

Vaikuttavuuden arviointi bayesiläisestä näkökulmasta

Esa Läärä

Oulun yliopisto, matemaattisten tieteiden laitos
& FinOHTA/Stakes <esa.laara@oulu.fi>

Suomen Tilastoseura ja MSM-tutkimusryhmä
Helsingin yliopisto 14.2.2007

1 / 32

Bayesiläinen on sellainen, joka

- kysyy, mitä ajattelit ennen koetta, kertoakseen mitä ajattelet kokeen jälkeen,
- odottaessaan hevosen tulevan vastaan kohtaakin aasin, ja päättelee nähneensä muulin.

Senn, S. (1997). *Statistical Issues in Drug Development*. Wiley

2 / 32

Bayesiläisyys terveydenhuollon arvioinnissa

"Ulkoisen näytön eksplisiittinen kvantitatiivinen käyttö terveydenhuollon arviointitutkimuksen suunnittelussa, sen tulosten monitoroinnissa, analyysissä, tulkinnassa ja raportoinnissa."

- Spiegelhalter ym. (1999). An introduction to bayesian methods in health tech'gy ass't. *BMJ* **319**: 508-12.
- Spiegelhalter ym. (2000). Bayesian methods in health technology assessment: a review. *HTA* **4**: 1-130.
<http://www.hta.nhsweb.nhs.uk/fullmono/mon438.pdf>
- Spiegelhalter ym. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley.
- ks. myös Bayes-artikkelit *Impakti* 4/2006, s. 16-19
<http://finohta.stakes.fi/FI/julkaisut/impakti/>

3 / 32

Intervention vaikuttavuus – PICO

Kuinka suuri on vasteen O odotusarvo π_I kohdepopulaatiossa P intervention I jälkeen verrattuna siihen, mitä sen arvo π_C olisi, jos olisi sovellettu vertailukäsittelyä C ?

Vaikutuksen vertailuparametreja erityisesti binäärivasteille:

$$\begin{aligned}RD &= \pi_I - \pi_C, \text{ riskiero,} \\RR &= \pi_I / \pi_C, \text{ riskisuhde,} \\RRR &= 1 - RR, \text{ riskin suhteellinen vähenemä,} \\OR &= (\pi_I / \pi_C) / [(1 - \pi_I) / (1 - \pi_C)], \text{ vetosuhde,} \\NNT &= -1 / RD, \text{ ns. NNT-luku (number needed to treat)}\end{aligned}$$

Ongelma: Samoilla yksilöillä ei voida havaita, mikä vaste olisi sekä intervention että kontrollikäsittelyn jälkeen.

4 / 32

Tilastotiede ja vaikuttavuuden arviointi

Lyhyt yhteinen historia

A. Tutkimuksen asetelmat

- Kokeiden suunnittelu ja analysointi, 1920- (Fisher ym.)
- Kliiniset kokeet, peräkkäistestit, 1950- (A.B. Hill ym.)
- Monikeskuskokeet, meta-analyysi, 1980- (R. Peto ym.)

B. Tilastollinen päättely

- Bayesiläinen (oikeasti klassinen!) koulukunta, 1800-
- Frekventistinen: Fisherin ja NPW- koulukunnat, 1920-
- Bayes-päättelyn renessanssi, 1990-

5 / 32

Tilastomenetelmät vaikuttavuustutkimuksessa

(A) Suunnittelu

- Asetelman määrittely
- Tutkimuspopulaation valinta
- Interventoryhmiin jako, satunnaistus
- Aineiston koon ennakoarviointi

(B) Analyysi

- Datan tiivistäminen ja kuvailu
- Mallitus
- Tilastollinen päättely

6 / 32

Asetelman määrittely – satunnaistettu koe

- Yksilö- vai ryvässatunnaistus
 - hoitointerventiot arvotaan aina yksilötasolla,
 - ehkäisyinterventiot (esim. seulonta, terveyskasvatus) usein ryvästasolla (esim. kunta, koulu, luokka).
- Rinnakkaisten ryhmien vai vaihtovuorokoe
 - tyypillisesti rinnakkaisten ryhmien koe.
- Yhden tekijän vai monitekijäkoe
 - useimmiten vain yksi koetekijä,
 - vrt. SETTI, 2 tekijää: α -tokoferoli ja β -karoteeni.
- Osittamaton vai ositettu satunnaistus
 - koko joukon kesken vai erikseen potilasryhmittäin,
 - monikeskustutkimus: kukin keskus oma ositteensa.

