærer å skille ut de proteinene vi er interessert i. Prøvene vi skulle analysere var muskelfibre fra muskelbiopsi. Vi benyttet to ulike buffere, Hepes og Laemmle, for å påvise ønskede proteiner. Først ble muskelfibre homogenisert med Hepes, og sentrifugert slik at supernatant og pellet ble dannet. Disse substansene ble skilt fra hverandre, og pellet ble deretter homogenisert med Laemmle og sentrifugert. Igjen dannet det seg supernatant og pellet, som ble separert fra hverandre. Proteinkonsentrasjonene i Hepes og Laemmle ble målt gjennom sammenligning med Bradford reagent (standardkurve).

**Western blotting** Når prøvene ble homogenisert og proteinkonsentrasjonen ble beregnet, så ble prøvene loadet i en gel (sds-page). Ladderen gikk i første brønn, og deretter prøvene. Hver prøve var todelt og bestod av en med hepes buffer og en annen med laemmle buffer. Deretter koblet vi på strøm i 40 minutter med 250 volt. Etter proteinene ble separert i gelen var neste fase en electro transfer, her skulle proteinene overføres fra gelen og over til en membran.

**Blocking og antistoffer** Membranen ble delt opp i flere deler og hver del ble blandet med et blokkeringsmiddel som bestod av tørrmelk, tween-20 og TBS buffer (Tris-Buffered Saline). Etter 60 minutter ble blokkeringsmiddelet fjernet. To ulike primære antistoffer, rRNA s6 og UBF, ble deretter blandet med membranene. Prøvene stod i et kjøleskap på et vippebrett i et døgn, for å sørge for at antistoffene festet seg til proteinene skikkelig. Etter å ha stått i et døgn ble membranene vasket. Sekundære antistoffer ble deretter tilsatt membranene. De sekundære antistoffene gjenkjenner og binder seg til de primære antistoffene. Fordi de sekundære antistoffene gir fra seg lys, kunnen vi indirekte detektere proteinene vi var ute etter når vi tok bilde av membranene. Membranene ble vasket etter å ha stått i 1-2 timer med sekundære antistoffer. Deretter tok vi bilde av hver membran. Bildene ble analysert på en datamaskin, der bildene av membranene ble lagt over bildene av ladderen som var vår referanse. Med linjal sammenlignet vi molekylærvekten til hvert protein.

## 2.3 Resultater

Resultatene viste at vi fant både UBF og s6 i forsøket. Vi fant både UBF og s6 i Laemmle bufferen. Selv om s6 ble påvist i større grad med Laemmle bufferen, kunne man til dels også se dette proteinet i Hepes bufferen.



## 2.4 Diskusjon

Som forventet fant vi UBF i Laemmle løsningen. Fordi Laemmle er en sterk løsning, skal denne bufferen kunne finne proteiner i cellekjernen. Vi regnet med å finne s6 med Hepes siden dette proteinet finnes i cytosol. Hepes er en svakere buffer enn Leammle og skulle potensielt ha funnet proteinene som befinner seg i cytosol. Vi fant mest av s6 i Leammle. Dette kan skyldes at s6 også finnes i cellekjernen (pga. transport inn og ut av cellekjernen). En annen mulighet er at ikke alle proteiner som ikke befinner seg i cellekjernen ble skilt ut gjennom en gjennomføring av homogenisering med Hepes. Hadde vi gjennomført dette en gang til (homogenisering med Hepes), kan det hende vi hadde funnet mer s6 i Hepes løsningen, fordi vi muligens hadde sikret oss at alle proteiner som ikke finnes i cellekjernen hadde blitt skilt ut. En annen mulig feilkilde kan være at maskinen som tok bilde av membranen ikke klarte å tilpasse seg det sterke lyset fra prøvene som lå i Laemmle bufferen. Dette kan muligens forklare de svake båndene som viste s6 i Hepes prøvene på bildet som ble tatt.

Western blotting er en metode som innebærer mange ulike steg, og er en kompleks prosess. For at resultatene skal være reliable og valide kreves nøyaktighet. Man må blant annet være konsekvent når det kommer til valg av buffer for homogenisering og ekstraksjon av ønskede proteiner. Gelsammensetningen har noe å si for hvordan proteinene vandrer etter molekylørstørrelse, og både spenningen man tar i bruk og hvor lenge elektroforese og elektrotransfer skal foregå kan påvirke utfallet for proteintransporten. Dersom varigheten er for lang kan små proteiner "forsvinne" fordi de potensielt kan vandre ut av gelen og membranen. I tillegg kan varigheten på blockingen påvirke bindingen av primære antistoffer til målprotein. For eksempel kan for lite varighet på blockingen føre til at primære antistoffer binder seg til ikke-spesifikke steder på membranen, slik at målproteinet ikke blir påvist. Bildeprogrammet og type kamera for analyse av båndene kan også påvirke resultatene, og er noe man må ha i bakhodet når man fremstiller resultatene. Western blotting må brukes med nøyaktig implementering og forståelse for gjennomføringen for å oppnå riktige og gode analyser av resultatene [Bass et al., 2017]



