



**Eksamen i Emne OPG3401**

## **Bacheloroppgave**

**«Sammenligning av bip-test og 3000 m resultater hos unge menn på seleksjon til Luftforsvarets Flygerskole»**

**av**

**Kadett Andreas Tuft**

**Godkjent for offentlig publisering**

## Publiseringsavtale

### En avtale om elektronisk publisering av bachelor/prosjektoppgave

Kadetten har opphavsrett til oppgaven, inkludert rettighetene til å publisere den.

Alle oppgaver som oppfyller kravene til publisering vil bli registrert og publisert i Bibsys Brage når kadetten har godkjent publisering.

Opgaver som er graderte eller begrenset av en inngått avtale vil ikke bli publisert.

Jeg gir herved Luftkrigsskolen rett til å gjøre denne oppgaven tilgjengelig elektronisk, gratis og uten kostnader	<input checked="" type="checkbox"/> Ja	<input type="checkbox"/> Nei
Finnes det en avtale om forsinket eller kun intern publisering? (Utfyllende opplysninger må fylles ut)	<input type="checkbox"/> Ja	<input checked="" type="checkbox"/> Nei
Hvis ja: kan oppgaven publiseres elektronisk når embargoperioden utløper?	<input type="checkbox"/> Ja	<input type="checkbox"/> Nei

## Plagiaterklæring

Jeg erklærer herved at oppgaven er mitt eget arbeid og med bruk av riktig kildehenvisning.

Jeg har ikke nyttet annen hjelp enn det som er beskrevet i oppgaven.

Jeg er klar over at brudd på dette vil føre til avvisning av oppgaven.

**Dato: 01 – 08 – 2022**

## Forord

Først og fremst vil jeg takke veilederen min, Vegard Finserås Rasdal, som har hjulpet meg gjennom hele prosessen. Jeg setter enorm pris på at han stiller opp for meg og avsetter tid under hele sommerferien. Jeg kunne ikke ha skrevet oppgaven uten han.

Takk til Per William Bertelsen ved Luftforsvarets Rekruttering- og Seleksjonssenter og sjefspsykolog i Forsvaret, Ole Christian Lang-Ree for forslag til oppgave, anskaffelse og sammenstilling av data.

Sist, vil jeg takke min Excel-ekspert og far Alexander som har lært meg å behandle data i Excel.

## Innhold

Forord.....	3
1. Innledning.....	5
1.1 Problemstillinger .....	6
1.2 Avgrensning og disposisjon .....	6
2 Metode.....	7
2.1 Kvantitativ metode .....	7
2.2 Utvalg.....	7
2.3 Valg av metode.....	9
2.4 Testprotokoller .....	11
2.5 Databehandling og prosedyre .....	13
2.6 Validitet og Reliabilitet .....	13
3 Teori .....	15
3.1 Utholdenhet og kondisjon.....	15
4 Seleksjon til Luftforsvarets flygerskole.....	18
4.1 Generelt om opptak .....	18
4.2 Bakgrunn for seleksjon.....	18
4.3 Seleksjonsmodell.....	18
4.4 Seleksjonen i praksis .....	19
4.4.1 Fysiske tester på LRSS .....	19
5 Resultater.....	20
5.1 Deskriptiv analyse .....	20
5.2 Korrelasjonsanalyse.....	21
5.3 Karakterer analyse .....	22
6 Diskusjon.....	23
6.1 Hovedproblemstilling .....	23
6.2 Underproblemstilling.....	25
6.3 Generalisering for resten av Forsvaret?.....	28
6.4 Oppsummering .....	29
7 Konklusjon .....	29
Referanser.....	31
Vedlegg A .....	33

## 1. Innledning

Tjeneste i Forsvaret har alltid forutsatt fysisk trening og arbeid, men standardisert testing av fysisk form er relativt nytt. Fysiske tester for Hæren ble beskrevet først i 1966 og de første testene på sesjon ble gjennomført i 1968 (Kirknes, Aandstad, & Stornæs, 2014). Det var først på 1970-tallet at den generelle fysiske testordningen for hele Forsvaret ble opprettet, og siden da har det vært flere reglement og endringer for å standardisere testene. Testbatteriet i dag består av både utholdenhetstester og styrketester. Av utholdenhetstester er det 3000 m løp som er det foretrukne i det norske forsvaret. Denne testen, eller tilsvarende utholdenhetstester (som 12 minutters løp (Cooper-test), 1.5 mile- eller 2 mile run), er mye brukt blant militære verden over (Kirknes, Aandstad, & Stornæs, 2014). Alternativt til 3000 m så kan bip-test, også kjent som *20 m Shuttle Run Test*, gjennomføres dersom forhold ikke tillater 3000 m (f.eks. ikke egnet rundløype, is eller snø, etc.). Bip-test har ikke samme historikk som 3000 m i det norske forsvaret, og ble først introdusert som et gyldig alternativ i 2017 (A. Aandstad, personlig kommunikasjon, 18. juli, 2022).

Forsvaret og Luftforsvaret har lange tradisjoner med bruk av testing i seleksjonssammenheng. Ved seleksjon til Luftforsvarets flygerskole (LFS) er fysiske tester en del av opptaket selv om det er kun en liten del av seleksjonen. I alle seleksjonssammenhenger er det viktig at prosessen er standardisert slik at ingen enkeltgruppe får en fordel. På opptak til LFS blir man testet i både Bip-test og 3000 m. Bakgrunnen for to utholdenhetstester på opptaket er hovedsakelig for statistikken, men også for å holde søkerne opptatt i seleksjonsprosessen (P.W. Bertelsen, personlig kommunikasjon, 28. juni, 2022). Ved å ha to tester kan man se hvilken av testene som er mest valid og om det er stor forskjell mellom dem. Seleksjonssjefen legger til at de som kun greier én test blir ikke nødvendigvis tatt ut av seleksjonen, men at vedkommende vil få oppfølgingsspørsmål i etterkant for å vurdere om søkeren er skikket (P.W. Bertelsen, personlig kommunikasjon, 28. juni, 2022).

Etter en revidering av Forsvarets fysiske tester i 2014 ble bip-test foreslått som et gyldig alternativ til 3000 m (Kirknes, Aandstad, & Stornæs, 2014). Det var ikke før 1. januar 2017 at bip-test ble implementert som et gyldig alternativ til 3000 m. Bip-test vurderes til å gi relativt reliable og valide estimater av  $VO_{2maks}$ , samt sterk korrelasjon mot andre distanseløp (Kirknes, Aandstad, & Stornæs, 2014, s. 60). Dette vil si at testen gir en god indikasjon på utholdenhetskapasitet, og kan likestilles med andre utholdenhetstester i så måte. Allikevel, er det flere soldater som foretrekker å løpe bip-test framfor 3000 m (Kirknes, Aandstad, & Stornæs, 2014). Av egen erfaring, med tre år i Forsvaret, er jeg en av disse og opplever selv at

jeg presterer bedre i bip-test enn 3000 m målt ved oppnådd karakter. Jeg merker at jeg må yte mye mer for å oppnå den beste karakteren (karakter 9) på 3000 m enn bip-test. I så fall er det kanskje ikke så rart at det er flere soldater som foretrekker å løpe bip-test? Med det kan man stille spørsmålet: Er det faktisk lettere å oppnå gode resultater på bip-test enn 3000 m selv om den er likestilt?

Med dette som bakgrunn skal jeg forsøke å besvare problemstillingene under.

### 1.1 Problemstillinger

#### Hovedproblemstilling:

Er det en sammenheng mellom prestasjon på 3000 m og bip-test hos unge menn på seleksjon til Luftforsvarets flygerskole?

#### Underproblemstilling:

Med bruk av Forsvarets karaktersettingssystem, er det noen forskjell i oppnådd karakter på 3000 m og bip-test hos unge menn i Forsvaret?

### 1.2 Avgrensning og disposisjon

I teoridelen vil jeg innledningsvis beskrive teori knyttet til arbeidsfysiologi, arbeidskrav og hvordan Forsvarets minstekrav (MK) er satt opp. Arbeidsfysiologi og minstekrav står sentralt i begge problemstillinger og en god forståelse for hvordan det henger sammen med testene er viktig for å forstå videre gjennomgang av resultater og drøfting. Innsamlingen av data er gjort over flere år med testing og registrering ved Luftforsvarets Rekruttering- og Seleksjonssenter (LRSS). Data er for formålet av denne oppgaven hentet ut av LRSS sitt register for retrospektive analyserer. Seleksjonsprosessen for opptaket til LFS vil også kort beskrives.

I metode og resultatdelen vil tabeller og funn fra databehandling bli presentert, med en påfølgende drøfting. Avslutningsvis vil jeg oppsummere og konkludere oppgaven.

## 2 Metode

### 2.1 Kvantitativ metode

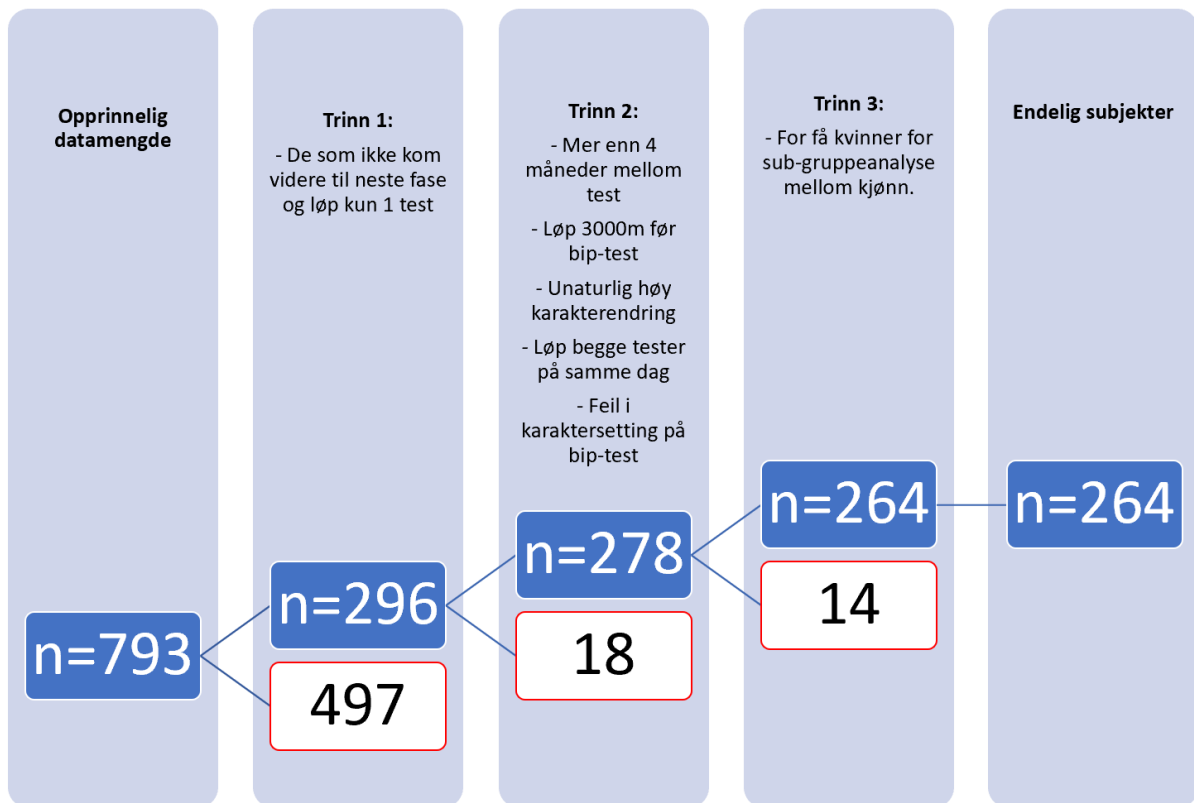
Kvantitativ metode omhandler det å samle inn og analysere kvantitativ data. Dette er data som foreligger i form av tall eller andre mengdetermer (Grønmo, 2022). Det vil også være forankret til spesifikke variabler, og innhentningsmetodene for datainnsamling er standardiserte (Kaiser, 2022). Kvantitativ metode tillater forskere å undersøke statistiske mål som gjennomsnitt, variasjon og korrelasjon hos målgruppen og analysere ved hjelp av ulike analyseverktøy. Resultatene i denne oppgaven er basert på utholdenhetstestene 3000 m og bip-test for søkerne ved flygeropptak til Luftforsvarets flygerskole (LFS). Dataen er hentet fra Luftforsvarets Rekrutterings- og Seleksjonssenter (LRSS).