7 / 32

Tutkimuspopulaation valinta

- Kelpoisuus- ja poissulku ehdot
 - Potilaiden lähde: paikka ja aika
 - Valinta
 - satunnaistonta vai peräkkäisten potilaiden poiminta,
 - rajaukset ja ositukset (tilastollinen tehokkuus)
 - Tarvittavien ryhmäkokojen arviointi
 - sisällöllisesti merkittäväväksi katsottu vaikuttavuus δ_0 ;
 - esim. vähimmäisriskiero hoitojen välillä tai max-NNT,
 - vasteen luonnollinen vaihtelu σ (teor. hajonta),
 - vaadittu tilastollinen tarkkuus tai osuvuus:
 - * "alfavirheen" todennäköisyys α , esim. 0.05,
 - * voimakkuus $1 - \beta$, esim. 0.90.
- NB.** Tätä pitäisi ajatella vähemmän testikeskeisesti.

8 / 32

Ryhmiiin jako – satunnaistus

(Huom! EI "randomisaatio" EIKÄ "randomointi"!)

- Hyvin dokumentoitu, toistettavissa oleva proseduur.
- Perustuu luotettavaan *satunnaislukugeneraattoriin* (RNG) lähtien annetusta siemenluvusta.
- *Satunnaistuslista*: RNG:n tuottama jono koodeja, joiden mukaan potilaat tulojärjestyksessä jaetaan hoitoryhmiin.
- Satunnaistuslista ja sen tekoprosessi pitää naamioida
 - potilailta,
 - tutkimusryhmän jäseniltä, potilaiden värväajiltä, hoitohenkilöiltä ja vasteen arvioijilta.
- Koodit suljetuissa kuorissa; avataan värväyksen jälkeen. Ks. CONSORT-ryhmän ohjeet, Moher ym. (2001)

9 / 32

Tilastollinen analyysi ja päättely

- (A) Aineiston analyysin vaiheet
- välianalyysit (*interim analyses*)
 - kokeen kuluessa ja uusia potilaita vielä värvättäessä,
 - päätössäännöt kokeen lopettamiseksi tai jatkamiseksi
 - lopullinen analyysi
- (B) Päättelyn frekventistiset menetelmät
- piste-estimointi,
 - tilastollinen testaus ja *P*-arvot,
 - väliestimointi
- Menetelmät ja toteutus päätetty tutkimusprotokollassa

10 / 32

Esim. GREAT-koe (BMJ 1992)

P: sydäninfarktin saaneet henkilöt,
I: varhainen liotushoito jo kotona,
C: sairaalaan vienti ja käypä hoito siellä,
O: kuolema 30 pv kuluessa.

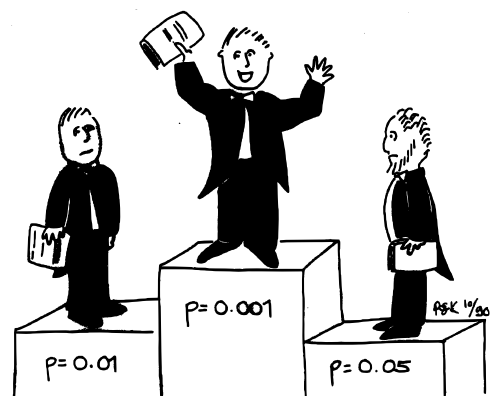
30 pv kuolleisuus, *I*: 13/163 = 8.0% vs. *C*: 23/148 = 15.5%

- RD = -7.5%-yks [95% CI: -15, -0.4], NNT = 13.2,
- RRR = 49% [2%, 73%], *P* = 0.04
- OR = 0.47, SE[log(OR)] = 0.37

Mitä nämä tulokset kertovat jo kotona aloitetun liotushoidon vaikuttavuudesta?