## Chapter 3

# Vitenskapetikk

### 3.1 Falsifikasjonisme

Poppers falsifiseringskriterium bygger på at de ulike teoriene som forskerne kommer opp med må være falsifiserbare. Det vil si at hvis enkelte av oppsamlede data strider mot teorien så skal det være mulig å motsi teorien. Teorier som er laget for å passe alle data og ikke kan motsies vil bli kalt for “psuedo-science” (“ikke vitenskap”), som forskere tror på, men har ufullstendige svar. Popper kritiserte Marx og Freuds sine teorier. Om teoriene til Marx møtte motstand i testingen, så endret han litt på teorien slik at den kunne forklare motstanden til dataene. På den måten unngikk han å måtte forkaste teorien sin. Freuds sine teorier mente Popper var myter som ikke var mulig å teste, og når teorier ikke kan testes, kan de heller ikke aksepteres som “science” (vitenskap). Popper mente at det var et klart skille mellom “science” og “psuedo-science”, og at hans falsifikasjonskriterium løste det han kaller “problem of demarcation” [Okasha, 2016, Popper, 1969]. Okasha mente at det ikke i det hele tatt er noe avgrensning på hva som kan kalles vitenskap, da han mener at vitenskap er en heterogen aktivitet. Han viser til at Wittgenstein mente at There is no fixed set of features that define what it is to be a ‘game’. Rather there is a loose cluster of features most of which are possessed by most games. But any particular game may lack any of the features in the cluster and still be a game. (Okasha, 2016, s. 15) Okasha mente at det samme som Wittgenstein forklarer kan også gjelde for vitenskap, da forskere ikke bare forkaster teorien sin så fort den møter motstand, men av naturlige grunner forsker videre og leter etter andre forklaringer for å støtte om teorien sin. Om teorier bare skulle blitt forkastet så fort den møtte motstand, så ville vitenskapelige funn gått mye tregere. I tillegg ville det vært svært vanskelig å finne teorier som passet til de dataene man har, uten at noen av de strider imot. (Okasha, 2016) Okasha hadde også et eksempel på at på at Poppers falsifiserbarhetkriterium var tvilsom. Adams og Leverrier oppdaget en ny planet ved å bygge videre på Newton’s gravitasjonsteori, som

på grunn av en feil kunne blitt forkastet. Okasha mente dermed at Popper ikke kunne klandre Marx sin teori, ettersom Adams og Leverrier gjorde akkurat det samme, og som resulterte i et viktig funn. Med dette mener Okasha at det ikke ser ut til at det finnes noe avgrensning om hva som kan kalles vitenskap. Jeg mener at Okasha har rett, da det ikke er logisk å bare deduktivt vise til at den ene eller andre teorien er riktig. Verden er for uforutsigbar til at man kan være 100% sikker på om teorier er riktige eller ikke. Det er heller ikke riktig å bare avskrive en teori bare fordi den møter litt motstand, for da vektlegger man feil altfor høy verdi i forhold til det riktige. Graden av motstand eller støtte fra dataene teoriene møter på, og antall studier som har undersøkt teorien, vil vise hvor mye statistisk styrke teorien får. Det vil igjen gjøre det lettere å finne ut hvor mye vi kan stole på om denne teorien er sann.

