KYS

Näkökantoja tieteellisiin julkaisuihin

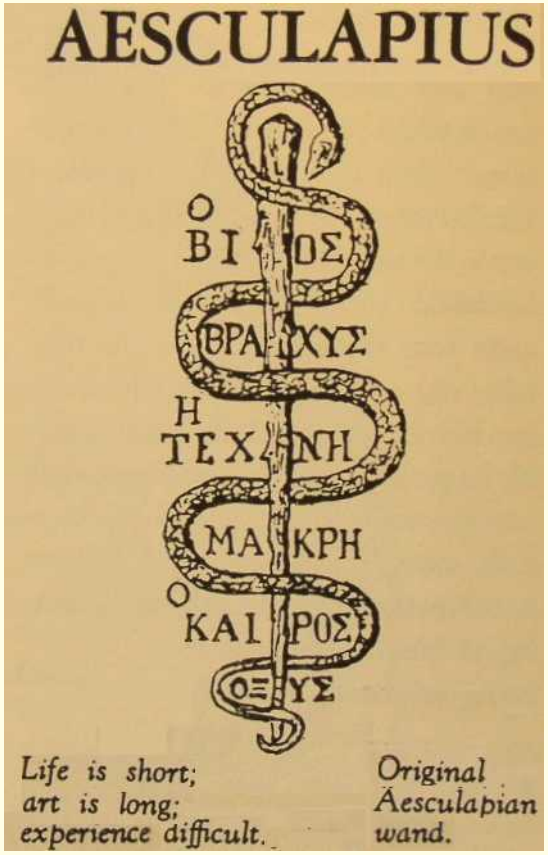
Vesa Kataja
ylilääkäri, professori
KYS, Syöpäkeskus

Erikoistuvien koulutuspäivät
25.-26.1.2013
Kuopio

Suurin osa julkaistusta tutkimustiedosta on epäluotettavaa!

Ars longa, vita brevis, occasio praeceps, experimentum periculosus, iudicatio difficile

Taide pitkä (ikuinen),
Elämä lyhyt,
Mahdollisuus hetkellinen,
Koe (kokemus)
hengenvaarallinen
Arviointi vaikeata



Role of gemcitabine-based combination therapy in the management of advanced pancreatic cancer: A meta-analysis of randomised trials

Domenico Ciliberto [a,e], Cirino Botta [a,e], Pierpaolo Correale [b], Marco Rossi [a], Michele Caraglia [c], Pierfrancesco Tassone [a,d], Pierosandro Tagliaferri [a,*]

European Journal of Cancer (2013) **49**, 593–603

Interpretation: The combination chemotherapy as compared to gemcitabine alone significantly improves OS in advanced pancreatic cancer (APC). However, this advantage is marginal whereas the treatment-related toxicity is increased, suggesting the use of gemcitabine-based combination regimens only in selected patient populations. New prospective trials, based on translational approaches and innovative validated biomarkers, are eagerly awaited on this topic.

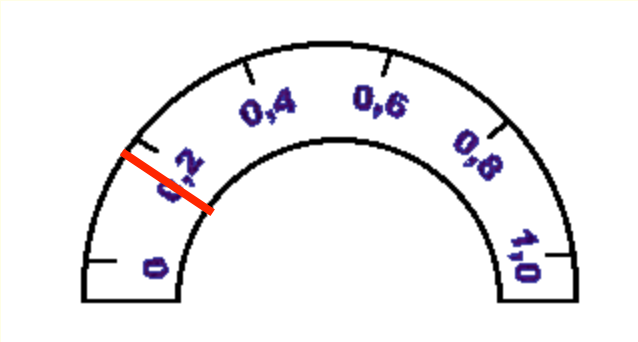
OS HR = 0.93 (95% CI 0.89 – 0.97); p=0.001

RR OR = 0.60 (95% CI 0.47 – 0.76); p<0.001

DCR OR = 0.79 (95% CI 0.66 – 0.93); p=0.006

Klassinen todennäköisyys

- Nopassa on kuusi eri silmälukua. Jokainen niistä on yhtä todennäköinen
- Todennäköisyys saada vaikkapa kuutonen on yksi mahdollisuus kuudesta