11 / 32

$P < 0.05$ – GREAT!



12 / 32

Frekventistisen analyysin ja päättelyn ongelmat

- P -arvon informatiivisuus ja luottamusvälin tulkinta?
 - Vastaavatko nämä kiinnostaviin kysymyksiin?
 - P -arvo yliarvioi näyttöä H_0 :a vastaan.
- Lopputuloksista tehtävien päätelmien riippuminen suoritettujen välianalyysien määrästä.
 - "Monitestausongelma", α :n "inflatoituminen"?
 - Miten näyttö voi riippua "kurkistusten" määrästä?
- Muun relevantin tiedon ja näytön ottaminen huomioon epävarmuuden määrällisessä arvioinnissa kokeen jälkeen.
 - Ei onnistu, vaikka tarvetta on.
 - Sama P implikoi eri asioita eri aineistoissa.

13 / 32

Kliinisten hoitokokeiden epidemiologia

Vain pieni osa tutkittavista uusista hoitokäytännöistä on vaikuttavampia kuin kulloinkin käypä hoito. Olkoon näitä 7%.

Uutta hoitoa testaava koe, otoskokolaskelmat:

- $\alpha = 5\%$, väärän positiivisen todennäköisyys,
- $1 - \beta = 90\%$, voimakkuus.

Koe tuottaa "merkittävän tuloksen" ts. $P < 0.05$.

– Mikä on tämän ennustearvo?

Bayesin kaava \Rightarrow "Merkittävän" tuloksen jälkeinen

Posterioritodennäköisyys, että koehoito on vaikuttava:

$$\frac{0.07 \times 0.90}{0.07 \times 0.90 + (1 - 0.07) \times 0.05} = 58\%$$

14 / 32

Esim. GREAT-koe (jatkuu)

Tulosten tulkinta ja päätelmät vaikuttavuudesta.

"*Liotushoito jo kotona alentaa kuolleisuutta liki 50%*"?

Vaikka tämä onkin RRR:n suurimman uskottavuuden estimaatti, tulos on liian hyvä ollakseen totta!

Relevantin ulkoisen informaation pohjalta asiantuntijat olisivat nimittäin voineet arvioida ennen koetta:

"15% – 20% väheneminen kuolleisuudessa on etukäteen uskottava (*plausible*), mutta $\leq 0\%$ tai $\geq 40\%$ alenema epäuskottava (*unlikely*)."

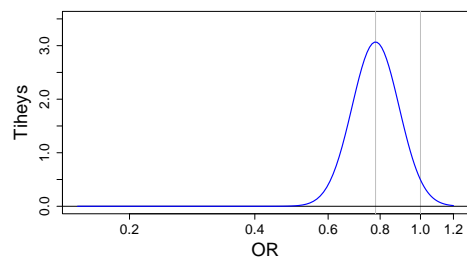
Kuinka ottaa tämä huomioon kokeen tuloksen tulkinnassa?

15 / 32

GREAT: Parametrin priorijakauma

Asiantuntijoiden ennakkokäsitys \Rightarrow **priorijakauma** OR:lle

$$\log(\text{OR}) \sim N(\log(0.78), 0.13^2)$$

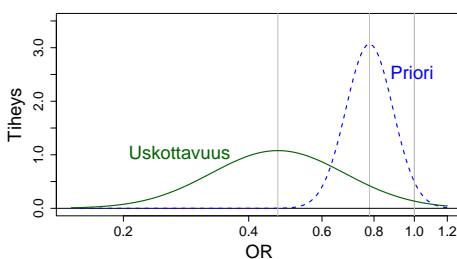


16 / 32

GREAT: Priori ja uskottavuusfunktio

Koetulokset \Rightarrow OR:n **uskottavuusfunktio** (likimääräinen)

$$\propto \exp\left\{-0.5[\log(\text{OR}) - \log(0.47)]^2 / 0.37^2\right\}$$

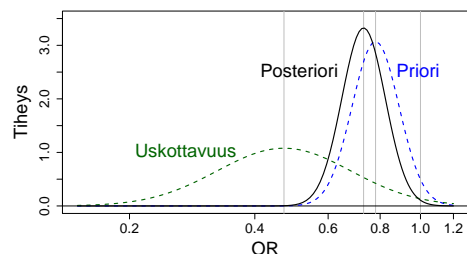


17 / 32

GREAT: Priori, uskottavuus, ja posteriori

Bayesin kaava \Rightarrow OR:n **posteriorijakauma** (likimääräinen)

