

Utvalgsstørrelse, styrke

Lise Lund Håheim DDS, PhD

Professor II, Forskerlinjen, UiO

Seniorforsker, Nasjonalt kunnskapssenter for helsetjenesten, Oslo

Seniorforsker, Institutt for oral biologi, UiO

Introduksjonskurset, forskerutdanningen

Det medisinske fakultet, UiO



- Hypotesetesting
- Teststyrke
- Beregning av utvalgsstørrelse

- Beregning nødvendig for
 - Planlegging
 - Pasienter
 - Ressursbruk
 - Tid
 - Bemanning
 - Kostnader
 - Søknader
- Litteratur: Laake P et al. Forskning i medisin og biofag. Kap.9

Å teste en hypotese

- H_0 – Den du forkaster
- H_a – Den du ønsker å vise

Hypotesetesting

	Ho er sann	Ho er feil
Ho forkastes ikke	A	B
Ho forkastes	C	D

- **A og D er ønskelige begivenheter**
 - $P(A)$ er sannsynligheten for at H_0 ikke forkastes når H_0 er sann
 - $P(D)$ er sannsynligheten for at H_0 forkastes når H_0 ikke er sann
- **B og C er feil konklusjoner**
 - $P(C)$ er sannsynligheten for at H_0 forkastes når den er sann
 - $P(B)$ er sannsynligheten for at H_0 ikke forkastes selv om den er feil

Hypotesetesting

	Ho er sann	Ho er feil
Ho forkastes ikke	Korrekt konklusjon $P(A)=1-\alpha$	Type 2 feil $P(B)=\beta$
Ho forkastes	Type 1 feil $P(C)=\alpha$	Korrekt konklusjon $P(D)=1-\beta$ Teststyrke/Power

- **Signifikansnivå**

- Sannsynligheten for å forkaste H_0 når H_0 er sann
- Den ønsker vi skal være lav
- Vanligst brukt nivå er $\alpha = 5\%$ eller 1%
 $P < 0.05$, $p < 0.01$

- **Teststyrke / Power**

- Sannsynligheten for å forkaste H_0 når den faktisk er feil
- Den ønsker vi skal være høy
- Vanligst brukt nivå er $1 - \beta = 80-95\%$

Effektstudier - styrkeberegning

- Hvor stor forskjell i effekt anser man for klinisk relevant?
- Hva er effektmålet?
- Hvor stort utvalg må man ha for å vise denne forskjellen med stor nok sikkerhet?
- Ved en styrkeberegning ønsker man å beregne en studie slik at man kan vise en sann effekt hvis den virkelig er det.

Størrelsen på kliniske studier – hvor mange trenger vi?

1. Statistisk/vitenskapelig tilnærming

- Hvor mange trengs for å få sikker evidens på behandlingseffekt hvis en slik effekt eksisterer?

2. Økonomisk tilnærming

- Hvor mange pasienter er tilgjengelig?
- Hvor mye tid, arbeid og kostnader vil medgå?

3. Etisk tilnærming

- Når kan vi stoppe studien for å unngå at pasienter får den minst effektive behandlingen

Utvalgsstørrelse, binomisk respons

- Antagelse:
- Standard behandling gir 40% overlevelse over 5 år
- Ny behandling forventes å gi 10 prosentpoeng bedret overlevelse
- P_s (0.40)
- P_n (0.50)
- $n = \frac{P_n(1-P_n) + P_s(1-P_s)}{(P_n - P_s)^2} \cdot c$
- = n i hver gruppe

$$n = \frac{0.50(1-0.50) + 0.40(1-0.40)}{(0.50 - 0.40)^2} \cdot c$$

$$n = \frac{0.25 + 0.24}{0.01} \cdot c$$

Velger her:

Signifikansnivå $\alpha = 5$

Teststyrke $1 - \beta = 80$

Dette gir $c = 7.9$

Vi får:

$n = 387$ i hver gruppe

Utvalgsberegning – konstanten c

	Teststyrke		
Signifikansnivå (tosidig)	0.80	0.90	0.95
0.10	6.2	8.6	10.8
0.05	7.9	10.5	13.0
0.01	11.7	14.9	17.8