1/6 = 0,17 = 17 %

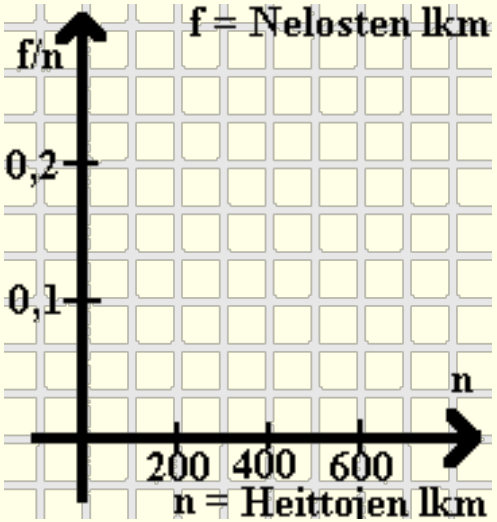
Tilastollinen todennäköisyys

- Klassisessa todennäköisyydessä kaikki ns. alkeistapahtumat ovat YHTÄ todennäköisiä
- Tilastollisessa todennäköisyydessä käytetään tilastollisia **KESKIMÄÄRÄISYYKSIÄ** apuna todennäköisyyksiä laskettaessa

Tilastollinen todennäköisyys

Heitetään noppaa useita satoja kertoja ja lasketaan vaikkapa nelosten lukumäärä. Klassisessa todennäköisyydessä nelosen todennäköisyys on 1/6 = 0,167 = 16,7 %.

Laskettaessa nelosten lukumäärän f suhdetta heittojen lukumäärään n eli f/n havaitaan, että tämä tilastollinen todennäköisyys **LÄHENEE** klassista todennäköisyyttä 16,7 % sitä mukaa, kun heittojen lukumäärä kasvaa.



Todennäköisyys sille, että tutkimuksen tulos on totuuden mukainen (oikein), riippuu

- Tutkimuksen voimasta (1-β)
  - todennäköisyys, että tulos on *tilastollisesti merkitsevä* (eli H_0 hylätään), kun vastahypoteesi (H_1) on totta

	Nolla hypoteesi (H_0) totta	**Nolla hypoteesi (H_0) ei totta**
Hylkää (H_0)	Tyypin I virhe Väärä positiivinen	Oikea lopputulos Todellinen positiivinen
Hyväksy (H_0)	Oikea lopputulos Todellinen negatiivinen	Tyypin II virhe Väärä negatiivinen

Tilastollisella testillä voidaan vain joko hyväksyä tai hylätä nollahypoteesi, ei koskaan osoittaa sitä todeksi!

Tutkimuksen tilastollinen voima = 1 – β

- Todennäköisyys sille, että löydetään ero, joka oikeasti on olemassa
  → yleisimmin käytetty 0.80
- Riippuu monista tutkimusasetelman tekijöistä, yleensä pyritään hallitsemaan vähintään kolme:
  (1) vaikutuksen koko / määrä tutkimusjoukossa / eron todellinen suuruus
  (2) jakauma / varianssi / SD [95% CI = +/- 1.96 x SD / √n]
  (3) Tutkimusjoukon n / otoksen koko
  → tarvittava tutkimusjoukon koko voidaan laskea tietyn eron luotettavaan osoittamiseen
- Testin tilastollinen merkitsevyys
  - kuinka epätodennäköinen positiivisen tuloksen täytyy olla, jotta H_0 = ”ei vaikutusta”, voidaan hylätä
  - yleisimmin käytetty 1/20 = 0.05
  - voimaa voidaan nostaa käyttämällä esim. 0.10 → H_0 hylkäämis-todennäköisyys kasvaa, eli tilastollisesti merkitsevän tuloksen saamisen todennäköisyys kasvaa H_0 ollessa väärä
    - tyyppi II virheen riski pienenee (väärä negatiivinen)
    - tyyppi I virheen riski kasvaa (väärä positiivinen)

Todennäköisyys sille, että tutkimuksen tulos on totuuden mukainen (oikein), riippuu

- Positiivisesta ennustearvosta (PPV)
  - *pre-study probability* – aikaisemmat vasteet hoidolle
  - *post-study probability* – arvioidaan tai arvataan
- Tärkeämpää, kuin tilastollisen merkitsevyyden metsästys, on arvioida etukäteen millä todennäköisyydellä yritetään saada esiin todellinen yhteys: hoito → vaste