$$\log(\text{OR}) \sim N(\log(0.73), 0.12^2)$$



Posteriori-tiheys sille, että kotiliotus pienentää kuolleisuutta, eli $\text{OR} < 1$, on 0.995. Loppuarvio vaikuttavuuden odotetusta suuruudesta silti havaittua piste-estimaattia pienempi.

18 / 32

GREAT: Skeptinen priori ja posteriori

Skeptisen klinikon H_0 : "kotona aloitetun liotuksen ei voi odottaa vaikuttavan kuolleisuuteen" $\Leftrightarrow E[\log(\text{OR})] = 0$.

Lisäksi ennakoarvioon sisältyy 95% epävarmuusväli OR:lle 0.5 – 2. Nämä lähtökohdat vastaavat priorijakaumaa

$$\log(\text{OR}) \sim N(\log(1.0), 0.35^2)$$

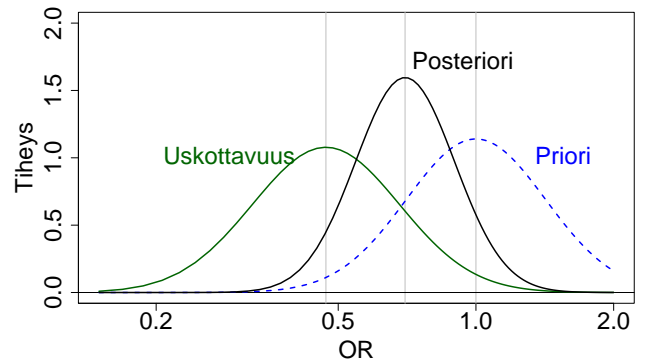
Yhdistettynä samoihin koetuloksiin posteriorijakaumaksi tulee

$$\log(\text{OR}) \sim N(\log(0.7), 0.25^2)$$

Posteriori-tn, että $\text{OR} \geq 1$, on 0.08. Kokeen antama näyttö ei siis aivan vakuuta skeptikkaa.

19 / 32

GREAT: skeptinen priori ja posteriori



20 / 32

Bayes-päätely

- Vaikuttavuutta koskeva epävarmuus ilmaistaan todennäköisyyksinä ja -jakaumina ennen ja jälkeen kokeen.
- Bayesin kaava päivittää ennakkokäsityksen/epävarmuuden uuden kokeen tulosten antamalla näytöllä.
- Päivitetty posteriori toimii priorina seuraavan empiirisen aineiston analyysissä.
- Ei ole yhtä oikeata priorijakaumaa, mutta on tarpeen selvittää, kuinka paljon päätelmät riippuvat priorista.
- Kun pätevää empiiristä näyttöä kertyy tarpeeksi, niin päätelmät ovat samat riippumatta priorikäsityksistä.

21 / 32

Priorikäsitysten muotoilu – erilaisia prioreja

Kelpoisten priorijakaumien heruttaminen (*elicitation*) klinikoilta ym. sisältöasiantuntijoilta ei ihan helppoa.

- *Innostuneen klinikon prior*: Heijastaa optimistista käsitystä uuden hoidon vaikuttavuudesta
- *Skeptikon prior*:
Vaikuttavuuden odotusarvo 0 ja hajonta ei kovin leveä.
- *Konsensusprior*: Neuvotellaan yhdessä.
- *Lattea, hämärä (flat, vague) t. "epäinformatiivinen" prior*: Odotusarvo 0 ja erittäin suuri hajonta.
 - Tarjoaa suuria todennäköisyyksiä vaikuttavuusparametrin arvojen täysin epärealistisille osajoukoille.
 - Voi olla hyvin informatiivinen parametrin muunnokselle.

22 / 32

Eri kokeiden tulosten yhdistely – meta-analyysi

Frekventistiset menetelmät toimivia yksittäisen, riittävän suuren kokeen tuloksiin liittyvän epävarmuuden tiivistämisessä.