Eksempel: medikament versus placebo i sekundærprofylakse etter hjerteinfarkt

- Vi ønsker å studere om medikamentet A reduserer risikoen for å dø
- Dødsfall innen et år
- Antar 20% et års dødelighet for pasienter på placebo
- Anses viktig å oppnå en 25% reduksjon i dødelighet
- Hvis sant, da ønsker vi å være 80 % sikre på å oppnå en p-verdi < 0.05

Utrekning

- $n = \frac{P_n(1-P_n) + P_s(1-P_s)}{(P_n - P_s)^2} \cdot c$

- = n i hver gruppe

- $P_s(0.20) = 20\%$

- $P_n(0.15) = 15\%$

- $\alpha = 0.05$

- $1-\beta = 80\%$ power

$$n = \frac{P_n(1-P_n) + P_s(1-P_s)}{(P_n - P_s)^2} \cdot c$$

$$n = \frac{0.2(1-0.2) + 0.15(1-0.15)}{(0.2-0.15)^2} \cdot 7.9$$

$$n = \frac{0.2(0.8) + 0.15(0.85)}{0.0025} \cdot 7.9$$

$$n = \frac{0.16 + 0.1275}{0.0025} \cdot 7.9$$

n = 909 i hver gruppe

Alt. 1: 2% reduksjon / endring

- $P_s (0.20) = 20\%$
- $P_n (0.18) = 18\%$
- $\alpha = 0.01$
- $\beta = 0.05$; $1 - \beta = 95\%$ power

$$n = \frac{P_n (1 - P_n) + P_s (1 - P_s)}{(P_n - P_s)^2} \cdot c$$

$$n = \frac{0.2(1-0.2) + 0.18(1-0.18)}{(0.2-0.18)^2} \cdot 17.8$$

$$n = \frac{0.16 + 0.1476}{0.0004} \cdot 17.8$$

$$n = 13,684 \text{ i hver gruppe}$$

Alt. 2: 50% reduksjon / endring

- $P_s (0.20) = 20\%$
- $P_n (0.10) = 10\%$
- $\alpha = 0.05 \%$
- $\beta = 0.5: 1 - \beta = 50 \%$ power

$$n = \frac{P_n (1 - P_n) + P_s (1 - P_s)}{(P_n - P_s)^2} \cdot c$$

$$n = \frac{0.2(1-0.2) + 0.10(1-0.10)}{(0.2-0.15)^2} \cdot 3.8$$

$$n = \frac{0.2 (0.8) + 0.1 (0.9)}{0.01} \cdot 3.8$$

$$n = \frac{0.16 + 0.09}{0.01} \cdot 3.8$$

$n = 95$ i hver gruppe

Vi ser:

- N øker når:
 - Effektmålet $p_n - p_s$ minker
 - Type 1 feil α minker
 - Styrke / power $1 - \beta$ øker

Styrke /power for å oppdage en forskjell hvis den virkelig er der

Utvalgsstørrelse, kontinuerlig respons.

Par av observasjoner

- Par av observasjoner (samme person); for eksempel crossover studie
- $n = (\sigma_d/\Delta)^2 \cdot C$
- Der Δ er forskjell i den behandlingseffekt man ønsker å se
- σ_d er standardavvik til differansen mellom behandlingseffektene
- n = totalt antall deltakere

Utregning. Par av observasjoner

- Intervensjon: Effekt av to legemidler mot astma.
- Effekt måles ved hjelp av PEF (peak respiratory flow rate i l/min)
- $\Delta = 30$ l/min
- $\sigma_d = 50$ l/min
- $n = (\sigma_d/\Delta)^2 \cdot C$
- $n = (50 / 30)^2 \cdot 7.9$
- $n = 22$ (totalt)

Utvalgsstørrelse, kontinuerlig respons.

Parallele grupper

- Parallele grupper, σ antas å være lik i gruppene:
- $n = 2 (\sigma/\Delta)^2 \cdot C$
- $n =$ antall i hver gruppe

Utregning. Parallele grupper

- Intervensjon: Effekt av to legemidler mot astma.
- Effekt måles ved hjelp av PEF (peak respiratory flow rate i l/min)
- $\Delta = 30$ l/min
- $\sigma_d = 60$ l/min
 - σ_d er større her fordi vi sammenlikner forskjellige personer og ikke samme personer.
- $n = 2 (\sigma/\Delta)^2 \cdot C$
- $n = 2(60 / 30)^2 \cdot 7.9$
- $n = 63$ (126 totalt)