Todennäköisyys sille, että tutkimuksen tulos on totuuden mukainen (oikein), riippuu

- Samaa hypoteesia selvittävien tutkimusten määrästä
  - mitä useampi tutkimus samasta aiheesta, sitä todennäköisemmin ainakin yksi tuottaa tilastollisesti merkitsevän tuloksen
  - **KÄÄNTÖPUOLI**: yksi positiivinen tutkimus saa kaiken huomion – myöhemmät negatiiviset tulokset jäävät tämän varjoon (tai jopa julkaisematta)
- Harhasta
  - syntyy erilaisista tutkimuksen suunnitteluun, tietoon, analysointiin ja esittämiseen liittyvistä tekijöistä, jotka ilmaistaan tai jätetään ilmaisematta tutkimuksen tuloksissa
    - esimerkki = selektiivinen raportointi

Tutkimuksen tuloksen luotettavuutta heikentävät

- Tutkimuksen aineiston pieni koko
  - → pieni aineisto = pieni tilastollinen voima
  - → mutta, jos ero on tarpeeksi suuri……
- Toimenpiteen vaikutuksen vähäisyys
- Testattavien yhteyksien suuri määrä (tai huono valinta)
- ”Joustavat” tutkimuksen suunnittelu, määritelmät, tulosmuuttujat ja analyysit
  - → esim. erilaiset PFS:n määritelmät vs kuolema

Tutkimuksen tuloksen luotettavuutta heikentävät

- Tulokseen vaikuttavat taloudelliset ja muut intressit ja ennakko-odotukset
- Paradoksaalisesti: tutkimusaiheen ajankohtaisuus – *”kuumuus”*
  - läpimurtotulosta seuraa pian aivan vastakkainen tulos *(”Proteus ilmiö”)*

PROTEUS ILMIÖ - esimerkki

- Ennen todella tehoavien antiviraalisten HIV lääkkeiden keksimistä:
  - 25 RCT:ssa identifioitiin tilastollisesti merkitsevä vaikutus HIV-potilaiden elossaoloon
  - Meta-analyysit julkaistuista tutkimuksista tukivat yleisesti ja vähemmän yleisesti hyväksyttyjen, ja nykyisin jopa kontraindikoitujen interventioiden vaikuttavuutta
  - Nyt, retroskooppisesti katsoen, useat näistä vaikuttavista hoidoista (immutioli, interferoni, immunoglobuliini) eivät vakuuta!
  - HIV potilaille tarjottiin täysin falskeja hoitoja, kunnes oikeasti vaikuttavat tulivat tarjolle.

Merkitsevyyden metsästys CAVE! $p < 0.05$

- Alun perin tilastollinen indeksi, joka kuvasi millä varmuudella nollahypoteesi voidaan hylätä *(Fisher, 1925)*
- $p=0.05$ (todennäköisyys 1/20 että, tulos on sattuman aikaansaamaa) valittu käytännöllisenä rajana
- $p<0.05$ ei tarkoita todennäköisyyttä, millä tutkimuksen tulos on ”totta”
  - parempi estimaatti on luottamusväli, joka kertoo minne (jatkumolla) tutkimuksen todellinen tulos asettuu

p-arvoon ei saa luottaa!

- Tutkimuksessa ilmoitettu p:n arvo, ei ole tutkimuksen tulos, vaan tutkimuksen työkalu!!
- Mikä tahansa triviaali ero ryhmien välillä on tilastollisesti merkitsevä, jos kasvate-taan aineiston kokoa

Pelkän p-arvon perusteella julkaistaan!