### 2.2 Utvalg

Undersøkelsen består av ett utvalg med subjekter: Søkere ved fase 2a og 2b (fase 2 og 4 fremover) av flyskoleopptak mellom 2018 og 2022. Bakgrunnen for valg av denne målgruppen er flerdelt. For det første er det sannsynlig at subjektene som deltar på flyskoleopptak har høy motivasjon for å prestere på de fysiske testene, da resultatet vil ha påvirkning på om de kommer gjennom opptaket eller ikke. For det andre sikrer gjennomføring og registrering av testresultatene ved LRSS en database som muliggjør en retrospektiv undersøkelse. Det opprinnelige datasettet gjort tilgjengelig fra LRSS til denne oppgaven bestod av 1386 testresultater fra totalt 793 subjekter. Av disse 793 ble 497 ekskludert grunnet at det var kun gjennomført én utholdenhetstest (disse kom ikke videre til neste fase av opptaket). Totalt 18 stykker ble deretter ekskludert grunnet ulike feilkilder og ekstremavvik (i.e. verdier som er ulik fra resten av datamengden og skiller seg ut (Field, 2013)). Av disse 18 var det 12 som hadde mer enn 4 måneder mellom gjennomført bip-test og 3000 m, 1 som gjennomførte testene i motsatt rekkefølge enn resten (3000 m før bip-test), 1 som gjennomførte begge utholdenhetstestene på samme dag, 2 som viste unaturlig store avvik i karakter mellom testene (mer enn 4 karakterer forskjell mellom testene) og 2 som hadde fått registrert resultat på bip-test (nivå og shuttle) som ikke eksisterer. Eksklusjonsprosessen er illustrert i Figur 1. Etter totalt 515 subjekter var ekskludert, var det for få kvinnelige subjekter (n=14) igjen til å kunne gjøre sub-gruppeanalyse mellom kjønn. Da det er fysiologiske forskjeller mellom kvinner og menn på gruppenivå som gir utslag i prestasjon på utholdenhetstester, og hvor det også er egne karakterskalaer for kvinner og menn, ble videre analyser derfor kun gjennomført for resterende n=264 mannlige subjekter 18-24 år for å unngå en potensiell feilkilde i tolkning av data.

### Figur 1: Oversikt over eksklusjonsprosessen

Figur 1 viser hvordan eksklusjonsprosessen ble gjennomført. Etter tre trinn i eksklusjonsprosessen ble totalt 529 subjekter fjernet fra utvalget. Mørkeblå bokser viser antall subjekter som kom videre, mens hvite bokser viser antall subjekter som ble ekskludert fra utvalget.



Testresultatene er fra tidsrommet 2018-2022. I 2020 endret derimot karakterkravene seg noe for bip-test (Aandstad, personlig kommunikasjon, 06. juli, 2022). Karakterene til alle subjektene som gjennomførte bip-test før karakterendringen i 2020 ble derfor justert til den nye karakterskalaen basert på den absolutte prestasjonen i antall fullførte nivå og lengder, for å kunne gjøre nødvendig sammenligning av karakter mellom bip-test og 3000 m.

Karakterskala for menn ved opptak til militær utdanning og tjeneste på bip-test før og etter karakterendringen er vist i Tabell 1a og 1b.



Tabell 1a: Minimumskrav for militært tilsatte menn  $\leq 29$  år og menn ved opptak til militær utdanning og tjeneste før 2020

			Minimumskrav før 2020								
Egenskap	Testøvelse	Enhet	1	2	3	4	5	6	7	8	9
Utholdenhet	3000 m løp	min:sek	18:00	16:30	15:00	14:30	14:00	13:30	13:00	12:30	12:00
	Bip-test	Level:shuttle	6:1	7:4	8:8	9:3	9:8	10:2	10:7	11:1	11:6

Tabell 1b: Minimumskrav for militært tilsatte menn  $\leq 29$  år og menn ved opptak til militær utdanning og tjeneste etter 2020

			Minimumskrav etter 2020								
Egenskap	Testøvelse	Enhet	1	2	3	4	5	6	7	8	9
Utholdenhet	3000 m løp	min:sek	18:00	16:30	15:00	14:30	14:00	13:30	13:00	12:30	12:00
	Bip-test	Level:shuttle	6:3	7:2	8:3	8:8	9:3	9:8	10:3	10:11	11:8

Innsamling av data blir gjennomført årlig og registrert av LRSS under opptak. I mai 2022 hentet jeg ut dataen i samarbeid med LRSS. Alderen til søkerne varierer mellom 18 og 24 år ved opptaksår. Luftforsvaret har satt et alderskrav slik at det er bare personer mellom 18 og 24 år som får lov til å delta på seleksjonen.

### 2.3 Valg av metode

For å undersøke en eventuell sammenheng mellom 3000 m tid og bip-test tid ble det gjennomført korrelasjonsanalyser med bruk av Pearsons  $r$  og Spearmans  $\rho$ .

Korrelasjonskoeffisienter, som Pearsons  $r$ , måler hvor sterkt forholdet er mellom to variabler. Det finnes ulike korrelasjonskoeffisienter, men Pearsons er en av de vanligste for kontinuerlige data (Glen, 2022). Slike korrelasjonskoeffisienter gir uttrykk for hvor sterk forholdet mellom to variabler er ved å utgi en verdi mellom -1 og 1. Hvis koeffisienten er -1 så er det en sterk negativ korrelasjon, og hvis den er 1 så er det en sterk positiv korrelasjon. Er den på 0, så er det ingen korrelasjon. I denne oppgaven har jeg benyttet Hopkins sin tabell for vurdering av korrelasjonskoeffisienten (Tabell 2) (Hopkins, 2006).

Tabell 2: Hopkins tabell over grad av effekt ved korrelasjon

Correlation Coefficient	Descriptor
0.0-0.1	trivial, very small, insubstantial, tiny, practically zero
0.1-0.3	small, low, minor
0.3-0.5	moderate, medium
0.5-0.7	large, high, major
0.7-0.9	very large, very high, huge
0.9-1	nearly, practically, or almost: perfect, distinct, infinite

Tabellen er hentet fra Hopkins sin nettside om effektstatistikk (Hopkins, 2006).

Ettersom at Pearsons korrelasjonskoeffisient er den vanligste for kontinuerlig data ble den forsøkt først. For å kunne gjennomføre en korrelasjonsanalyse med Pearsons  $r$ , må man undersøke om alle antakelsene som ligger til grunn for utregningen av Pearsons korrelasjonskoeffisient er oppfylt. For at Pearsons  $r$  skal gi et presist svar må følgende forutsetninger være møtt: det er to kontinuerlige variabler, variablene må være parete, det må være et lineært forhold mellom variablene, begge variabler bør være normalfordelt og det må heller ikke være noen «outliers» / ekstremavvik (Field, 2013). Bip-test resultatene (nivå og shuttle) ble regnet om til gjennomført tid (minutter) for at begge variablene ble kontinuerlig og parete. Det lineære forholdet mellom variablene ble sett visuelt og alle ekstremavvik ble fjernet ved utvalget (Figur 1, Trinn 2). For å sjekke om dataen var normalfordelt ble tidene for 3000 m og tidene for bip-test testet med Jarque-Bera test. Jarque-Bera testen ser om dataen har kurtosen og skjevheten for å kunne bestemme om det er en normalfordeling (Zach, 2021). Testen benytter seg av en nullhypotese hvor  $H_0$  betyr at dataen er normalfordelt og  $H_A$  betyr at dataen ikke er normalfordelt. Teststatistikken  $JB$  blir definert som  $JB = \left(\frac{n}{6}\right) * \left(S^2 + \left(\frac{C^2}{4}\right)\right)$  hvor  $n$  = antall,  $S$  = skjevhet og  $C$  = kurtose. For å finne p-verdien blir kjikvadratfordeling med to frihetsgrader benyttet. Hvis p-verdien er mindre enn signifikantnivået ( $p \leq 0.050$ ) kan vi forkaste nullhypotesen og konkludere med at dataen ikke er normalfordelt (Zach, 2021). Jarque-Bera testen ble gjennomført på både 3000 m tidene og bip-test tidene. 3000 m tidene viste seg å være normalfordelt ( $p = 0.650$ ). Dette betyr at 3000 m dataen besto nullhypotesen

og er normalfordelt. Den samme testen ble gjennomført på bip-test tidene, men bip-test besto ikke nullhypotesen ( $p \leq 0.001$ ) som betyr at bip-test dataen ikke er normalfordelt.

Grunnet at bip-test ikke var normalfordelt ble det utført Spearmans  $\rho$  i tillegg til Pearsons  $r$  siden det er en ikke-parametrisk korrelasjonsanalyse. Spearmans korrelasjonskoeffisient, ofte forkortet til Spearmans  $\rho$ , er statistisk relatert til Pearsons  $r$ , men basert på verdiene til den relative rangeringen av observasjonene og ikke de observerte verdiene (Pripp, 2018). Dette betyr at Spearmans  $\rho$  er godt egnet til analyse av ikke-normalfordelte variabler som bip-test tidene. Dermed ble både Pearsons  $r$  og Spearmans  $\rho$  gjennomført for å vurdere korrelasjonen mellom 3000 m tid og bip-test tid.

