

# **Otoskoon arviointi**

**Tero Vahlberg**

# Otoskoon arviointi

- Otoskoon arviointi (sample size calculation) ja tutkimuksen voima-analyysi (power analysis) ovat tilastollisen tutkimuksen suunnittelussa keskeisiä kysymyksiä
- Otoskokolaskelmat tulisi esittää jokaisessa hyvin tehdyssä tutkimussuunnitelmassa. Otoskokolaskelmat kertovat, että tutkijalla tai tutkimusryhmällä on ennen aineiston keräämistä selkeä käsitys, mitä tutkimushypoteesia (tutkimushypoteeseja) ollaan tutkimassa. Myös monet lehdet vaativat näitä laskelmia.
- Otoskoon on oltava riittävän suuri, jotta kliinisesti merkittävä ero pystytään suurella todennäköisyydellä havaitsemaan
- Suurilla aineistoilla kliinisesti hyvin pienetkin erot tulevat usein tilastollisesti merkitseviksi, koska tutkimuksen voima on niin suuri (estimaattien keskivirheet pienenevät ja tarkkuus paranee, jolloin estimaattien luottamusvälit ovat pieniä).
- Liian suuren potilasmäärän kerääminen ei ole taloudellisesti ja ajallisesti järkevää. Lisäksi se voi olla epäeettistä, jos tutkitaan uutta hoitomuotoa.
- Pienillä aineistoilla kliinisesti suuriakaan eroja ei kyetä havaitsemaan tilastollisesti merkitseviksi, koska tutkimuksen voima on niin pieni (estimaattien keskivirheet ovat suuret, jolloin estimaattien luottamusvälit ovat suuria)
- Liian pienen potilasmäärän kerääminen on tutkimusresurssien tuhlaamista. Sekin voi olla epäeettistä, jos altistetaan potilaat turhaan uudelle hoitomuodolle.

- Laskelmat tehdään yleensä yhden tai muutaman tärkeimmän vastemuuttujan (primary outcomes) mukaan. Eri vastemuuttujille saadaan erilainen vaatimus aineiston koosta. Usein näistä valitaan suurin otoskokovaatimus.
- Tarvittavan otoskoon määrittämiseen on olemassa laskentakaavoja, taulukkokokoelmia ja tilasto-ohjelmistoja
  - *esim. tilasto-ohjelmistot SAS, NQuery, JMP ja R.*
  - *esim. Prof. S. Sarnan Excel-pohjaiset laskukaavat osoitteessa:  
“<http://www.kttl.helsinki.fi/sarna/Stats/Aineistokoko/>”*
- Otoskoon arvioinnissa on käytännössä hyvä tehdä laskelmia useammille vaihtoehtoisille tilanteille, jotta eri tekijöiden vaikutus otoskoon hahmottuu
- Otoskokoa arvioitaessa on huomioitava kadon arviointi. Jos esim. oletetaan katoa tulevan 20%, niin tarvittava otoskoko on  $N / (1 - 0.2)$  (=1.25-kertainen aineiston koko)
- Otoskokolaskelmat ovat tärkeitä, mutta käytännössä ne toimivat kuitenkin vain yleisenä ohjenuorana tehtäessä päätöstä kerättävän aineiston koosta
- Lopullisen aineistokoon valintaan vaikuttavat otoskokolaskelmien lisäksi myös potilaiden saatavuus sekä eettiset, taloudelliset resurssit, käytettävissä oleva ja käytännölliset syyt