Box 2. Refuting a *Nature* Cover Page Story and a *p*-Value of 9×10^{-16}

In 1994, the cover page of *Nature* announced the discovery of the osteoporosis gene. A study of a few hundred subjects claimed that polymorphisms in the vitamin D receptor (*VDR*) gene could explain 75% of the genetic variability of bone mineral density [18]. This paper has received more than 1,000 citations to date. Three years later, the research team published an erratum in *Nature* acknowledging a laboratory error. The revised results showed a weaker, but still formally statistically significant association. Over 100 studies, mostly of small sample size, were performed trying to replicate this association. Several meta-analyses tried to synthesize the published data and concluded that the association was statistically significant. One of them [19] reached a $p = 9 \times 10^{-16}$ by comparing the proportion of studies with formally statistically significant results against an "expected" 5%— a questionable approach. A subsequent meta-analysis of individual level data from a small subset of studies also found an odds ratio of 4 for osteoporosis [20]. However, recent studies questioned the association. A very large study found a statistically significant association, but in the very opposite direction [21]. Finally, a large collaborative study with sample size about 100-fold larger than the original *Nature* study did not detect any association with either bone mineral density or fractures. All odds ratios in the main analyses were between 0.98 and 1.02, and not even 1% of the genetic variability of bone mineral density could be explained [52]. Nevertheless, this may not be the end of the story. *VDR* is a long gene with many haplotype blocks. In general, as the amount of data increases exponentially (e.g., genomic testing in biobanks with linkage to health outcomes), the potential for both discoveries and errors also increases exponentially.

Milloin onkologinen tutkimustulos on ”positiivinen”?

- PFS ja OS hyväksyttyjä pääte-tapahtumia
- PFS surrogaatti OS:lle; ainakin joissakin tuumoriryhmissä
  - PFS (TTP) ”määritelmäherkkä”, ”proxy”
  - OS = kovin mittari; sekään ei ongelmaton
    - tutkimuksellisen hoidon jälkeinen mahdollinen muu aktiivinen hoito, esim. rintasyövässä
    - tutkimuksen aikainen / myöhempi X-over

Milloin onkologinen tutkimustulos on ”positiivinen”?

- RCT:ssa tulee etukäteen määrittää tavoiteltava ero, δ, primaarimuuttujassa tutkimusryhmän ja kontrolli-ryhmän välillä
  - tästä riippuu tutkimusjoukon koko sekä tutkimuksen kesto (→ kustannukset)
  - δ tulisi kuvastaa kliinisesti merkittävää eroa (myös huomioiden lääkityksen siedettävyys ja toksisuus)
  - mielekkään eron suuruudesta monia mielipiteitä
    - onkologi, potilas, omaiset, yhteiskunta

Milloin onkologinen tutkimustulos on ”positiivinen”?

- Riippumatta δ:sta, tulos voi olla tilastollisesti merkitsevä, vaikka paras arvio eron suuruudesta jää pienemmäksi, kuin etukäteen arvioitu δ
- Tällaisten tulosten julkaiseminen hyvin yleistä
- Saattaa johtaa ”outoihin” tilanteisiin:
  - erlotinibi haimasyövässä: tavoite 25 %:n riskin vähenemä, saavutettiin 18 % (HR 0.82, 95%CI 0.69-0.99; p=0.038), mediaani survival ero = 10 päivää; mediaani PFS ero= 6 päivää *(Moore et al., 2007)*
  - lääke sai FDA:n hyväksynnän!

Milloin onkologinen tutkimustulos on ”positiivinen”?

- Kliinisesti merkittävä ero?
  - PFS ≥ 4 – 6 kk; HR n. 0.5
  - OS ≥ 3 kk; HR n. 0.75
- Tilastollisesti merkittävä ero
- Hoidon toksisuus huomioitava!

Odds ratio (OR)? Hazard ratio (HR)? Mitä ihmettä?

- Piste-estimaatteja = vertailevan tutkimuksen”data” yhdellä numerolla ilmaistuna
- Piste-estimaatin käyttö riippuu muuttujista
  - OR → binaarisissa muuttujissa
  - HR → Aika-Tapahtuma (survival) muuttujissa
- Onkologisissa tutkimuksissa yleisesti
  - OR
  - HR
  - Relatiivinen riski (RR)

Odds Ratio (OR) esimerkki

Age, Sex, and Racial Differences in the Use of Standard Adjuvant Therapy for Colorectal Cancer", Potosky, Harlan, Kaplan, Johnson, Lynch. JCO, vol. 20 (5), March 2002, p.1192.

- Vaikuttaako sukupuoli tavanomaisen liitännäis-hoidon käyttöön (SAT) stage II/III rektaali- tai stage III kolonsyöpäpotilialla?
  - 53 % miehistä sai SAT
  - 62 % naisista sai SAT
- Miten arvioidaan ero?