For å undersøke om det var forskjell i karakter mellom bip-test og 3000 m ble Wilcoxon rangsum-test utført. Wilcoxon rangsum-test er et ikke-parametrisk alternativ til den parametriske  $t$ -testen (Scheff, 2022), og ble valgt da karakterene er på en ordinal skala (1-9) og ikke kontinuerlig. Siden datautvalget er «stort nok» (ikke en bred enighet om hva stort nok er, men minimum 16 benyttes i *Statistical Methods* av George Snedecor (2022)), ble en normalfordelingstilnærming  $N(\mu, \sigma)$  brukt, hvor  $\mu = \frac{n(n+1)}{4}$  og  $\sigma^2 = \frac{n(n+1)(2n+1)}{24}$ . I datasettet var det totalt 4 sett med like rangeringer; 87 stk 44-ranger, 35 stk 105-ranger, 9 stk 127-ranger og 1 stk 132-rang. Derfor ble en justering for antall like rangeringer gjort for et bedre estimat av variansen;  $\sigma^2 = \frac{n(n+1)(2n+1)}{24} - \frac{1}{48} \sum_t (f_t^3 - f_t)$ , hvor  $t$  varierer over settene med like rangeringer og  $f_t$  er antall ganger rangering  $t$  inntreffer. Z-skåren ble kalkulert med følgende formel  $z = \frac{|U - \mu| - 0.5}{\sigma}$ , hvor  $U$  er minste sum av rangering og  $-0.5$  utgjør korreksjon for normalfordelingstilnærming.

Den kritiske verdien ( $T_{crit}$ ) ble regnet ut ved en normalfordelingstilnærming i Microsoft Excel (Microsoft 365 Apps for enterprise, versjon 2202) og bruk av funksjonen NORM.S.INV;  $T_{crit} = \mu + \sigma \times NORM.S.INV\left(\frac{\alpha}{2}\right) - 0.5$ . Etter å ha kalkulert z-skåren,  $z = 6.47$ , kan man regne ut p-verdien gjennom *two-tail test* og funksjonen *NORM.S.FORDELING*.

## 2.4 Testprotokoller

I innsamling av data ble det benyttet to testprotokoller; 3000 m løp og bip-test. Disse utholdenhetstestene er, sammen med protokoll for styrketester, gjeldene for alle førstegangstjenestegjørende, elever og søkere ved opptak til militær utdanning og tjeneste og de fleste militært tilsatte i Forsvaret. Karaktersettingen gitt til prestasjon i utholdenhetstestene

er basert på en normbasert modell og et estimert mål av  $VO_{2maks}$ , som videre sier om man er fysisk kapabel til å jobbe i en gitt stilling i Forsvaret (Aandstad, 2016).

Alle fysiske tester i Forsvaret blir regulert av *Reglement for fysisk test* (2020). Formålet med reguleringen er å kunne fastsette gjennomføringen av de ulike fysiske testene (Forsvaret, 2020). Ved enhver gjennomføring av test skal en testleder være til stede. Testledere skal ha tilstrekkelig kunnskap om *Reglement for fysisk test* og ha mottatt opplæring i praktisk ledelse av fysisk testing. I tillegg sier reglementet at ytre forhold skal tas i betraktning før testen. Dette vil si at testene skal avlegges under tilfredsstillende klimatiske forhold. Dette innebærer blant annet temperaturgrensen som er satt til minus 15 °C og pluss 25 °C (Forsvaret, 2020).

Før testing skal det også avsettes minimum 15 minutter til oppvarming. De første ti minuttene består av generell oppvarming og de siste fem spesifikk oppvarming. Den enkelte testpliktige velger selv om ønsker å delta på en ledet oppvarming, eller gjøre det på egenhånd. For et korrekt estimat av en persons  $VO_{2maks}$  er begge testprotokollene avhengig av testpersonens motivasjon og evne til å yte maksimalt på testene.

#### 2.4.1 3000 m

I *Reglement for fysisk test* blir 3000 m løp vurdert å gi et valid mål på  $VO_{2maks}$  uttrykt per kg kroppsvekt så lenge testpersonen yter maksimalt under løpet (Forsvaret, 2020). Det er spesielt viktig at testpersonen yter maksimalt under hele 3000 m løpet. Siden man løper over en lengre avstand vil det ikke være testledere til stede for å kunne gi signal på hvor langt eller hvor fort man løper som ved en bip-test hvor man følger lydsignalene hele tiden. Reglementet sier videre at prestasjonen til testpersonene skal ikke bli vesentlig påvirket av vær og føreforhold og at de klarer å disponere løpet (Forsvaret, 2020).

Ved punkt 6.2.1 i *Reglement for fysisk test* blir det videre spesifisert krav til gjennomføring av 3000 m løp. Løypen skal bestå av enten vei, bred sti eller friidrettsbane og det skal være maksimalt ti meter i høydeforskjell mellom løypas høyeste og laveste punkt (Forsvaret, 2020). Løs grus, gress og is sees på som et ikke akseptabelt underlag til testing. I tillegg skal løypen være kontrollmålt for lengde, med en feilmargen på maksimalt  $\pm 10$  m.

3000 m skal ha start og mål på tilnærmet samme sted. Dette er for å sikre at eventuelle høydeforskjeller, mot- eller medvind blir utlignet underveis (Forsvaret, 2020). Løypa som benyttes ved opptaket til LFS er en rett og lang asfaltvei. Den strekker seg 1500 m, slik at man snur i enden og løper de siste 1500 m på vei tilbake.

## 2.4.2 Bip-test

*Reglement for fysisk test* spesifiserer også krav til gjennomføring av Bip-test. Under punkt 6.2.2 blir det spesifisert krav til underlag, organisering og utførelse (Forsvaret, 2020). Testen skal gjennomføres innendørs og underlaget skal bestå av parkett, gummi/syntetisk sportsgulv eller lignende underlag. Gulvet skal gi en god friksjon slik at vendinger er mulig uten å skli. De som gjennomfører testen, skal løpe mellom to vendestrekere som er tydelig oppmerket og ha 20 m mellomrom. Testleder vil sette i gang testen ved bruk av en utlevert og godkjent bip-test CD, eller godkjent app på telefonen, som inneholder lydsignaler. Deltagerne vil dermed starte å løpe etter instruksjon fra lydfilen og tilpasse hastigheten basert på lydsignalene («bip'ene»). De vil løpe frem og tilbake mellom vendestrekene og snu ved lydsignalet. Testen avsluttes når deltagerne ikke lenger greier å følge lydsignalene.

## 2.5 Databehandling og prosedyre

I datasettet jeg mottok var det  $n=793$  subjekter som ble opprinnelig inkludert. Etter eksklusjonsprosessen beskrevet i kapittel 2.2 ble test-data tilhørende  $n=264$  subjekter tatt videre til analyse. All data som ble uthentet var anonymisert. Ved godkjenning og hjelp fra seleksjonssjefen ved LRSS ble dataen overført til et Excel-dokument og deretter sendt via militær gradert e-post til meg. All databehandling i etterkant har blitt utført gjennom Microsoft Excel i Office 365.

## 2.6 Validitet og Reliabilitet

### 2.6.1 Validitet

Validitet går ut på om en test klarer å måle det den er designet til å måle. Det vil si at resultatet samsvarer med den egenskapen vi ønsker å få vurdert (Kirknes, Aandstad, & Stornæs, 2014). Innenfor begrepet validitet finnes det ulike kategorier. I denne oppgaven vil vi kun benytte oss av kriterievaliditet og dermed kun gå inn på det. Kriterievaliditet handler om hvorvidt testresultatet samsvarer med et «gullstandardmål» (Kirknes, Aandstad, & Stornæs, 2014). Når man sammenligner eksempelvis 3000 m tid og  $VO_{2maks}$  sjekker man kriterievaliditeten ved å sammenligne den direkte målte  $VO_{2maks}$  og den estimerte fra 3000 m testen. Hvis det er et godt samsvar, gir testen et valid mål på  $VO_{2maks}$ . Kriterievaliditet blir dermed også kalt for prediktiv validitet, ettersom at en enklere test prøver å predikere noe som er vanskeligere å måle eller gjennomføre.

3000 m test har vist seg, fra tidligere studier, å ha en høy validitet sammen med flere lignende utholdenhetstester (Kirknes, Aandstad, & Stornæs, 2014, ss. 57-60). I Aandstad sin undersøkelse av estimert  $VO_{2maks}$  ved 3000 m hos kvinner og menn i Forsvaret, fikk han en

korrelasjonskoeffisient på  $r = 0.74$  hos menn og  $r = 0.79$  hos kvinner mellom 3000 m prestasjon og målt  $VO_{2maks}$  (Aandstad, 2021). I tillegg fant han ingen signifikant forskjell mellom målt og predikert  $VO_{2maks}$  for menn ( $p = 0.98$ ) eller kvinner ( $p = 1.00$ ).

I Aandstad sin undersøkelse *Validity and Reliability of the 20 Meter Shuttle Run Test* viser bip-test til å ha en lavere validitet ved individuell måling opp mot direktemålt  $VO_{2maks}$  (2011, s. 517). Den estimerte  $VO_{2maks}$  varierer svært mye hos den enkelte subjekten fra ulike studier. Eksempelvis ville en kadett med målt  $VO_{2maks}$  på  $56 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ , ha en statistisk estimert  $VO_{2maks}$  alt imellom  $49.5$  og  $62 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . Et avvik på  $12.5 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$  blir relativt stort, og dermed anbefales det ikke å benytte bip-test som en prediktiv test opp mot  $VO_{2maks}$ . Allikevel regnes bip-test på gruppenivå å ha en relativt god validitet (Kirknes, Aandstad, & Stornæs, 2014).

### 2.6.2 Reliabilitet

Reliabilitet betyr i hvilken grad testen eller instrumentet gir samme resultat ved repeterte målinger (Kirknes, Aandstad, & Stornæs, 2014). Det kan være flere grunner til hvorfor en test utgir ulike resultat for hver gjennomføring. Ved fysisk testing som 3000 m kan biologiske variasjoner som dagsform eller sykdom spille inn. Hvis testen foregår ute, vil vær og føre kunne påvirke resultatene. Er testen mer teknisk vanskelig kan det hende at flere får bedre resultat på andre gjennomføring på grunn av læringseffekten (Kirknes, Aandstad, & Stornæs, 2014). Alle tester har med andre ord noen feilmarginer, og 100 % reliabilitet vil sjeldent kunne oppnås. Likevel er reliabilitet veldig viktig for å kunne si om testen er god. Hvis testen har lav reliabilitet vil ikke testen være valid (Kirknes, Aandstad, & Stornæs, 2014).

3000 m test har ved tidligere studier vist seg å ha en høy reliabilitet, sammen med flere lignende utholdenhetstester (12 minutters løp, 1.5 mile- eller 2 mile run) (Aandstad, 2021). Allikevel har 3000 m og lignende tester et par fallgruver som må tas hensyn til. Den første er at soldatene må være motivert til å presse seg for at prestasjonen skal gi et riktig bilde av soldatens  $VO_{2maks}$  (Kirknes, Aandstad, & Stornæs, 2014, s. 59). Dette er sjeldent et problem ved opptaksprøver som seleksjon til flyskolen. Her er søkerne motivert til å prestere og avhengig av å konkurrere seg inn. Den andre fallgruven er at testen gjennomføres ute hvor vær, føre og terreng kan være ulikt mellom test og retest. *Reglement for fysisk test* er skrevet for å kunne standardisere slike ulikheter. Med dette som bakgrunn er det grunn til å anta at 3000 m test har en høy reliabilitet under denne undersøkelsen.