Usein tarjolla on eri puolilla tehtyjä pieniä kokeita, joiden tuloksissa suuri satunnaisvaihtelu. Entä jos kaikkien näiden piste-estimaattit uutta hoitoa puoltavia mutta "ei-merkitseviä"?

Meta-analyysi: Lasketaan koekohtaisista luvuista yhdistetty estimaatti (& luottamusväli) vaikuttavuusparametrille:

- *kiinteän vaikutuksen* malli: usein epärealistinen, ei ota huomioon kokeiden välistä heterogeenisuutta,
- *satunnaisvaikutusten* malli: frekventistiset menetelmät vielä keskeneräisiä.

23 / 32

Esim. Magnesium ja sydäninfarkti

1980-luvulla 7 pientä koetta, joiden meta-analyysi (1991):

- Magnesiumilla hoidetuilla potilailla ($n_1 = 657$) sairaala/30pv-kuolleisuus oli 54% pienempi kuin lumeryhmillä ($n_0 = 644$), 3.8% vs. 8.2%
 $\text{OR} = 0.45$, 95% CI 0.28–0.71 (kiinteä vaikutus)

1992 LIMIT-2-koe ($n_1 \approx n_0 \approx 1160$):

- kuolleisuus 7.8% vs. 10.2%; RRR = 24%
 $\text{OR} = 0.74$, 95% CI 0.56–0.98

1990-luvulla 6 uutta pientä koetta.

1995 ISIS-4-megakoe ($n_1 \approx n_0 \approx 29000$):

- kuolleisuus 7.6% vs. 7.2%; RRR = –5.5%
 $\text{OR} = 1.06$, 95% CI 0.996–1.13

24 / 32

Magnesium ja sydäninfarkti

Mistä pienten kokeiden ja megakokeen tulosten ristiriita?

Higgins ja Spiegelhalter (*Int J Epid* 2002; **31**: 96-104):

"Pienten, julkaisukynnyksen ylittäneiden kokeiden tulosten innostunutta tulkintaa olisi viilentänyt

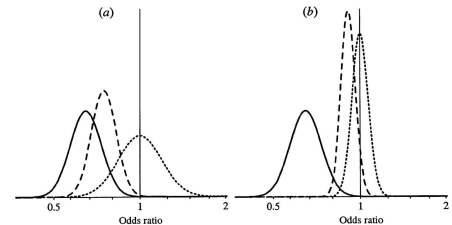
- riittävän skeptinen priorii,
- satunnaisvaikutusten mallin käyttö kiinteän sijaan,
- havaitun vaikutuksen ja kokeen koon välisen yhteyden tarkastelu – julkaisuharhan mahdollisuuden huomiointi."

Bayesiläinen meta-analyysi, jossa nämä otettiin huomioon, poisti havaitun ristiriidan.

25 / 32

Magnesium ja sydäninfarkti

8 ensimmäisen kokeen (ml. LIMIT-2) meta-analyysi kiinteän vaikutuksen mallilla: Kaksi eri tavalla skeptistä prioria



Pisteviiva = priorijakauma, yhtenäinen = uskottavuus, katkoviiva = posteriorijakauma

Higgins, J. P. et al. *Int. J. Epidemiol.* 2002 31:96-104;

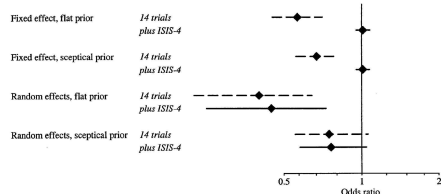
International Journal of
Epidemiology

Copyright restrictions may apply.

26 / 32

Magnesium ja sydäninfarkti

ISIS-4:n vaikutus eri oletuksien tehtyjen meta-analyysien tuloksiin. Mukana kaikki 14 pientä koetta (ml. LIMIT-2)



Higgins, J. P. et al. *Int. J. Epidemiol.* 2002 31:96-104

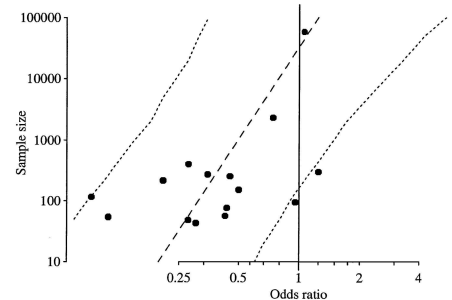
International Journal of
Epidemiology

Copyright restrictions may apply.