Odds Ratio (OR) esimerkki

- Odds = p/(1-p)
- miehelle tulee SAT 0.53/(1 - 0.53) = 1.13
- naiselle tulee SAT 0.62/(1 - 0.62) = 1.63
- OR = 1.63/1.13 = 1.44
- OR = 1.13/1.63 = 0.69

Tulkinta: Nainen saa 1.44 x useammin (todennäköisemmin?) SAT:n kuin mies

Odds Ratio (OR) esimerkki

- Joskus käytetään *log* OR

log OR naiset:miehet on log(1.44) = 0.36
log OR miehet:naiset on log(0.69) = -0.36

log OR > 0: suurentunut riski
log OR = 0: riskeissä ei eroa
log OR < 0: pienentynyt riski

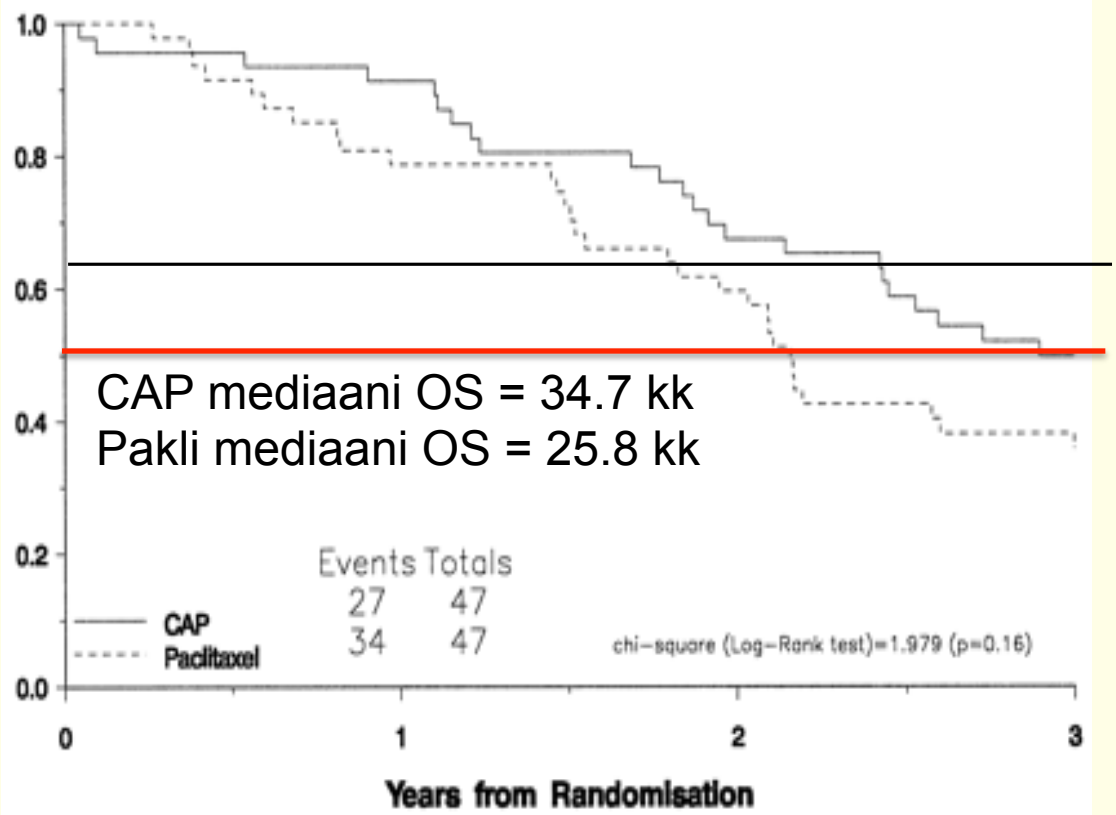
Relatiivinen riski (RR)

- RR = p_1/p_2
- RR = 0.62/0.53 = 1.17
- Käytännössä sama asia (tulkinta) kuin OR MUTTA:
  - p_1 = 0.50, p_2 = 0.25
  - RR = 0.5/0.25 = 2
  - OR = (0.5/0.5)/(0.25/0.75) = 3

Hazard Ratio - vaarasuhde

Mikä on CAP:n vaikutus kokonaiselossa-oloaikaan vertailussa paklitakseliin?