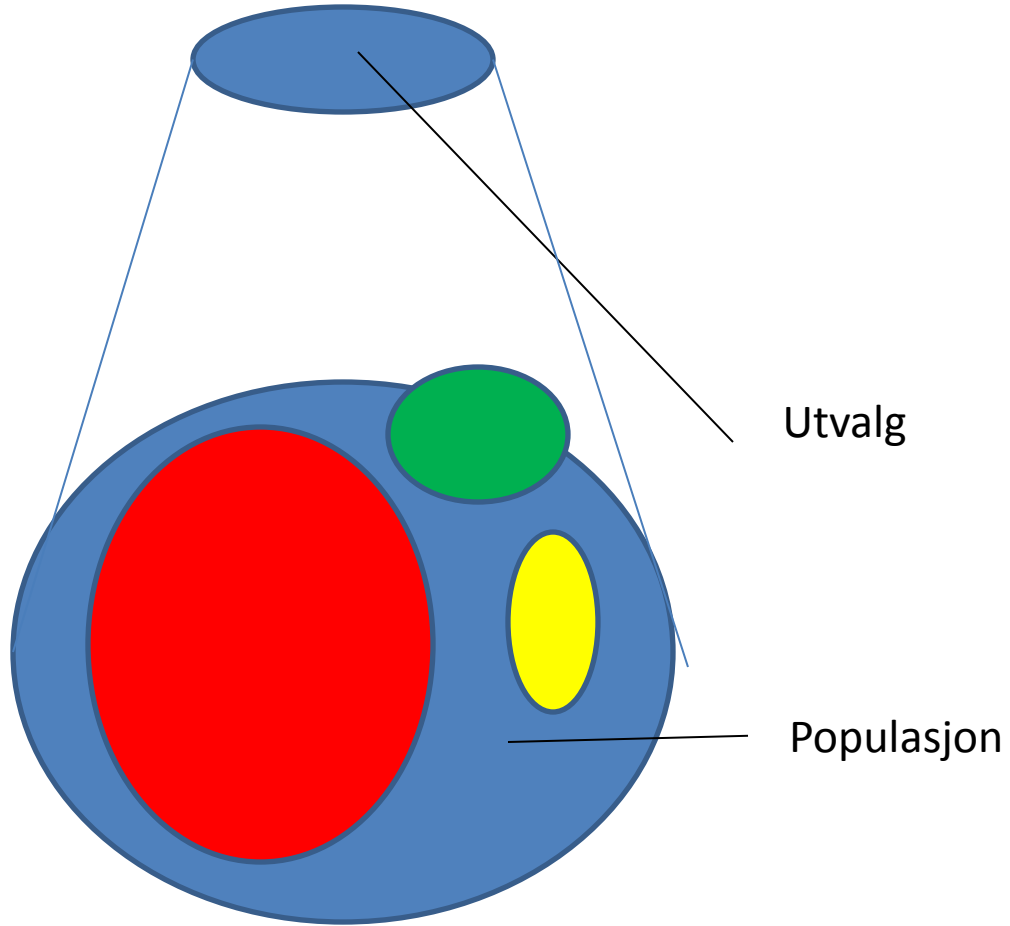
2 X 2 faktoriell design

- Fordeler:
 - Kan teste to behandlingshypoteser uten å øke antall
 - Få erfaring av kombinasjonen
- Ulemper:
 - Problem med tolkning hvis det er tegn til interaksjon
 - Praktiske begrensninger

”Non-inferiority” studier. Likeverdighet

- Spesielt ved legemidler:
- Nok å vise at medikamenter er like:
- H_0 = behandlingene er ulike
- H_a = behandlingene er like
- Ved effektstudier ønsker å vise forskjell:
- H_0 = intervensjons- og kontrollgruppene er like
- H_a = intervensjons- og kontrollgruppene er forskjellige

Utvalg versus populasjon



Hvor representativ er studiepopulasjonen?

- Generalisering / ekstern validitet

Målpopulasjon

Kildepopulasjon

Inkluderbare
personer

Studiedeltakere

Referanser for beregning av utvalgsstørrelse

- Hein Stigum, FHI, UiO:

<http://folk.uio.no/heins/>

- Google – "Sample size", power, calculator

<http://stat.uiowa.edu/~rlenth/Power/>