27 / 32

Magnesium ja sydäninfarkti

Estimoidun OR:n ja kokeen koon välinen yhteys (kaikki 15 koetta, ml. ISIS-4)



Higgins, J. P. et al. *Int. J. Epidemiol.* 2002 31:96-104;

International Journal of
Epidemiology

Copyright restrictions may apply.

28 / 32

Vaikuttavuuden arviointi ja epäkokeellinen tutkimus

Satunnaistettuja ja riittävän isoja kokeita voi käyttää vain rajattuun osaan terveydenhuollon vaikuttavuuskysymyksiä.

Useimpiin ongelmiin löytyy vain epäkokeellista dataa, jonka pohjalta tehtäviä vaikuttavuuden arvioita haittaavat

- sekoittuneisuus (*confounding*); eri interventioiden kohteeksi joutuvat ryhmät eivät ole vertailukelpoisia,
- puuttuvat, epäsuorat ja virheelliset havainnot ja mittaustulokset

Päätöksiä on tehtävä datan puutteista ja näytön harhaisuudesta huolimatta. Miten siis menetell?

29 / 32

Vaikuttavuuden arviointi ja epäkokeellinen tutkimus

Frekventistinen tilastotiede tarjoaa niukasti työkaluja.

Bayesiläisestä paradigmasta käsin on alettu kehittää malleja ja menetelmiä, joilla pyritään

- yhdistämään erilaatuisten aineistojen informaatio,
- ottamaan piilevät harhat huomioon niitä koskevin eksplisiittisin osamallein, jotka sisältävät realistisiksi arvioitavia priorijakaumaoletuksia harhojen suuruudesta.

Eri osamalleista kootaan **yleistetty näytön synteesi**. Kokonaisuudessaan mutkikas ja haasteellinen tehtävä!

Ades & Sutton (2006). Multiparameter evidence synthesis in epidemiology and medical decision making. *JRSS A* **16**: 5-35.

30 / 32

Pari kriittistä kysymystä – ja vastaukset

1. *Priorin tuoma epäobjektiivisuus – ongelma?*

- Subjektiivisuus aina mukana inhimillisen tiedon hankinnassa ja päätelmien teossa ml. tilastanalyysit.
- Bayes-päätelyssä sitä ei kätketä vaan esitetään avoimesti ja selvitetään eri priorien seuraukset.
- Päätöstilanteessa eri osapuolilla muutenkin erilaiset utiliteetit ym. subjektiiviset lähtökohdat.

2. *Vahva informatiivinen prior ja satunnaistus – ristiriita?*

- Vaikka innostunut klinikko olisi hyvinkin vakuuttunut uuden hoidon paremmuudesta, myös skeptikoiden perustellut ennakkokäsitykset pitää ottaa huomioon.
- Pätevin tapa päästä kaikkia vakuuttavaan posterioriin on riittävän suuri satunnaistettu koe.

31 / 32

Loppupäätelmiä (Spiegelhalter ym.)

Bayesiläinen lähestymistapa on frekventististä

- joustavampi mukautumaan eri sovellustilanteisiin,
- tehokkaampi käyttämään kaikkea olevaa tietoa ja näyttöä,
- hyödyllisempiä työkaluja päätöksentekijälle tarjoava,
- eettisempi, koska ei ole sokea datan ulkopuoliselle tiedolle.

Bayesiläinen analyysi on frekventististä vaativampaa:

- priorijakaumien määrittely, tulosten herkkyysoanalyysi,
- kompleksisten hierarkkisten mallien rakentaminen,
- posteriorijakaumien simulointi & monitorointi.

Terveydenhuollon todellisuus on itsessään kompleksinen, joten hyvien vastausten saaminen ei voikaan olla yksinkertaista.

32 / 32