Mediaani survival ei kerro kaikkea!

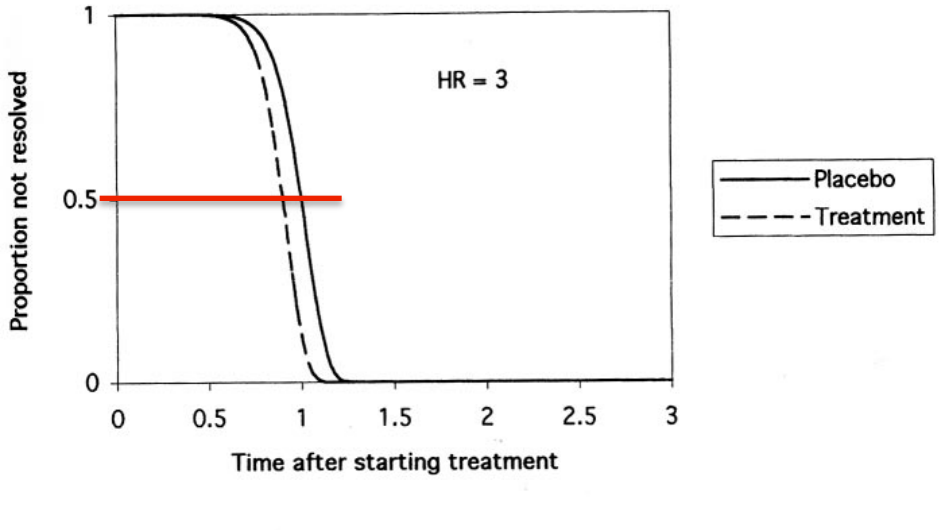


"Randomized Controlled Trial of Single-Agent Paclitaxel Versus Cyclophosphamide, Doxorubicin, and Cisplatin in Patients with Recurrent Ovarian Cancer Who Responded to First-line Platinum-Based Regimens", Cantu, Parma, Rossi, Floriani, Bonazzi, Dell'Anna, Torri, Colombo. JCO, vol. 20 (5), March 2002, p. 1232.

Hazard ratio – vaarasuhde

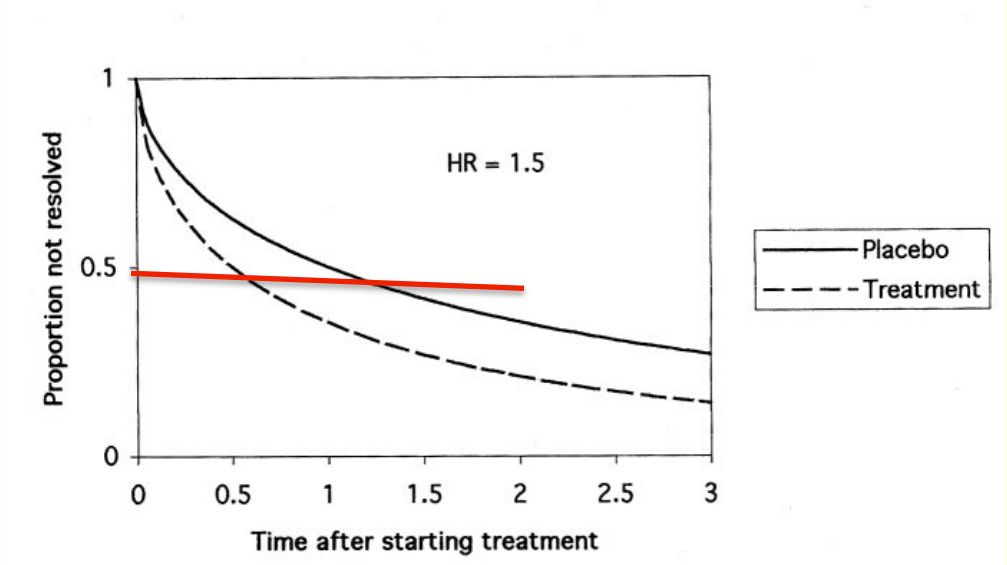
- HR vertaa riskejä kahdessa eri joukossa
- Riski1/Riski2 = suhteellinen riski
- Ajankohdasta riippumaton – riski on vakio ajan funktiona
- *Vaarafunktio = todennäköisyys kuolla ajanhetkellä t, jos pysyit elossa ajanhetkeen t*
- HUOM! Vaarasuhde (HR) ei ole survival käyrien välinen suhde!