Bip-test har også vist seg å ha en svært høy reliabilitet hos militært personell (Aandstad, Berntsen, Holme, & Anderssen, 2011). I studien til Aandstad & Berntsen ble 38 heimevernssoldater kryssvalidert med 28 luftkrigsskole kadetter for å sjekke reliabiliteten og validiteten til bip-test (2011). Både intraklassekorrelasjonskoeffisienten (ICC) og korrelasjonskoeffisient var mellom 0.95 og 0.96 for test-retest uansett om prestasjonen ble målt i antall gjennomførte shuttle, nivåer eller estimert  $VO_{2maks}$  (Aandstad, Berntsen, Holme, & Anderssen, 2011).

## 3 Teori

### 3.1 Utholdenhet og kondisjon

Kirknes et al (2014) beskriver at det finnes flere begreper som kan benyttes for å beskrive evnen og/eller forutsetningen for å utføre arbeid som 3000 m eller bip-test. Eksempler på slike begreper er: aerob kapasitet, aerob utholdenhet, utholdenhet, kondisjon og  $VO_{2maks}$ . Begrepet utholdenhet er tradisjonelt bruk i Forsvaret og kan defineres som organismens evne til å arbeide med relativt høy intensitet over lengre tid (> 4-5 minutter) (Kirknes, Aandstad, & Stornæs, 2014). Kirknes et al nevner også at utholdenhet favner videre enn begrepet kondisjon. Kondisjon blir gjerne sett synonymt med  $VO_{2maks}$ , mens utholdenhet inkluderer flere av faktorene som bestemmer prestasjonsevnen i et lengre varig arbeid med relativt høy intensitet (Kirknes, Aandstad, & Stornæs, 2014, s. 57). Faktorene består blant annet av  $VO_{2maks}$ , utnyttingsgrad, teknikk (arbeidsøkonomi) og motivasjon. Ved å gå ut ifra denne definisjonen kan man si at tester som 3000 m og bip-test bør kalles for utholdenhetstester og ikke kondisjonstester, ettersom at testresultatene ikke bare er avhengig av  $VO_{2maks}$ . Reliabiliteten til 3000 m er blant annet avhengig av løperens motivasjon og evne til å presse seg under testen, som viser til at det ikke bare er  $VO_{2maks}$  som avgjør løperens resultat.

#### 3.1.1 $VO_{2maks}$

$VO_{2maks}$ , også kjent som maksimalt oksygenopptak eller maksimal aerob kapasitet, er kroppens maksimale evne til å ta opp og utnytte oksygen under fysisk arbeid (Athlepedia, 2022).  $VO_{2maks}$  er anerkjent standardmål på kardiorespiratorisk kondisjon og blir ofte uttrykt i milliliter per kilo kroppsmasse per minutt (Athlepedia, 2022). Fra hvile vil oksygenopptaket øke i takt med en økning i arbeidsintensitet (f.eks. økning i hastighet på tredemølle), til et punkt hvor en videre økning i arbeidsintensitet ikke blir møtt med en videre økning i oksygenopptak. Det er ved det punktet, når oksygenopptaket stagnerer til tross for økende intensitet, at  $VO_{2maks}$  blir identifisert.

I idretter som løping, sykling og langrenn der man bruker store deler av kroppen, og er aktiv i mer enn 30 minutter, vil minst 95 % av energien komme fra aerob energiomsetting (Olympiatoppen, 2022). Aerob kapasitet vil av den grunn være den viktigste faktoren for å holde en høy gjennomsnittsfart gjennom aktiviteten. 3000 m og bip-test varer vanligvis bare rundt 12 minutter (ikke 30), allikevel vil ca. 85-90 % av energien komme fra aerobe prosesser ifølge Kirknes et al (2014, s. 59). 3000 m og bip-test kan ikke sidestilles helt derimot. 3000 m kan defineres som en *steady-state* hvor løperen vil forsøke å holde en konstant fart, mens bip-test kan defineres som *incremental* ettersom man stadig øker hastigheten i løpet av testen. Uavhengig om det er en *steady-state* eller *incremental* test, vil maksimal aerob kapasitet eller  $VO_{2maks}$  fortsatt være den viktigste faktoren. Allikevel er det flere faktorer som påvirker prestasjonen ved utholdenhetstester som 3000 m og bip-test. Mental robusthet (her relatert til evnen å tolerere ubehag og smerte), løpsøkonomi (hvor effektivt man løper) og laktatterskel (raskeste man kan løpe uten å akkumulere store mengder melkesyre i blodet) er faktorer utenom  $VO_{2maks}$  som påvirker subjektets prestasjon (Athlepedia, 2022). Dermed er  $VO_{2maks}$  i en ellers homogen gruppe mer en indikator på en persons aerob potensial enn noe som forutser suksess i utholdenhetstester. Hvis det derimot er store forskjeller mellom to subjekters  $VO_{2maks}$  (i.e. heterogen gruppe) kan man relativt trygt forutse hvem som kommer til å komme i mål først.

Å måle  $VO_{2maks}$  nøyaktig krever en maksimal fysisk belastning av et subjekt (fortrinnsvis utført på tredemølle, alternativt ergometersykel) med en streng protokoll mens trent personell benytter åpen krets spirometri (Athlepedia, 2022). Protokollene består av trinnvise økninger i intensitet og en samling og måling av volum og oksygenkonsentrasjonen ved inspirasjon og ekspirasjon. Differansen mellom  $O_2$  inspirert og ekspirert utgjør  $VO_2$ , altså hvor mye oksygen subjektet benytter seg av. Testen avsluttes når oksygenopptaket stagnerer og ikke lenger øker med farten og intensiteten.

### 3.1.2 Estimat av $VO_{2maks}$

Det er svært upraktisk og ressurskrevende å ta direkte mål av  $VO_{2maks}$  av store grupper som soldater. På 1960-tallet påpekte Kenneth Cooper, en amerikansk major, viktigheten av å måle  $VO_{2maks}$  på soldater, men siden direkte måling ikke var praktisk mulig utviklet han en indirekte test som kunne predikere  $VO_{2maks}$  (Kirknes, Aandstad, & Stornæs, 2014). Denne testen er kjent som Cooper-test og går ut på å løpe så langt man klarer på 12 minutter. Tilbakelagt distanse har vist å ha en svært god sammenheng med direkte målt  $VO_{2maks}$  ( $r =$



0.90) (Kirknes, Aandstad, & Stornæs, 2014). Dette muliggjorde en enkel utholdenhetstest for store grupper mennesker som også kan gi et indirekte mål på  $VO_{2maks}$ .

I dag er det mange land som har gått over til andre utholdenhetstester for militært personell. Dagens tester tar gjerne utgangspunkt i måling av tid heller enn distanse. Dette forenkler administrasjonen ved gjennomføringen da alle subjekter er ferdig på samme sted. Det norske Forsvaret har benyttet 3000 m testen siden slutten av 1970-tallet og i 2017 ble bip-test sidestilt som et gyldig alternativ til 3000 m.

### 3.1.3 Arbeidskravsanalyse og krav til $VO_{2maks}$

Bakgrunnen for at soldater og personell i Forsvaret gjennomgår fysiske tester er for å sikre at de har nødvendig fysisk form for hva stillingen deres i Forsvaret krever. Kravsettingen til ulike stillinger i Forsvaret er basert på en arbeidskravsanalyse på hvilke fysiske krav tjenesten stiller. Arbeidskravsanalyser inkluderer vanligvis flere beskrivelser: type aktivitet (for eksempel forflytning til fots), frekvens (hvor mange ganger i uken), varighet (hvor mange minutter/timer aktiviteten pågår), intensitet (f.eks. hastighet på forflytning), forhold mellom arbeid og hvile (antall pauser eller intervaller), teknologiske hjelpemidler tilgjengelig og rammefaktorer jobben utføres under (temperatur, terreng, bekledning etc.) (Kirknes, Aandstad, & Stornæs, 2014, s. 29). I tillegg vil arbeidskravsanalysen beskrive kravet til fysisk kapasitet som er nødvendig for å gjøre jobben (Kirknes, Aandstad, & Stornæs, 2014). For eksempel så kan et arbeidskrav for en soldat være å gå med tung pakning eller å grave stillinger. Gjennom arbeidskravsanalysen kan det komme frem at det å grave en stilling vil tilsvare en  $VO_2$  på  $20 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . For at soldaten skal klare å gjennomføre denne aktiviteten en hel arbeidsdag er det vist at intensiteten ikke bør være høyere enn ca. 40 % av maksimal aerob kapasitet (Kirknes, Aandstad, & Stornæs, 2014). Det vil si at en soldat som har et arbeidskrav om å grave stillinger hele dagen bør ha en  $VO_{2maks}$  på minst  $50 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . Ut ifra arbeidskravsanalysen blir minimumskravene for Forsvarets opptak og seleksjon satt. Hver Driftsenhet i Forsvaret (DIF) har satt minimumskrav (MK) for hver tjenestestilling knyttet til sin DIF (Aandstad, 2016). Disse minimumskravene, sammen med krav til  $VO_2$ , har vært bakgrunnen for karakterskalaen til utholdenhetstesten som 3000 m og bip-test.

## 4 Seleksjon til Luftforsvarets flygerskole

### 4.1 Generelt om opptak

For å kunne bli flyger eller navigatør i Forsvaret må man først gjennom en seleksjonsprosess. Seleksjonen varer omtrent ett år og består av ulike faser. Etter seleksjonen vil man begynne i et utdanningsløp som befinner seg i både Norge og USA. Til sammen vil man bruke omtrent 4-5 år fra man søker til man begynner på arbeidsplassen. På veien til å bli flyger eller navigatør vil man få grunnleggende offisersutdanning (GOU) med bachelorgrad innen luftmakt og luftmilitær ledelse, 150-200 timer flytid, en rekke kurs, opplevelser og personlig utvikling som oppstår under utdanning, og ansettelse som fenrik (OF-1) (Forsvaret, 2022).

### 4.2 Bakgrunn for seleksjon

Da brødrene Wright skulle fly for første gang kastet de kron og mynt om hvem som skulle fly den første turen (Martinussen, 2005). Siden den gang har det gått 120 år, og mye har endret seg i utvelgelsen av flygere. I dag benyttes det blant annet moderne teknologi og databaserte tester for å velge ut de som skal fly i Luftforsvaret. Veien til seleksjonsmetoden Luftforsvaret benytter seg av i dag har vært lang og er i kontinuerlig endring (Bertelsen, 2022). Det finnes antageligvis ikke en yrkesgruppe som har blitt utsatt for mer forskning, undersøkelser og testing enn flygere (Harlem, 2016). Bakgrunnen for all seleksjonen er av sikkerhetsårsaker, men også økonomiske da både materiellet flygerne opererer og utdannelsen til flygerne er ekstremt kostbart (Martinussen, 2005). Regjeringen beregnet at det koster 71 milliarder 2017-kroner å anskaffe våre F-35 jagerfly, som betyr at det koster over én milliard per fly (Regjeringen, 2017). Jagerflygerne skal fly og operere flymaskinen i over 1500 km/t, levere svært avanserte våpen og trene på å ta liv (Lockheed Martin, 2022). Det er et enormt ansvar å ha, slik at ikke hvem som helst kan fly et slikt fly. Ved nøyere utvelgelse av flygere vil man kunne redusere risikoen for uhell og bedre sikkerheten ved flyging.