## Määritelmiä

- **Hylkäämisvirhe (väärän positiivisen tuloksen riski), tyypin I virhe,  $\alpha$ -virhe, p-arvo:** Testin perusteella todetaan, että eroa on eli hylätään nollahypoteesi, vaikka todellisuudessa eroa ei ole
- **Hyväksymisvirhe (väärän negatiivisen tuloksen riski) , tyypin II virhe,  $\beta$ -virhe:** Testin perusteella todetaan, että eroa ei ole eli hyväksytään nollahypoteesi, vaikka todellisuudessa eroa on
- Tutkimuksen (testin) voima  $1 - \beta$  (= havaitaan testin perusteella ero, joka on myös todellinen ero) määritellään tutkimuksen otoskokoa laskettaessa

|                           |             | totuus<br>erosta                    |                                   |
|---------------------------|-------------|-------------------------------------|-----------------------------------|
|                           |             | eroa                                | eroa ei ole                       |
| testin<br>tulos<br>erosta | eroa on     | ei virhettä<br>$1 - \beta$ (=voima) | tyypin I virhe<br>$\alpha$ -virhe |
|                           | eroa ei ole | tyypin II virhe<br>$\beta$ -virhe   | ei virhettä                       |

## **Kysymyksiä otoskoon laskemista varten:**

1. Mikä on kokeen päätarkoitus? Määritellään tutkimushypoteesi.
2. Mikä on keskeisin vastemuuttuja?
3. Mitä tilastollista analyysimenetelmää käytetään eron testaamiseksi?
4. Millaisia keskimääräisiä tuloksia on ennakoitavissa standardikäsitteilyllä (eli millaisia tuloksia oletetaan vertailuryhmässä)?
5. Mikä on pienin kliinisesti merkittävä ero, joka on tärkeätä paljastaa suurella varmuudella (=vaikutuksen suuruus eli effect size)?
6. Mikä on tutkijan sopivaksi katsoma tilastollinen merkitsevyystaso ja testin voima?

## **Tilastollinen merkitsevyystaso (p-arvo):**

- Todennäköisyys, että hylätään nollahypoteesi, vaikka se on totta (tyypin I virhe)
- Välttämätön edellytys kliiniselle merkittävyydelle, mutta ei kerro mitään vaikutuksen suuruudesta
- Merkitään symbolilla  $\alpha$  ja useimmiten käytetään arvoa 0.05
- Päätetään, käytetäänkö yksi- vai kaksisuuntaista vaihtoehtoista hypoteesia

## **Tutkimuksen/testin voima:**

- Todennäköisyys, että olemassa oleva ero havaitaan eli nollahypoteesi hylätään, kun vaihtoehtoinen hypoteesi on tosi
- Merkitään ”symbolilla”  $1-\beta$  ja yleensä käytetään 80-90% voimaa

Esimerkki taulukosta, josta nähdään otoskoon laskentakaavoissa tarvittavan funktion  $f(\alpha, \beta)$  arvoja valittujen virhetasojen mukaan:

| <b>Tutkimuksen voimakkuus (1-<math>\beta</math>)</b>            |             |             |             |             |             |
|---|-------------|-------------|-------------|-------------|-------------|
|   |             | <b>0.95</b> | <b>0.90</b> | <b>0.80</b> | <b>0.50</b> |
| <b><math>\alpha</math></b><br><b>(tyypin</b><br><b>I virhe)</b> | <b>0.1</b>  | 10.8        | 8.6         | 6.2         | 2.7         |
|   | <b>0.05</b> | 13.0        | 10.5        | 7.9         | 3.8         |
|   | <b>0.02</b> | 15.8        | 13.0        | 10.0        | 5.4         |
|   | <b>0.01</b> | 17.8        | 14.9        | 11.7        | 6.6         |