Kummassa tapauksessa hoito on hyödyllisempää?



Hoidon hyödyllisyys ei riipu ainoastaan HR:sta, vaan myös taustalla olevista todennäköisyys-jakaumista, jotka ovat tautiriippuvaisia

HR = 3
mediaani survival
+ 10 %
HR = 1.5, mediaani survival + 50 %



NO 16966

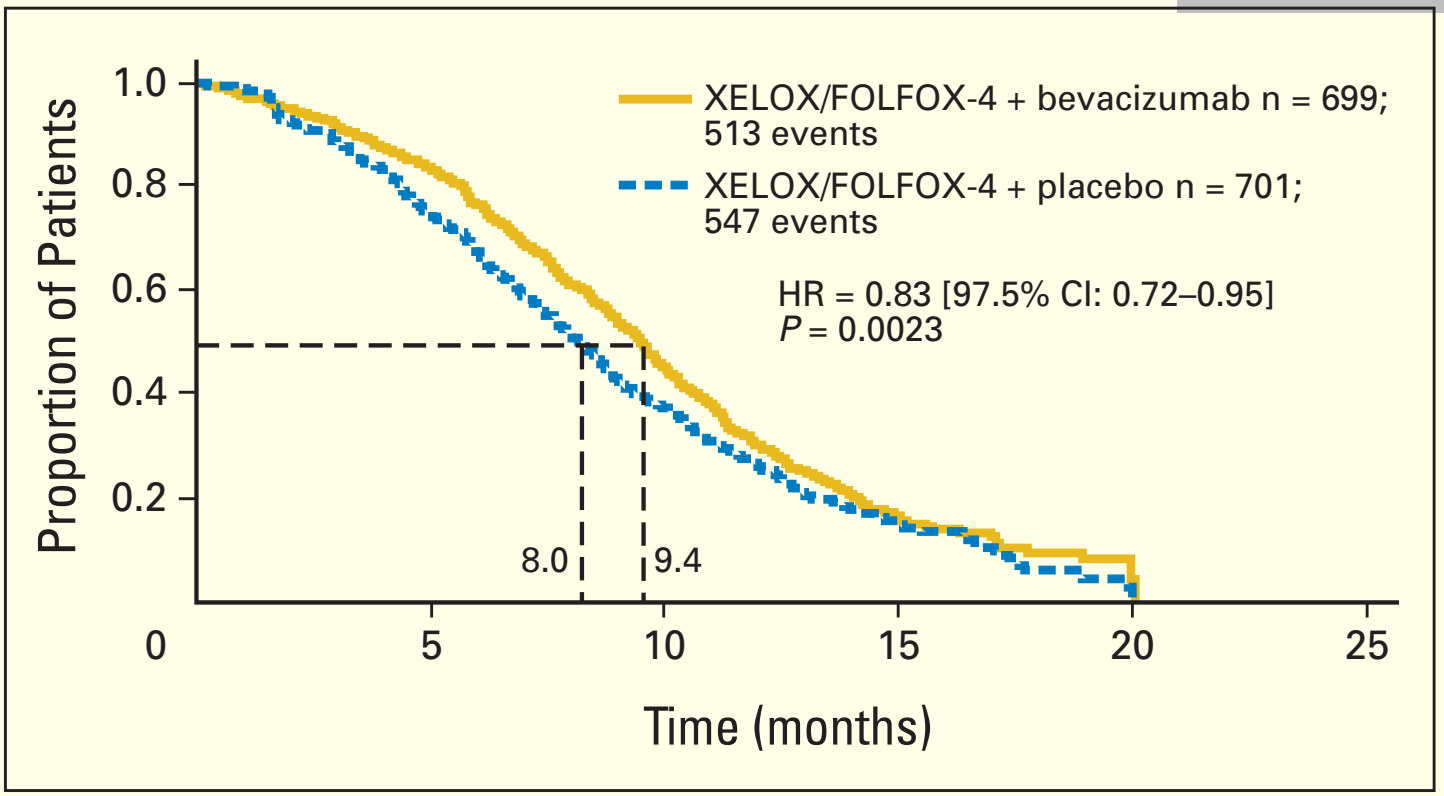


Fig 2. Progression-free survival (intent to treat population). XELOX, capecitabine and oxaliplatin; FOLFOX-4, infused fluorouracil, folinic acid, and oxaliplatin; HR, hazard ratio.