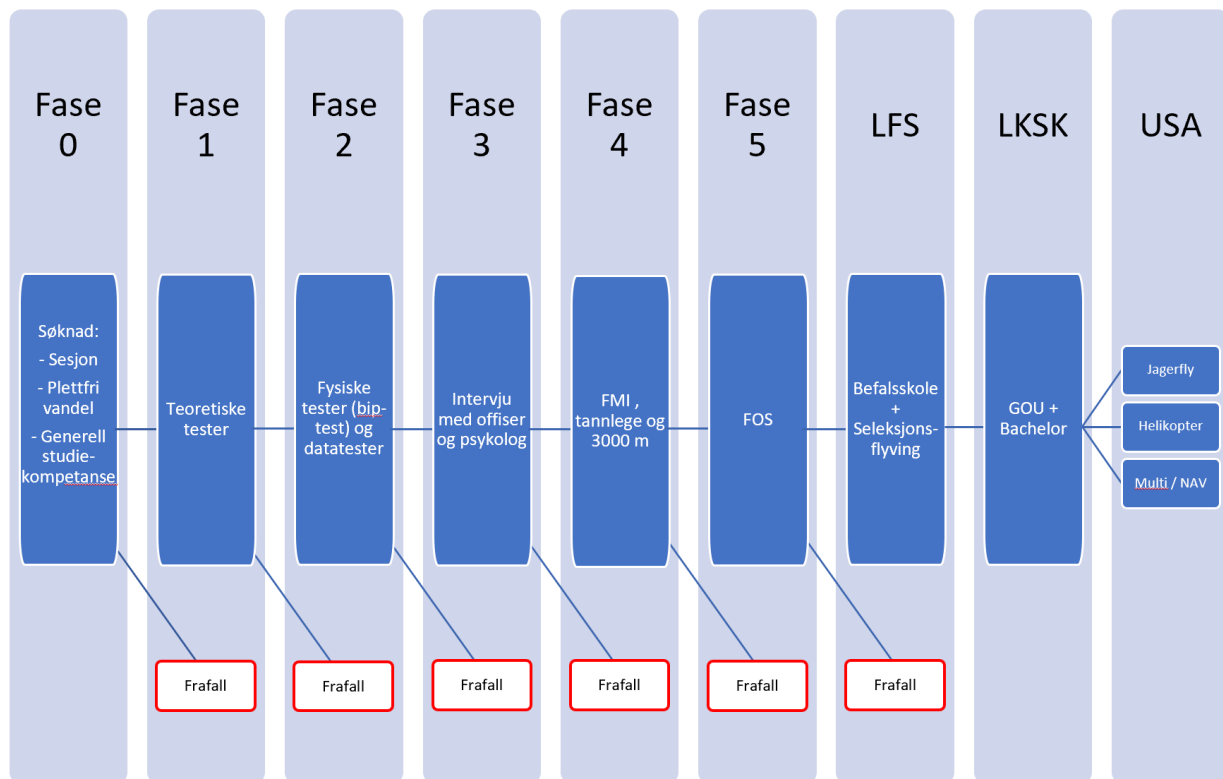
### 4.3 Seleksjonsmodell

Seleksjon for flygere og navigatører i Luftforsvaret er en grundig prosess.

Seleksjonsprosessens form og farge endrer seg årlig for å forbedre og effektivisere prosessen mest mulig, men hovedelementene og metoden er likt hvert år (P.W. Bertelsen, personlig kommunikasjon, 16. juni, 2022). Figur 2 illustrerer alle fasene hver søker må igjennom for å bli flyger eller navigatør fra og med 2022.

*Figur 2: Seleksjonsmodell*

Modell av hvordan seleksjonen av flygere og navigatører blir gjennomført. Søkerne til LFS må gjennom fem faser med seleksjon før seleksjonsflygning og utdanning begynner. Mørkeblå bokser viser til hva som blir gjennomført ved hver fase, mens hvite bokser indikerer hvor søkerne kan ryke ut fra seleksjonen.



LFS; Luftforsvarets Flygerskole, LKSK; Luftkrigsskolen, FMI; Flymedisinsk Institutt, FOS; Forsvarets Opptak og Seleksjon, GOU; Grunnleggende Offisersutdanning, Multi; Flermotor flytyper (C130-J Hercules, P-8 Poseidon), NAV; Navigatør

#### 4.4 Seleksjonen i praksis

Hvert år tar Forsvaret inn nye søkere til Luftforsvarets flygerskole. Søknaden leveres inn i løpet av høsten det året man søker, og fristen utgår som oftest første desember samme år (Forsvaret, 2022). Av de som søker og kommer seg videre er det totalt fem faser som kandidatene må gjennomgå før de begynner seleksjonsflygningen på LFS. De fem fasene skjer i løpet av det første halvåret (januar-juli) og inneholder ulike tester.

##### 4.4.1 Fysiske tester på LRSS

I løpet av seleksjonen vil det være fysiske tester for kandidatene som søker seg inn. På fase 2 vil søkerne gjennomgå styrketester (medisinball støt, stille lengde, pull-up) og én utholdenhetstest (3000 m eller bip-test). Utholdenhetstesten som ikke ble gjennomført ved fase 2, vanligvis 3000 m, vil gjennomføres i fase 4 for søkerne som kommer seg videre dit. Alle søkere må oppnå MK 5 for å bestå de fysiske testene (se tabell 1b for minstekrav), men hvis en kandidat ikke greier å oppnå de fysiske minstekravene, vil han/hun bli fjernet fra seleksjonsprosessen det året. Søkeren vil få som regel mulighet til å kunne prøve på nytt året

etter, så lenge søkeren fortsatt oppfyller de andre seleksjonskravene (Bertelsen, 2022). Karakterkravet for å komme videre til LFS ble endret i 2022 for å kunne mulig øke antall kandidater (P.W. Bertelsen, personlig kommunikasjon, 26. juli, 2022). Kravet ble senket fra MK 6 til MK 5 ved seleksjonsprosessen, men Bertelsen mente at kravet ved LFS var fortsatt MK 6 og var urørt. I tillegg, for de som allerede er flygere ved avdeling er kravet noe annerledes. Jagerflygere skiller seg ut med å ha MK 7, mens resterende har MK 6 eller MK 5 (P.W. Bertelsen, personlig kommunikasjon, 26. juli, 2022).

## 5 Resultater

### 5.1 Deskriptiv analyse

For å lettere se sammenhenger og forskjeller i resultat fra 3000 m og bip-test ble det gjennomført en deskriptiv analyse. Resultatene fra analysen i tabell 2 viser at subjektene løper i gjennomsnitt ett minutt lengre og over én kilometer i timen raskere på 3000 m, men får allikevel i snitt 0.4 dårligere i karakter.

*Tabell 2: Gjennomsnitt +/- SD (min-maks) fra 3000 m og bip-test (n=264)*

<b>Variabler</b>	<b>3000 m</b>	<b>Bip test</b>
Tid (min)	12.2 ± 0.9 (9.6–14.3)	11.3 ± 1.2 (8.9–14.3)
Nivå : Shuttle		11:3 ± 1:2 (9:5–15:1)
Fart (v)*	14.9 ± 1.1 (12.5–18.6)	13.6 ± 0.6 (12.5–15.5)
Karakter**	7.8 ± 1.4 (4.0–9.0)	8.2 ± 1.0 (5.0–9.0)

\* *Fart (v) for 3000 m er gjennomsnittlig løpehastighet i km/t, mens fart (v) for bip test er gjennomsnittlig maksimal hastighet oppnådd i km/t.*

\*\* *Karakter 1 og 9 er henholdsvis den dårligste og beste mulige karakteren*

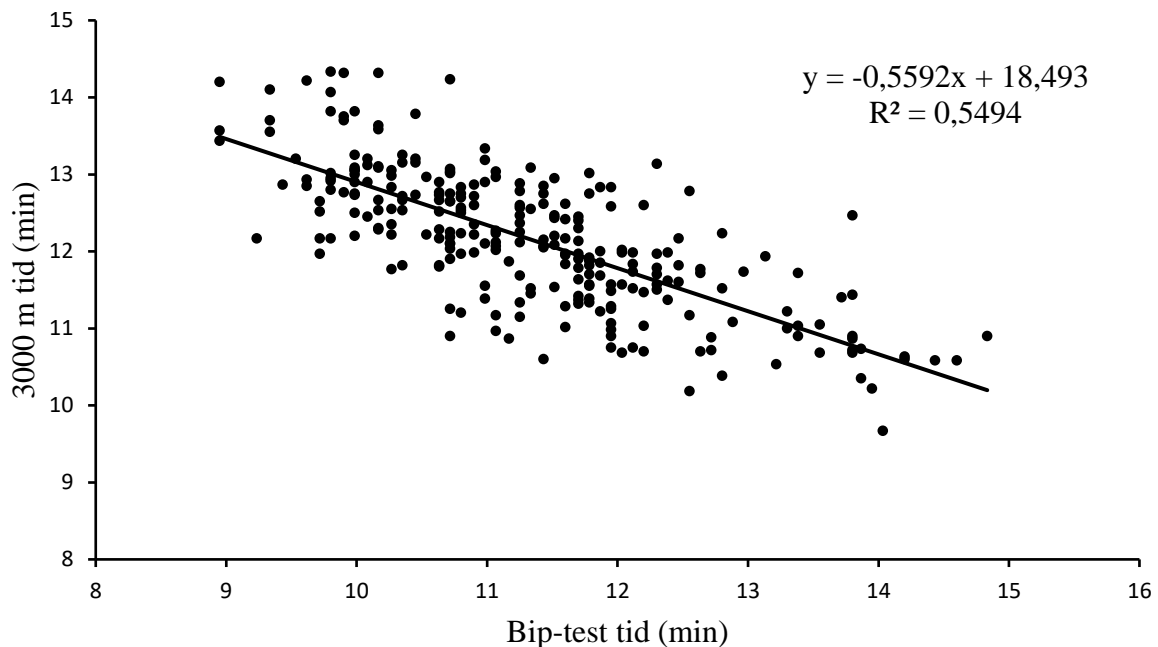
## 5.2 Korrelasjonsanalyse

Pearsons korrelasjonskoeffisient ( $r$ ) ble benyttet til å vurdere forholdet mellom tid brukt på 3000 m og bip-test hos unge menn på opptak til Luftforsvarets Flygerskole ( $n=264$ ).

Innledende analyser viste at 3000 m tidene var normalfordelt, mens bip-test var ikke normalfordelt. Dette ble testet gjennom Jarque-Beras test ( $p \leq 0.05$ ) og alle ekstremverdier ble fjernet ved utvalget i forkant av analysen. Siden bip-test tidene ikke var normalfordelt ble Spearmans  $\rho$  kjørt i tillegg og sammenlignet med Pearsons  $r$ .

Det var en svært sterk negativ korrelasjon mellom tid brukt på 3000 m løp og bip-test, uavhengig av om Pearsons  $r$  ( $r = -0.74$ ,  $p < 0.001$ ) eller Spearmans  $\rho$  ( $r = -0.73$ ,  $p < 0.001$ ) ble brukt. Resultatet på bip-test kunne forklare 55 % av variasjonen i prestasjon på 3000 m.