## 3.2 HD-metoden og abduksjon

HD-metoden handler om at vi ut ifra tidligere data lager en teori/hypotese som kan føre til nye funn. Om vi kan dedusere/trekke slutninger på om utvalgte data fra teorien er riktig. Hvis de utvalgte dataene gir et bekreftende svar på at teorien stemmer, så vil de dataene gi en induktiv bekreftelse på at teorien er sann. En induktiv bekreftelse vil si at teorien bekreftes til en viss grad. Grunnen til at vi sier at teorien bekreftes til en viss grad er teoriene aldri kan bekreftes til å være 100% sann. Hempel [1966] forklarer dette med at fremtidige relevante data alltid kan gi andre funn enn det dataene hittil har vist. Et eksempel på dette kan være at styrketrening med tunge vekter og få repetisjoner gjør musklene våre sterkere enn styrketrening med lettere vekter og flere repetisjoner. Det er høyt sannsynlig at dette stemmer i og med at mye av forskningen viser dette, men man kan aldri bevise 100% sikkert at det er sant! For alt vi vet så kan teorien i framtiden få andre resultater som vil gi mindre sjanse for at den teorien stemmer. En kan aldri vite 100% sikkert med induksjon, men man kan tillegge teorien mer eller mindre troverdighet ut ifra hvor mye støtte den får. HD-metoden har også svakheter, for vi kan ikke dedusere data fra teorien hvis teorien handler om frekvensen for at noe kan skje. Et eksempel kan være frekvensen å få diabetes type 2 etter år med inaktivitet. Hvis man lager en graf som viser at sjansen for å få diabetes type 2 øker med for eksempel 5% for hvert år med inaktivitet, så vil man få stabil linje. Problemet er at det også er andre teorier som kan spille en rolle i utviklingen av diabetes type 2, blant annet inntak av fet og kaloririk mat. Dette kan gjøre at frekvensen av diabetes type 2 vil se helt annerledes ut i grafen. HD-metoden tar ikke hensyn til at ulike faktorer kan påvirke dataene. Den tar en empirisk test som viser om dataene er riktig overfor teorien om inaktivitet. Det finnes metoder som er ment for å løse problemet til HD-metoden, kalt Abduksjon. Abduksjon er nesten lik som HD-metoden, hvor forskjellen ligger i at Abduksjon utformer teorier som forsøker å gi den beste forklaringen til de aktuelle dataene, istedenfor å dedusere. I teoriene som blir laget kan det legges til ulike årsaker som kan forklare flere data. Teorien vil

da få bedre forklaringskraft, men blir til gjengjeld mer avansert jo flere årsaker som legges til grunn for teorien. Meningen med Abduksjon er at teorien som utformes skal ha den beste forklaringskraften, men samtidig være så enkel som mulig (få årsaker i teorien). Om én teori har én årsak som forklarer to data, mens en annen teori har to årsaker som forklarer to data, så er det teorien med færrest årsaker som vinner, da den er enklest, og har like god forklaringskraft som teorien med to årsaker. Om en teori med to årsaker kan forklare tre data, så vil den teorien ha bedre forklaringskraft, men til gjengjeld være mer avansert enn teorien med en årsak. Oppsummert, hvis man skal forklare et fenomen så vil HD-metoden gå ut ifra en teori, dedusere for å se om dataene er riktig, og så gi en induktiv bekreftelse basert på hvor sannsynlig det er at teorien er sann. HD-metoden tar for seg bare en teori om gangen som skal forklare fenomenet, mens Abduksjon sammenligner flere teorier basert på enkelthet og forklaringskraft, og velger den teorien med færrest årsaker og kan forklare mest data.

### 3.3 Replikasjonskrisen

Vitenskapen har vært utsatt for en krise når det gjelder å kunne replisere originale forskningsartikler. Mange artikler er rett og slett ikke repliserbare, og Bird [2020] har analysert dette og kommet med sin forklaring. Det er en generell frykt for å grave opp i de artiklene som ikke er repliserbare, da det vil svekke tilliten til vitenskapen i offentligheten. For det er mange store studier som viser seg å inneholde falsk positive resultater, når hypotesene som blir brukt i seg selv gir veldig lav sjanse for å gi riktig positive resultater. Det kommer av at forskeren utelukkende baserer seg på et spesielt bevis på det fenomenet som blir forsket på, uavhengig av frekvensen av at fenomenet oppstår. Dette kalles basfrekvensfeilen. I studier som gir falske positive resultater, har ofte ikke forskeren re-testet flere ganger, så da er det stor sjanse for at forskeren kommer fram til et positivt resultat som viser seg å være falsk positiv. Om forskeren derimot velger å re-teste hypotesene, så vil nok forskeren få flere negative resultater, men til gjengjeld kunne få mer riktige resultater. Å replisere studier kan være krevende og dyrt, og som Bird nevner er det også mulighet å feile med å replisere slik at de ikke støtter suksessfulle teorier. At det repliseres lite studier innen medisin mener Bird har en sammenheng med økonomisk og moralsk press for selskaper som skal finne opp medisiner som gir resultater. Andre grunner som Bird diskuterer er lav statistisk styrke, for høy “alpha” (signifikantgrense p-verdi), publikasjonsskjevhet, tvilsom forskningspraksis og svindel. Med lav statistisk styrke så blir studier med for få forsøkspersoner trukket frem, for det vil gjøre det vanskeligere for replikasjonsstudiene å finne samme effekt. Det gjør også at vi må ha høy “alpha”. Med høy statistisk styrke så kan vi senke “alpha”, og dermed kunne øke PPV (sjansen for at hypotesen faktisk er sann, gitt at testingen har fått positivt utslag). Publikasjonsskjevheten viser at det er for mange studier med positivt resultat som er publisert i forhold til studier med negativt resultat. Grunner til det er at både tidsskrifter og forskere kan bare