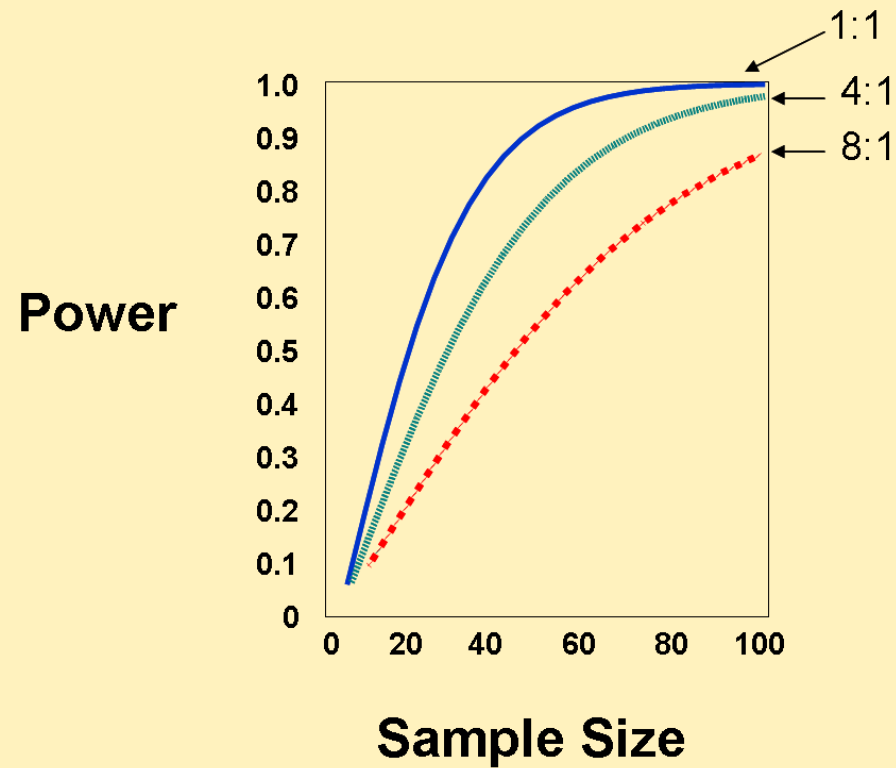
Funktion  $f$  arvon millä tahansa  $\alpha$ :n ja  $\beta$ :n arvoilla voi määrittää standardoidun normaalijakauman mukaisesti (esim. taulukon avulla) seuraavan kaavan avulla:

$$f(\alpha, \beta) = \left( z_{1-\alpha/2} + z_{1-\beta} \right)^2$$

# Esimerkkikuvia tutkimuksen voimaan vaikuttavista tekijöistä:

## Ryhmien potilasmäärän suhteiden vaikutus tutkimuksen voimaan

### Power Curves by Ratio of Group Sample Sizes



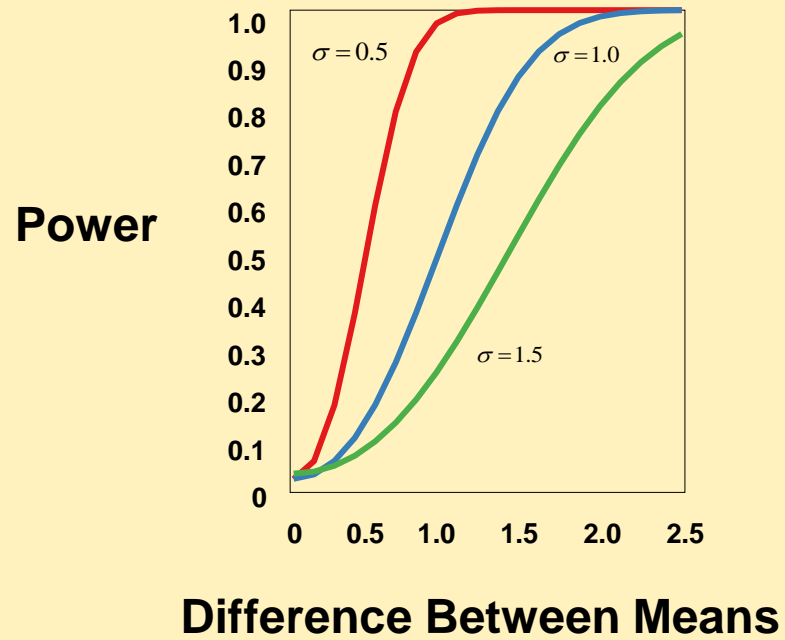
2

Lähde: SAS Institute



# Vastemuuttujan keskihajonnan vaikutus tutkimuksen voimaan

## Power Curves by Standard Deviation

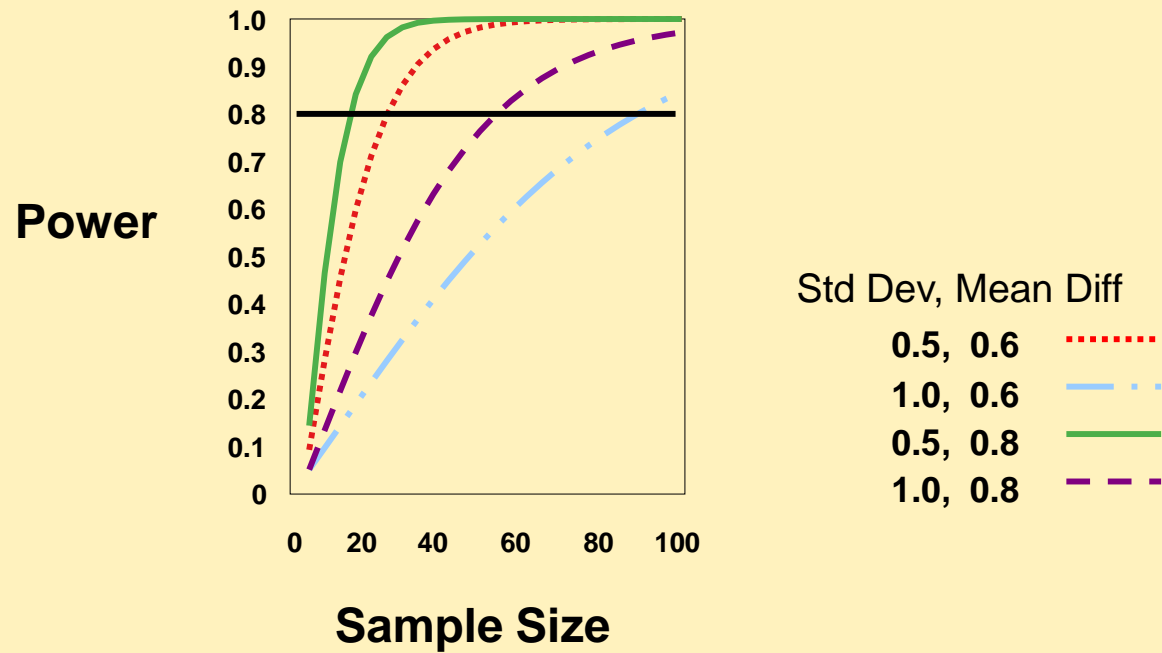


5

Lähde: SAS Institute

# Vastemuuttujan keskiarvoeron ja keskihajonnan vaikutus tutkimuksen voimaan

## Sensitivity Analysis



6

Lähde: SAS Institute

## Kahden ryhmän suhteellisten osuuksien vertailu

Kun halutaan vertailla kahden ryhmän suhteellisia osuuksia (prosentteja), tarvittava N (per ryhmä) voidaan laskea seuraavalla kaavalla:

$$N = \frac{P_1 \times (100 - P_1) + P_2 \times (100 - P_2)}{(P_1 - P_2)^2} \times f(\alpha, \beta)$$

Kaavassa  $P_1$  on hoidon onnistumisen osuus prosentteina vertailuhoitoa saaneilla ja  $P_2$  uutta hoitoa saaneilla. Funktion  $f(\alpha, \beta)$  arvo määräytyy valittujen virhetasojen perusteella.