*Figur 3: Sammenheng mellom 3000 m tid og bip-test tid i minutter*



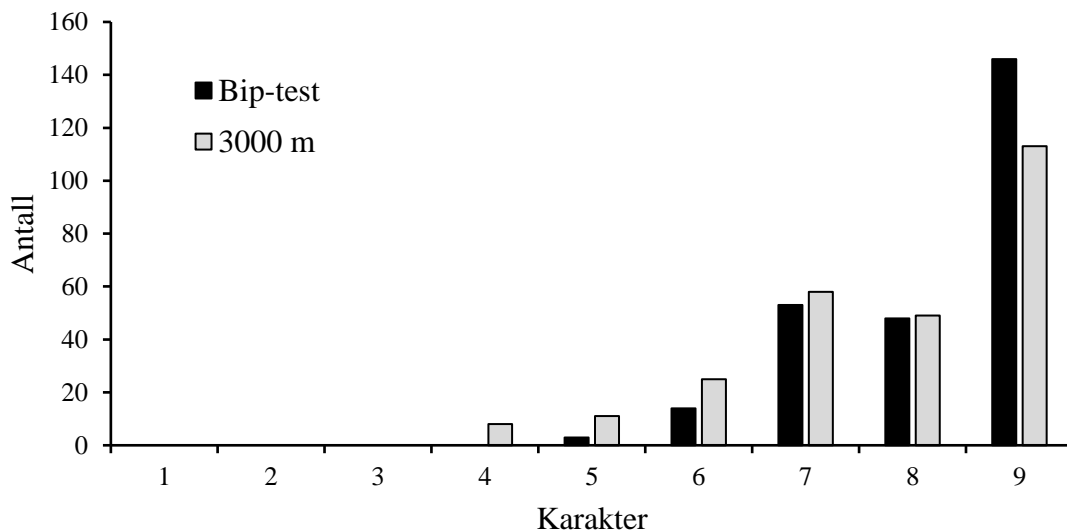
### 5.3 Karakterer analyse

Bip-test er en nyere form for utholdenhetstest i Forsvaret som ikke har blitt forsket like mye på som 3000 m, og karakterene fra bip-test og 3000m ble sammenlignet med en Wilcoxon rangsum-test for å undersøke eventuelle forskjeller i karakter. Median og 1.-3. kvartil for var 9 (7-9) for bip-test og 8 (7-9) for 3000 m. Fordelingen av oppnådd karakter i de to utholdenhetstestene er illustrert i Figur 4, mens differansen er illustrert i Figur 5.

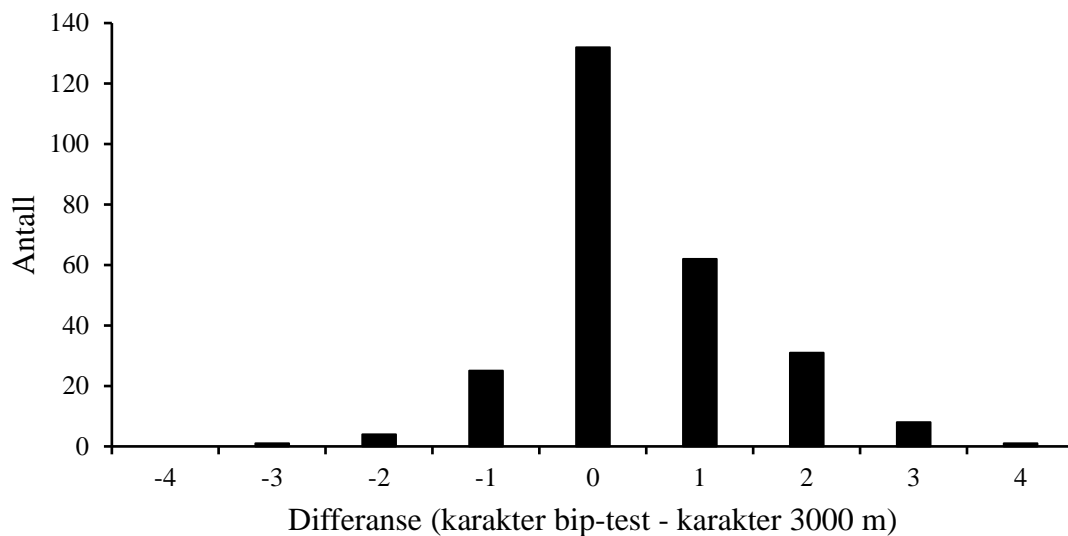
Differanseskårene var ikke symmetrisk fordelt og hadde en overvekt med positive ranger, som vil si bip-test-karakter generelt var bedre enn 3000m-karakter. Differanseskårene ble vurdert ved et histogram med en normal kurve lagt over.

Av de 264 deltakerne som ble analysert, fremkalte bip-test en økning i karakter hos 102 deltakere (39%) sammenlignet med 3000 m, mens 132 deltakere (50%) ikke så noen forbedring og 30 deltakere (11 %) fikk dårligere karakter på bip-test enn 3000 m. Subjektene oppnådde statistisk signifikant bedre karakter i bip-test (*median* = 9) enn 3000 m (*median* = 8) ( $z= 6,47, p < 0,001$ ).

Figur 4: Søylediagram over karakterer oppnådd i utholdenhetstestene 3000 m og bip-test



Figur 5: Søylediagram over differanse i karakterer mellom bip-test og 3000 m



## 6 Diskusjon

Testresultatene til 264 unge menn ved opptak til Luftforsvarets Flygerskole ble analysert for sammenligning av 3000 meter test og bip-test. Ved bruk av korrelasjonsanalyse og Wilcoxon rangsum-test ble testresultatene analysert for å vurdere om det er en sammenheng i prestasjon mellom bip-test og 3000 m, og om det eventuelt er en forskjell i oppnådd karakter mellom de to utholdenhetstestene.

Resultatene viser at det er en tydelig sammenheng mellom prestasjon i utholdenhetstestene, samtidig som at det også er forskjell i oppnådd karakter ved Forsvarets karakterskala.

### 6.1 Hovedproblemstilling

Det ble funnet en svært sterk negativ korrelasjon mellom tid på 3000m og bip-test. Ved maksimal ytelse er mesteparten av energifrigjøringen (85-90%) aerob for begge testprotokoller (Kirknes, Aandstad, & Stornæs, 2014). Siden begge tester også har en god validitet opp mot  $VO_{2maks}$  ( $r_{3000m} = 0.74$ ,  $r_{bip} = 0.82$ ) var det som forventet at testene ville korrelere godt. Dette støtter også oppunder tidligere studier på det norske forsvaret (Aandstad & Berntsen, 2011; Aandstad, 2021), samt studier fra det amerikanske og britiske forsvaret på lignende testprotokoller.

Canino et al. gjennomførte i 2018 en studie på soldater i U.S. Army hvor resultater fra bip-test og 2 mile run (3.2 km; svært likt 3000 m) ble sammenlignet hos 531 soldater (403 menn og 128 kvinner) (Canino, et al., 2018). Alderen til soldatene var  $24 \pm 4$  år, som er noe eldre enn aldersspennet i denne oppgaven, men vil trolig utgjøre lite til ingen forskjell. Pearsons korrelasjonskoeffisient og en regresjonsanalyse ble gjennomført for å se forholdet mellom tid

og antall bip-test nivå gjennomført. Det ble funnet en signifikant, negativ korrelasjon mellom tid og nivå gjennomført ( $r = -0.75$ ,  $p < 0.001$ ).

Wilkinson et al. (2014) gjennomførte også en korrelasjonsstudie på soldater. The British Army ble testet i bip-test og 2.4 km run (noe kortere enn 3000 m) hos  $n=156$  soldater (100 menn og 56 kvinner). Alderen til soldatene var  $21 \pm 3$  år, som er svært likt aldersspennet i denne oppgaven. Ved korrelasjonsanalyse ble det funnet en signifikant, positiv korrelasjon mellom gjennomsnittlig løpehastighet på 2.4 km run og antall shuttle gjennomført på bip-test ( $r = 0.91$ ,  $p < 0.01$ ).

I begge studiene ble det funnet en sterk korrelasjon mellom bip-test og en tilsvarende 3000 m test. Resultatene fra denne oppgaven var helt lik som Casino et al. ( $-0.75$ ), men noe lavere enn Wilkinson et al. ( $0.91$ ). Canino et al. benyttet 2 mile run (3.2 km) som er mer lik 3000 m enn Wilkinson et al. sin 2.4 km run er, noe som kan forklare like resultater. Alle resultater er bygd på solide utvalg; 531 (403 menn), 156 (100 menn) og 264 ved oppgaven min. Distansen på testprotokollene er noe ulikt, men alle er allikevel lange nok for at aerob kapasitet er den viktigste bidragsyteren. Det er vanskelig å avgjøre hva som var årsaken til ulik korrelasjon mellom studiene, men uavhengig av årsak støtter begge studier resultatene mine fra denne oppgave. Det er lignende korrelasjonskoeffisienter mellom bip-test og tilsvarende 3000m løp hos soldater både i US Army og The British Army.

En viktig forskjell mellom studiene til Canino et al. og Wilkinson et al. er at de begge inneholder kvinnelig subjekter, noe som min oppgave ikke gjør. Dette vil påvirke resultatene og kan skape feilkilder ved sammenligning. På gruppenivå er det fysiologiske forskjeller mellom menn og kvinner som gir utslag på fysisk prestasjonsevne. Dette gjør at man kan forvente på normalfordelte utvalg at kvinner er i nedre sjikt av prestasjon, mens menn i øvre. Potensielt kan dette skape to ulike prestasjonsgrupper i utvalget som først og fremst har med kjønn å gjøre, og ikke eksempelvis treningsmengde. Ytterste konsekvens er at man får en stor spredning mellom disse gruppene innad i utvalget man analyserer, noe som ved korrelasjonsanalyse blir en feilkilde da det kan forsterke korrelasjonskoeffisienten. Selv om denne forskjellen vil sannsynligvis påvirke resultatene, er det fortsatt et betydelig flertall av menn i begge studier som gjør de relevante. I tillegg er det få studier som er så lik min oppgaven som Canino et al. og Wilkinson et al. sin.



## 6.2 Underproblemstilling

Subjektene oppnådde en statistisk signifikant høyere karakter ved bip-test enn 3000 m. Dette var forventet basert på egen erfaring, men delvis uforventet basert på tidligere studier og litteratur.

Minstekravskalaen, som Forsvaret bruker som karaktersystem (1-9), er bygget på en normbasert modell fra prestasjonen til mannlige og kvinnelige soldater ved innrykk til rekruttskole i perioden 2010-2014 (Aandstad, 2016). Denne gruppen ble valgt grunnet de er mindre selektert på fysisk form enn det militært tilsatte og søkere til GBU og GOU er. I tillegg gjør det lettere å vurdere hvor mange prosent av normalbefolkningen som klarer å kvalifisere seg for videre karriere i Forsvaret (Aandstad, 2016). Det ble bestemt at ca. 20 % av mannlige rekrutter skal klare MK 9, og at 95 % skal bestå MK 1. MK 8 skal 30 % bestå, MK 7 skal 40 % bestå og videre helt til MK 2 hvor 90 % skal bestå. Kravene blir også delvis justert for kjønn og alder, men dette er ikke relevant for problemstillingen ettersom at utvalget mitt består av kun menn under 29 år.