droppe å publisere artikler som ikke gir et positivt resultat. For hvis studier med negative resultater blir publisert, så vil offentligheten tvile på at hypotesen er sann. Det som er uheldig med det er at falske positive studier kan bli mer akseptert. Tvil som forskningspraksis og svindel er også grunner til at studier ikke kan repliseres. Først og fremst er det standardiseringen studiene har hatt som er kritikkverdig, men også P-hacking hvor forskerne bruker egne programmer til å lage falske p-verdier for å kunne støtte hypotesen. I psykologiforskningen er det ofte færre folk i forskningsgruppene, så ifølge Bird får studiene i dette fagfeltet ofte gå gjennom uten å bli forøkt replisert. Bird har mye rett i det han kommer med. Det handler mye om skjevhet i publiseringen. De fleste forskere ønsker å finne noe av stor verdi, så da kommer det ikke som noe sjokk at mange legger for stor vekt på de positive resultatene, som oftest vil føre til falsk positive resultater. Som nevnt over er det også mye økonomisk press på selskaper til å finne positive resultater, som vil bidra til å kunne legge for mye vekt på de positive resultatene, samt kunne ty til svindel. Lav statistisk styrke på grunn av for få forsøkspersoner bidrar til at det blir vanskelig replisere studiene, som kan gi både type-I-feil og type-II-feil.

# Bibliography

- Bruce Alberts. *Molecular Biology of the Cell*. Garland Science, Taylor and Francis Group, 2015. Google-Books-ID: tLSuoQEACAAJ.
- J. J. Bass, D. J. Wilkinson, D. Rankin, B. E. Phillips, N. J. Szewczyk, K. Smith, and P. J. Atherton. An overview of technical considerations for western blotting applications to physiological research. *Scandinavian Journal of Medicine and Science in Sports*, 27(1):4–25, 01 2017. doi: 10.1111/sms.12702. URL <https://onlinelibrary.wiley.com/doi/10.1111/sms.12702>.
- 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>.
- Gregory P. Copenhaver, Christopher D. Putnam, Michael L. Denton, and Craig S. Pikaard. The rna polymerase i transcription factor ubf is a sequence-tolerant hmg-box protein that can recognize structured nucleic acids. *Nucleic Acids Research*, 22(13):2651–2657, 1994. doi: 10.1093/nar/22.13.2651. URL <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/22.13.2651>.
- Israel Halperin, David B. Pyne, and David T. Martin. Threats to internal validity in exercise science: A review of overlooked confounding variables. *International Journal of Sports Physiology and Performance*, 10(7):823–829, 2015. doi: 10.1123/ijsp.2014-0566. URL <http://journals.humankinetics.com/doi/10.1123/ijsp.2014-0566>.
- Carl Gustav Hempel. *Philosophy of Natural Science*. Prentice-Hall, 1966. Google-Books-ID: Iau78s98cqWC.
- Will G. Hopkins. Measures of reliability in sports medicine and science:. *Sports Medicine*, 30(1):1–15, 2000. doi: 10.2165/00007256-200030010-00001. URL <http://link.springer.com/10.2165/00007256-200030010-00001>.
- Samir Okasha. *Philosophy of Science: A very Short Introduction*. Oxford University Press, 2 edition, 2016.

- Roger D. Peng, Francesca Dominici, and Scott L. Zeger. Reproducible epidemiologic research. *American Journal of Epidemiology*, 163(9):783–789, 2006. doi: 10.1093/aje/kwj093. URL <http://academic.oup.com/aje/article/163/9/783/108733/Reproducible-Epidemiologic-Research>.
- Karl R. Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge (3. utg.)*. London: Routledge. Routledge and Kegan Paul, 1969.
- Emma Puighermanal, Anne Biever, Vincent Pascoli, Su Melser, Marine Pralong, Laura Cutando, Stephanie Rialle, Dany Severac, Jihane Boubaker-Vitre, Oded Meyuhas, Giovanni Marsicano, Christian Lüscher, and Emmanuel Valjent. Ribosomal protein s6 phosphorylation is involved in novelty-induced locomotion, synaptic plasticity and mrna translation. *Frontiers in Molecular Neuroscience*, 10:419, 12 2017. doi: 10.3389/fnmol.2017.00419. URL <http://journal.frontiersin.org/article/10.3389/fnmol.2017.00419/full>.
- E. T. Schroeder, Y. Wang, C. Castaneda-Sceppa, G. Cloutier, A. F. Vallejo, M. Kawakubo, N. E. Jensky, S. Coomber, S. P. Azen, and F. R. Sattler. Reliability of maximal voluntary muscle strength and power testing in older men. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 62(5):543–549, 2007. doi: 10.1093/gerona/62.5.543. URL <https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/62.5.543>.