## Kahden ryhmän suhteellisten osuuksien vertailu, esimerkki:

1. Tutkimuksessa päätarkoituksena on selvittää, onko testiläkkeellä vaikutusta sydäninfarktin jälkeiseen kuolleisuuteen
2. Ensisijainen vastemuuttuja on kuolleisuus kahden vuoden kuluessa
3. Valitaan käytettäväksi tilastolliseksi testiksi  $\chi^2$ -testi.
4. Plasebo-käsittelyssä oletetaan noin 10% potilaista kuolevan kahden vuoden sisällä ( $P_1$ )
5. Jos testilääke vähentää kuolleisuutta puoleen (5%), eroa voidaan pitää kliinisesti merkittävänä ( $P_2$ )
6. Käytetään 5% merkitsevyystasoa (2-suuntainen testaus) ja 90% voimakkuutta ( $1-\beta$ ) eli haluttu 5% ero käsittelyissä paljastuu 90% varmuudella

$$N = \frac{10 \times (100 - 10) + 5 \times (100 - 5)}{(10 - 5)^2} \times f(0.05, 0.10) = \frac{900 + 475}{25} \times 10.5$$

= **577.5** potilasta/ryhmä

# Esimerkki otoskoon laskemisesta SAS-ohjelmalla, suhteellisten osuuksien vertailu

## SAS-koodi:

```
proc power;  
  twosamplefreq test=pchi  
    groupproportions = (.10 .05)  
    npergroup = .  
    power = 0.80 0.90;  
run;
```

## SAS-tulostaulukot:

| Fixed Scenario Elements    |                      |
|----------------------------|----------------------|
| Distribution               | Asymptotic normal    |
| Method                     | Normal approximation |
| Group 1 Proportion         | 0.1                  |
| Group 2 Proportion         | 0.05                 |
| Number of Sides            | 2                    |
| Null Proportion Difference | 0                    |
| Alpha                      | 0.05                 |

| Computed N per Group |               |              |             |
|----------------------|---------------|--------------|-------------|
| Index                | Nominal Power | Actual Power | N per Group |
| 1                    | 0.8           | 0.801        | 435         |
| 2                    | 0.9           | 0.900        | 582         |

## Eri tekijöiden vaikutukset otoskokoihin, kahden ryhmän suhteellisten osuuksien vertailu:

Otoskoot on laskettu SAS-ohjelmalla ja kerätty taulukkoon.

### Suhteellinen osuus %

| Alpha | 1-beta (power) | Plasebo | Lääke | N/ryhmä | N    |
|-------|----------------|---------|-------|---------|------|
| 0.05  | 0.8            | 10      | 5     | 435     | 870  |
| 0.05  | 0.9            | 10      | 5     | 582     | 1164 |
| 0.05  | 0.8            | 10      | 7     | 1356    | 2712 |
| 0.05  | 0.9            | 10      | 7     | 1814    | 3628 |
| 0.05  | 0.8            | 20      | 10    | 199     | 398  |
| 0.05  | 0.9            | 20      | 10    | 266     | 532  |
| 0.05  | 0.8            | 20      | 5     | 76      | 152  |
| 0.05  | 0.9            | 20      | 5     | 101     | 202  |

## Kahden ryhmän keskiarvon vertailu

Kun halutaan vertailla kahden ryhmän keskiarvoja kahden otoksen t-testillä (ryhmillä sama hajonta) tarvittava  $N$  (per ryhmä) voidaan laskea seuraavalla kaavalla:

$$N = \frac{2SD^2}{(\bar{x}_2 - \bar{x}_1)^2} \times f(\alpha, \beta)$$

Kaavassa  $SD$  on vastemuuttujan keskihajonta ( $SD^2$  on vastemuuttujan varianssi) ja  $(\bar{x}_2 - \bar{x}_1)$  on ryhmien keskiarvojen kliinisesti merkittävä erotus. Funktion  $f(\alpha, \beta)$  arvo määräytyy valittujen virhetasojen perusteella.