For 3000 m ble MK-skalaen utviklet ut ifra resultatene til rekruttene ved innrykk til GSU 2010-2014, hvor datagrunnlaget besto av totalt 29 025 menn og 3 678 kvinner (Aandstad, 2016). Den 20. prosentilet hos de mannlige rekruttene var 12:08 på 3000m. MK 9 ble dermed justert til 12:00 (for å forenkle testkravene og gjøre de mer logiske) og hvert 10. prosentil nedover skalaen økte ca. med 30 sekunder (Tabell 1a/1b).

Bip-test var derimot ikke et gyldig alternativ eller praktisert i Forsvaret før 2017, slik at MK-skalaen til bip-test kunne ikke bli utviklet på samme måte som for 3000 m. I Aandstad sin rapport om fastsettelse av det nye MK-skalaen skrev han:

*«For bip-testen har vi beregnet MK med utgangspunkt i antatt sammenheng mellom bip-test prestasjon og 3000 m løpsprestasjon (...) Det skal da i prinsippet være like vanskelig å bestå et bestemt MK, uavhengig om man gjennomfører bip-test eller 3000 m løp. (Aandstad, Minimumskravskalaer for fysiske tester i Forsvaret, 2016, s. 7)*

Videre blir det beskrevet at bip-test er validert opp mot direktemålt  $VO_{2maks}$  fra flere studier. MK-skalaen til bip-test baserer seg på er Aandstad et al. sitt studie hvor forfatterne konkluderer med at prediksjonslikningen estimerer en  $VO_{2maks}$  på en god måte (2011). Det utviklingsarbeidet er basert på 38 mannlige HV-soldater ( $35 \pm 4$  år), og er kryssvalidert med 28 mannlige kadetter fra Luftkrigsskolen ( $23 \pm 4$  år). Aandstad utviklet videre tre ulike

prediksjonsligninger fra studiet for å skape en bredere dekning på MK-skalaen uavhengig av god eller dårlig prestasjon på bip-test (én for de som løper under 40 shuttles, én for de som løper mellom 40 og 80 og til slutt én for de som løper over 80 shuttles) (Aandstad, 2016). Prediksjonsligningen for de som løper over 80 shuttles er kun bygd på resultatene fra de 26 mannlige kadetter som ble målt i 2006. 80 shuttles tilsvarende bip-test-resultat 9:8 eller karakter 6 på dagens MK-skala. Slik at 261 (99 %) av subjektene i denne oppgaven baserer seg på prediksjonsligningen for over 80 shuttles.

Ettersom at tilnærmet 40 % av subjektene fra LFS-opptaket hadde høyere karakter i bip-test enn 3000 m er det rimelig å anta at det ikke er like vanskelig å oppnå MK 9 på bip-test som 3000 m (ref. direkte sitat fra Aandstad). En mulig grunn til dette er at prediksjonsligningen som bestemmer de øverste minstekravene (MK 6-9) er bygd på en begrenset mengde med data (n=28) og har dermed et svakere grunnlag til å gi riktige MK enn det 3000 m har med sin datamengde (n=32 703). I tillegg, siden bip-test MK er basert på prediksjonsligninger og ikke direkte målt prosenttilgrenser som 3000 m har, vil MK til bip-test sannsynligvis ikke ha samme nøyaktighet som 3000 m MK har.

Argumentet bygger på et par antakelser ved resultatene fra de 28 mannlige kadettene. Første antakelse er at kadettene var motivert under testing og presset seg maksimalt. Hvis kadettene ikke var motivert til å løpe bip-test er det svært sannsynlig at søkerne ved LFS-opptaket presterer bedre på bip-test ettersom skalaen, i så fall, er lagd for «umotiverte» løpere. Den andre antakelsen er kadettene hadde like testforhold og kunnskap om bip-test som LFS-søkerne hadde. Reglement for fysisk testing i Forsvaret hadde sannsynligvis ikke samme standardiserte krav for bip-test som det er i dag ettersom at bip-test ikke var en standard praksis i 2006. Jeg går ut ifra at disse antakelsene stemmer og at de ikke er faktorer som påvirker resultatene.

En annen antakelse er at unge menn er generelt like spreke eller godt trent i dag som i 2006. For å opprettholde intensjonen om prosentfordeling ved MK-skalaen kreves det en kontinuerlig kontroll og justering av kravene ettersom at fysisk form i samfunnet vil sannsynligvis utvikle seg. Ifølge Statistisk sentralbyrå (SSB) har mannlige vernepliktige blitt tyngre for hvert år, samtidig som høyden har stagnert (Kjelvik, 2011). Dette betyr at enten har unge menn trent på seg mer muskelmasse eller pådratt seg mer fett gjennom et dårligere kosthold. SSB påstår at unge menn lever på mange områder en sunnere livsstil, men dette er data fra 2009 som kan være utdatert for subjektene ved LFS-opptak (Kjelvik, 2011).

Allikevel, hvis flere unge menn blir tyngre grunnet økning av muskelmasse, kan det forklare hvorfor unge menn får bedre resultater på bip-test. Bip-test har flere anaerobe momenter i seg enn det 3000 m har (vendinger, akselerasjon og negativ akselerasjon). Hvis unge menn i dag er mer vant til å trene på slike bevegelser vil de ha en stor fordel på bip-test. I populære idretter som fotball eller håndball er disse momentene viktig. Norges Idrettsforbund hadde 1.93 millioner registrerte medlemmer ved utgangen av 2019, med fotballforbundet som det klart største særforbundet generelt og blant barn og ungdom (Ingebrigtsen, 2020). Den økende populariteten blant unge å drive med styrketrening og idretter som fotball og håndball kan forklare hvorfor unge menn presterer bedre i bip-test, og foretrekker denne over 3000 m.

MK-skalaen til bip-test ble endret i 2020, men spørsmålet er om den endringen likestilte kravene mer eller ikke. Jeg har ingen kilder på hvorfor kravene ble endret, men ut ifra hvordan kravene var før og hvordan de er nå kan vi se om de påvirket resultatene våre på en positiv måte eller ikke. På tabell 1a og 1b ser vi at MK 1 og MK 9 ble vanskeligere og økte med 2 shuttles, mens MK 2 – MK 8 ble lettere og senket kravet med 2-5 shuttles. Ut ifra resultatene til denne oppgaven er endringen positiv ettersom at det er et høyere antall med karakter 9 på bip test (n=146) enn på 3000 m (n=113). Hvis kravet hadde vært lavere, som det var før 2020, ville differansen vært enda større siden 3000 m kravet ikke har endret seg. Det kan likevel stilles spørsmål om endringen var stor nok. Figur 4 viser at betydelig flere fikk MK 9 på bip-test, men at det var relativt likt ved de andre karakterene. Dette kan være grunn til å øke kravet til MK 9 på bip-test.

Hvis man også sammenligner kravene for å oppnå MK 9 på hver test ser man at det er ulikheter. MK 9 på bip-test er 11:8 som tilsvarer 2 040 m i avstand,  $\geq 11$  minutter og 15 sekunder varighet og en maksimalfart på 13.5 km/t. MK 9 på 3000 m er 12:00 som naturligvis er 3000 m i avstand,  $\leq 12$  minutters varighet og en gjennomsnittsfart på 15 km/t. Bare på tallene alene kan man se at ved bip-test løper man nærmest en kilometer kortere, kan bruke 45 sekunder mindre og på det raskeste må man bare løpe 13.5 km/t. Selvfølgelig er testene ulike, hvor bip-test er en *incremental* test og har flere anaerobe momenter (vendinger, akselerasjon og negativ akselerasjon), men hvis unge menn er trent på dette området, gjennom idretter som fotball og håndball, er det kanskje ikke så rart om de prefererer bip-test over 3000 m. Både av egen erfaring og HV-soldatene ved Aandstad sitt studie foretrakk bip-test over 3000m (Kirknes, Aandstad, & Stornæs, 2014).

Allikevel er det flere andre preferanser som kan spille inn. En «mental belønning» for bedre prestasjon på 3000 m er at jo raskere man løper jo forttere blir man ferdig med testen, derimot ved bip-test må man løpe så lenge man gidder. Mental innstilling og egne preferanser har mye å si, men basert på funnene og tidligere studier er det en tendens at unge menn preferer bip-test over 3000 m.

### 6.3 Generalisering for resten av Forsvaret?

Hvis vi skal kunne si noe om at resultatene kan generaliseres for alle unge menn i Forsvaret må flere faktorer diskuteres. Generalisering kan defineres som påstander om at det som gjelder på ett sted eller én tid, også vil gjelde et annet sted eller til en annen tid (Payne & Williams, 2005). Først og fremst må man se om resultatene er signifikante eller ikke. I denne oppgaven var det to statistiske signifikante funn: Korrelasjonen mellom tidene på utholdenhetstestene og om det er forskjell i karakterer oppnådd på utholdenhetstestene. Siden korrelasjonsanalysen var forventet å være signifikant og støttet opp med tidligere forskning, velger jeg å fokusere å se om «forskjell i karakter oppnådd» kan generaliseres.

Det er hovedsakelig utvalget jeg har benyttet meg av som bestemmer om resultatet kan generaliseres. Som nevnt i metodekapittelet har flere subjekter blitt ekskludert fra utvalget av ulike årsaker. Utvalget besto opprinnelig av  $n=793$  subjekter, men 497 subjekter ble fjernet fordi de ikke kom videre til neste fase hvor man løp neste test. Ved en ny gjennomgang av dataen var det kun 87 av de 497 som fikk MK 4 eller mindre, som vil si at det var bare 87 av 497 subjekter som ble ekskludert fra utvalget grunnet for dårlige resultater på bip-test. De resterende 410 subjektene ble ekskludert siden de ikke løp test nummer 2 (3000 m). De 410 subjektene ble sannsynligvis selektert bort fra opptaket av årsaker som databaserte tester, helsesjekk eller intervju. Det gjør at utvalget er mer representativt siden jeg ikke har fjernet kun de som presterer dårlig på utholdenhetstester. I tillegg til de 497 subjektene ble en ekstra 32 stykk fjernet av andre årsaker. Disse var enten kvinner (for få for å være representativt), ekstremverdier eller feilkilder. Utvalget besto dermed av 264 unge menn. Utvalget er dermed representativt for unge menn som presterer godt (MK 5 – MK 9) på utholdenhetstester i Forsvaret.

I tillegg er det svært sannsynlig at de som er på opptak til LFS har enten tjent i Forsvaret tidligere eller skal inn i en annen stilling i Forsvaret hvis de ryker ut av seleksjonen. Det er bare søkere mellom 18 og 24 som kan delta på opptaket, og hvis de er interessert i LFS vil tjeneste i Luftforsvaret mulig også være interessant for dem. Det er vanskelig å si om 264

unge menn er representativt for eksempelvis ca. 6 500 førstegangstjenestegjørende (antall mannlige vernepliktige i 2021), men av de som presterer godt på utholdenhetstester er det sannsynlig at et flertall presterte bedre på bip-test enn 3000 m (Forsvaret, 2022).

#### 6.4 Oppsummering

Det fremkommer ved drøftingen at årsaken til korrelasjon mellom testprotokollene, kan være et resultat av at testprotokollene er knyttet opp til  $VO_{2maks}$  ettersom at mesteparten (85-90%) av energien kommer fra aerob energiomsetning. Ved to tidligere studier på soldater ser vi at begge tester har en god validitet opp mot  $VO_{2maks}$  og vil dermed korrelere godt. Ved videre drøfting fremkommer det mulige årsaker til hvorfor det er forskjell på oppnådd karakter ved testprotokollene. MK-skalaen til bip-test er basert på prediksjonsligninger fra en mye tynnere datamengde enn det 3000 m er, som betyr at validiteten til bip-test MK-skalaen ikke nødvendigvis er like god som 3000 m skalaen. Ved sammenligning av MK 9 krav for begge tester ser vi at ved bip-test løper man en kilometer kortere, bruker 45 sekunder mindre og må kun akselerere til en hastighet på 13.5 km/t. Hvis løperen er trent på og vant til vendinger og å akselerere, blir bip-test enklere å gjennomføre enn et 3000 m løp, noe som kan forklare årsaken til 40 % forbedring i bip-test kontra 3000 m. Til slutt fremkommer det ved drøfting at resultatene kan generaliseres for alle unge menn i Forsvaret som generelt sett presterer godt på utholdenhetstester.

### 7 Konklusjon

Denne oppgaven har tatt for seg bip-test og 3000 m prestasjoner og sammenhengen mellom disse. Oppgavens hensikt har vært å besvare følgende problemstillinger: «Er det en sammenheng mellom prestasjon på 3000 m og bip-test hos unge menn på seleksjon til Luftforsvarets flygerskole?» og «Med bruk av Forsvarets karaktersettingssystem, er det noen forskjell i oppnådd karakter på 3000 m og bip-test hos unge menn i Forsvaret?».

Resultatene viser en svært sterk negativ korrelasjon mellom tid på bip-test og 3000 m. Dette henger sammen med at begge tester er knyttet opp til  $VO_{2maks}$  og at 85-90 % av energien kommer fra aerob energiomsetning. Det var også en signifikant forskjell på karakterer oppnådd ved testprotokollene, hvor 40 % av subjektene oppnådde bedre karakter på bip-test. Dette fremkommer sannsynligvis av at bip-test-kravene, spesielt MK 9, er satt for lavt og er ikke likestilt med kravene til 3000 m på en god nok måte.

Denne studien har avdekket at det er en sterk samvariasjon mellom bip-test og 3000 m, samt at det er enklere for unge menn i Forsvaret å oppnå MK 9 på bip-test enn 3000 m.

Forslag til videre studier er å analysere karakterforskjeller hos rekrutter ved innrykk til GSU fra 2017 til i dag. Med den datamengden vil man kunne avsløre om det er karakterforskjell ved alle MK og ikke bare de øvre kravene. Samtidig vil en mulig løsning være å benytte seg av 20. prosentilgrensen på samme måte som MK-skalaen til 3000 m ble satt. Det vil sørge for at samme prosentilgrense vil oppnå de samme minstekravene uavhengig av hvilken test soldaten tar.

## Referanser

- Athlepedia. (2022, Juli 7). *Maximal Oxygen Uptake*. Hentet fra Athlepedia, The Athletics Wiki: [https://athletics.fandom.com/wiki/Maximal\\_Oxygen\\_Uptake](https://athletics.fandom.com/wiki/Maximal_Oxygen_Uptake)
- Bertelsen, P. W. (2022, Juli 22). Powerpoint brief Pilot & navigator selection. Rygge, Norge.
- Canino, M., Cohen, B. S., Redmond, J. E., Sharp, M. A., Zambraski, E. J., & Foulis, S. A. (2018, Mai-Juni). The Relationship Between Soldiers Performance on the Two-Mile Run and the 20-m Shuttle Run Test. *Military Medicine*, ss. e182-e187.
- Davis, J. A. (2006). Direct Determination of Aerobic Power. I P. J. Maud, & C. Foster, *Physiological Assessment of Human Fitness* (ss. 9-18). Human Kinetics.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. London: SAGE Publications.
- Forsvaret. (2020). *Reglement for fysisk test*. Oslo: Sjef Forsvarets høgskole.
- Forsvaret. (2022, Juni 12). *Flyger- og navigatørutdanning*. Hentet fra Forsvaret: <https://www.forsvaret.no/utdanning/utdanninger/flyger>
- Forsvaret. (2022, april 29). *Forsvaret i tall*. Hentet fra Forsvaret: <https://www.forsvaret.no/om-forsvaret/forsvaret-i-tall>
- Glen, S. (2022, juli 2). *Correlation Coefficient: Simple Definition, Formula, Easy Steps*. Hentet fra StatisticsHowTo.com: <https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/>
- Grønmo, S. (2022, juni 22). *Kvantitativ metode*. Hentet fra Store norske leksikon: [https://snl.no/kvantitativ\\_metode](https://snl.no/kvantitativ_metode)
- Hagen, R. (2009). *Innføring i personlighetspsykologi*. Trondheim: Tapir akademisk forlag.
- Harlem, T. O. (2016). *Seleksjon flygere F-35 - Neste generasjon jagerfly, en ny generasjon flygere?* Forsvarets Høgskole.
- Hopkins, W. G. (2006, august 7). *A Scale of Magnitudes for Effect Statistics*. Hentet fra A New View of Statistics: <https://www.sportsci.org/resource/stats/effectmag.html>
- Ingebrigtsen, P. (2020, oktober 6). *Nøkkeltallene for norsk idrett i 2019*. Hentet fra Norges Idrettsforbund: <https://www.idrettsforbundet.no/nyheter/2020/nokkeltallene-for-norsk-idrett-i-2019/>
- Kaiser, M. (2022, juni 22). *Kvantitativ metode*. Hentet fra De nasjonale forskningsetiske komiteene: <https://www.forskningsetikk.no/ressurser/fbib/metoder/kvantitativ-metode/>
- Kirknes, J., Aandstad, A., & Stornæs, A. V. (2014). *Revidert fysisk testordning for Forsvaret*. Oslo: Forsvarets Høgskole.
- Kjelvik, J. (2011, februar 28). *Vernepliktige opp i vekt*. Hentet fra Statistisk sentralbyrå: <https://www.ssb.no/helse/artikler-og-publikasjoner/vernepliktige-opp-i-vekt>

- Lockheed Martin. (2022, august 2). *Everything You Need to Know about the F-35C*. Hentet fra F-35 Lightning II: <https://www.lockheedmartin.com/f35/news-and-features/everything-you-need-to-know-about-the-f-35c.html>
- Martinussen, M. (2005). Seleksjon av flygere og flygeledere. *Tidsskrift for Norsk psykologforening, Vol 42, nummer 4*, ss. 291-300.
- Olympiatoppen. (2022, juli 17). *Prestasjonsbestemmende faktorer*. Hentet fra Olympiatoppen: <https://olympiatoppen.no/fagomrader/utholdenhet/fagstoff/prestasjonsbestemmende-faktorer2/>
- Payne, G., & Williams, M. (2005, April). Generalization in Qualitative Research. *Sociology Vol 39*, ss. 295-314.
- Pripp, A. H. (2018, mai 8). *Pearsons eller Spearmans korrelasjonskoeffisienter*. Hentet fra Tidsskriftet - Den Norske Legeforening: <https://tidsskriftet.no/2018/05/medisin-og-tall/pearsons-eller-spearmans-korrelasjonskoeffisienter>
- Regjeringen. (2017, november 7). *Hva koster egentlig nye kampfly?* Hentet fra Regjeringen: <https://www.regjeringen.no/no/tema/forsvar/innsikt/kampfly/hva-koster-egentlig-nye-kampfly/id710435/>
- Scheff, S. W. (2022, Juli 2). *Wilcoxon Signed Ranks Test*. Hentet fra ScienceDirect: <https://www.sciencedirect.com/topics/medicine-and-dentistry/wilcoxon-signed-ranks-test>
- Science Direct. (2022, juli 17). *Signed Rank Test*. Hentet fra ScienceDirect: <https://www.sciencedirect.com/topics/mathematics/signed-rank-test>
- Sport Fitness Advisor. (2022, Juli 7). *VO2 Max, Aerobic Power & Maximal Oxygen Uptake*. Hentet fra Sport Fitness Advisor: <https://www.sport-fitness-advisor.com/VO2max.html>
- Thrane, C. (2018). *Kvantitativ Metode*. Oslo: Cappelen Damm Akademisk.
- Torjussen, T. M., & Hansen, I. (1999). Forsvaret, best i test? *Tidsskrift for Norsk psykologforening, Vol 36, Nr 8*, ss. 772-779.
- Wilkinson, D. M., Blacker, S. D., Richmond, V. L., Rayson, M. P., & Bilzon, J. L. (2014, Februar). Relationship Between the 2.4-km Run and Multistage Shuttle Run Test Performance in Military Personnel. *Military Medicine*, ss. 203-207.
- Zach. (2021, Mars 1). *How to Perform a Normality Test in Excel (Step-by-Step)*. Hentet fra Statology: <https://www.statology.org/normality-test-excel/>
- Aandstad, A. (2011). *Fysiske arbeidskrav for militært personell*. Oslo: Norges idrettshøgskole Forsvarets institutt.
- Aandstad, A. (2016). *Minimumskravskalaer for fysiske tester i Forsvaret*. Sognsvann: Norges idrettshøgskole Forsvarets institutt.
- Aandstad, A. (2021). Estimation of maximal oxygen uptake from the 3,000 m run in adult men and women. *Journal of Sports Sciences*, 1-8.



Aandstad, A., Berntsen, S., Holme, I., & Anderssen, S. A. (2011, Mai 14). Validity and Reliability of the 20 Meter Shuttle Run. *Military Medicine Vol. 176*, ss. 513-518.

## Vedlegg A

Level	Shuttles	Running speed (km/h)	time per shuttle (s)	Total level time (s)	Cumulative Time (min:seconds)	Distance per level (m)	Cumulative Distance (m)
1	7	8.0	9.00	63.0	1:03	140	140
2	8	9.0	8.00	64.0	2:07	160	300
3	8	9.5	7.58	60.6	3:08	160	460
4	9	10.0	7.20	64.8	4:12	180	640
5	9	10.5	6.86	61.7	5:14	180	820
6	10	11.0	6.55	65.5	6:20	200	1020
7	10	11.5	6.26	62.6	7:22	200	1220
8	11	12.0	6.00	66.0	8:28	220	1440
9	11	12.5	5.76	63.4	9:32	220	1660
10	11	13.0	5.54	60.9	10:32	220	1880
11	12	13.5	5.33	64.0	11:36	240	2120
12	12	14.0	5.14	61.7	12:38	240	2360
13	13	14.5	4.97	64.6	13:43	260	2620
14	13	15.0	4.80	62.4	14:45	260	2880
15	13	15.5	4.65	60.4	15:46	260	3140
16	14	16.0	4.50	63.0	16:49	280	3420
17	14	16.5	4.36	61.1	17:50	280	3700
18	15	17.0	4.24	63.5	18:53	300	4000
19	15	17.5	4.11	61.7	19:55	300	4300
20	16	18.0	4.00	64.0	20:59	320	4620
21	16	18.5	3.89	62.3	22:01	320	4940

Hentet fra: [Beep Test Table \(beep-test-table.com\)](http://beep-test-table.com)