## Kahden ryhmän keskiarvon vertailu, esimerkki:

1. Tutkimuksessa päätarkoituksena on selvittää, laskeeko uusi kolesteroliaineenvaihduntaan vaikuttava lääke kolesterolia
2. Ensisijainen vastemuuttuja on kokonaiskolesteroliarvon muutos 3 kk aikana
3. Valitaan käytettäväksi tilastolliseksi testiksi kahden otoksen t-testi
4. Plasebo-käsittelyn ei oleteta laskevan kolesterolia
5. Jos kolesteroli laskee testilääkkeen avulla 0.5 mmol/l, laskua voidaan pitää kliinisesti merkittävänä. Aiempien tutkimusten perusteella tiedetään, että kolesterolin keskihajonta tällaisessa potilasryhmässä on 1.2 mmol/l
6. Käytetään 5% merkitsevyystasoa (2-suuntainen testaus) ja 90% voimakkuutta (0.5 mmol/l lasku paljastuu 90% varmuudella)

$$N = \frac{2 \times 1.2^2 \text{ mmol}^2}{0.5^2 \text{ mmol}^2} \times f(0.05, 0.1) = \frac{2.88}{0.25} \times 10.5$$

= **121** potilasta/ryhmä



# Esimerkki otoskoon laskemisesta SAS-ohjelmalla, kahden ryhmän keskiarvon vertailu

## SAS-koodi:

```
proc power;  
  twosamplemeans test=diff  
    meandiff = 0.5  
    stddev = 1 1.2 1.4  
    npergroup = .  
    power = 0.80 0.90;  
run;
```

## SAS-tulostaulukot:

| Fixed Scenario Elements |        |
|-------------------------|--------|
| Distribution            | Normal |
| Method                  | Exact  |
| Mean Difference         | 0.5    |
| Number of Sides         | 2      |
| Null Difference         | 0      |
| Alpha                   | 0.05   |

| Computed N per Group |         |               |              |             |
|----------------------|---------|---------------|--------------|-------------|
| Index                | Std Dev | Nominal Power | Actual Power | N per Group |
| 1                    | 1.0     | 0.8           | 0.801        | 64          |
| 2                    | 1.0     | 0.9           | 0.903        | 86          |
| 3                    | 1.2     | 0.8           | 0.803        | 92          |
| 4                    | 1.2     | 0.9           | 0.902        | 123         |
| 5                    | 1.4     | 0.8           | 0.803        | 125         |
| 6                    | 1.4     | 0.9           | 0.900        | 166         |

## Eri tekijöiden vaikutukset otoskokoihin, kahden ryhmän keskiarvon vertailu:

Otoskoot on laskettu SAS-ohjelmalla ja kerätty taulukkoon.

| Alpha | 1-beta (power) | Keskiarvo muutokselle |       |      | Keskihajonta | N/ryhmä | N   |
|-------|----------------|-----------------------|-------|------|--------------|---------|-----|
|       |                | Plasebo               | Lääke | Ero  |              |         |     |
| 0.05  | 0.8            | 0                     | -0.5  | -0.5 | 1            | 64      | 128 |
| 0.05  | 0.9            | 0                     | -0.5  | -0.5 | 1            | 86      | 172 |
| 0.05  | 0.8            | 0                     | -0.5  | -0.5 | 1.2          | 92      | 184 |
| 0.05  | 0.9            | 0                     | -0.5  | -0.5 | 1.2          | 123     | 246 |
| 0.05  | 0.8            | 0                     | -0.5  | -0.5 | 1.4          | 125     | 250 |
| 0.05  | 0.9            | 0                     | -0.5  | -0.5 | 1.4          | 166     | 332 |
| 0.05  | 0.8            | 0                     | -0.3  | -0.3 | 1.2          | 253     | 506 |
| 0.05  | 0.9            | 0                     | -0.3  | -0.3 | 1.2          | 338     | 676 |
| 0.05  | 0.8            | 0                     | -1    | -1   | 1.2          | 24      | 48  |
| 0.05  | 0.9            | 0                     | -1    | -1   | 1.2          | 32      | 64  |