Potilaskysymykset

NO 16966

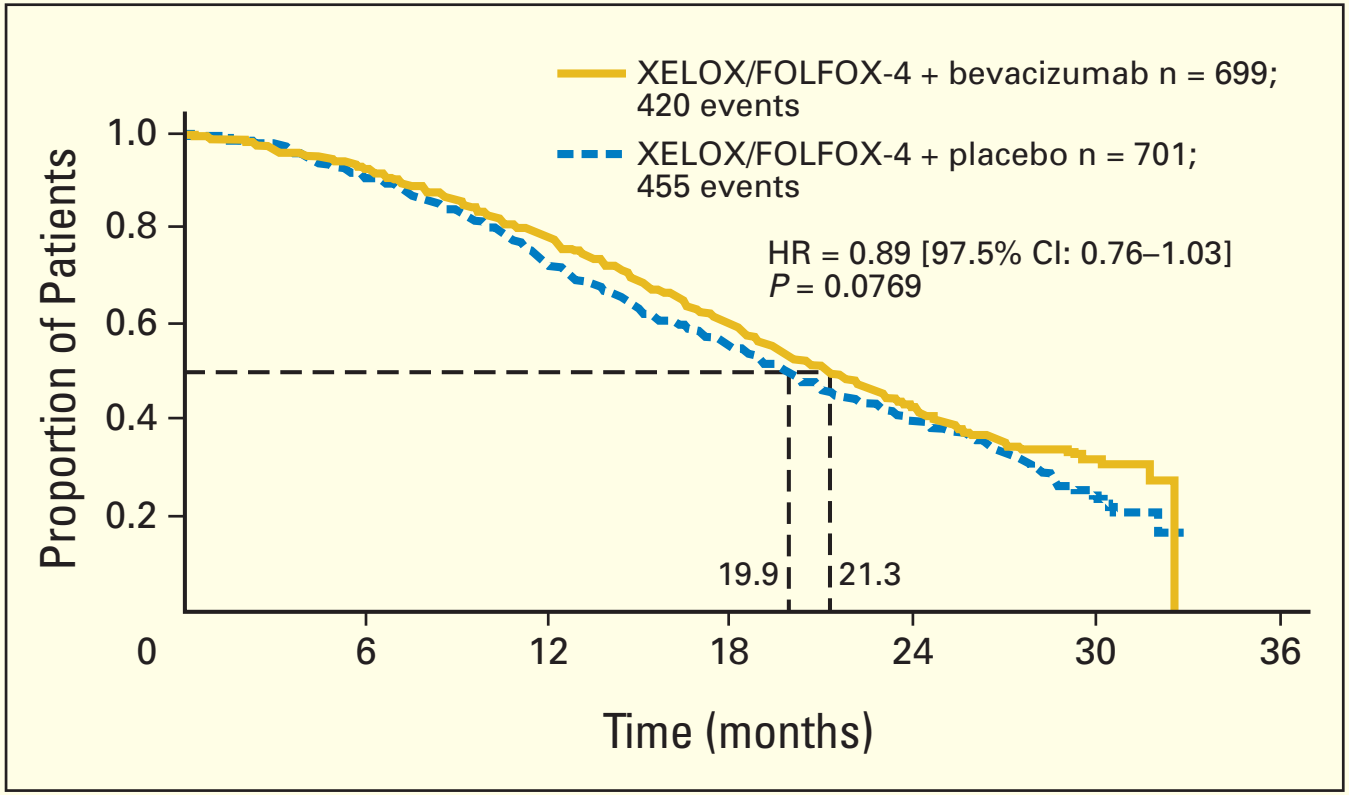


Fig 3. Overall survival (intent to treat population). XELOX, capecitabine and oxaliplatin; FOLFOX-4, infused fluorouracil, folinic acid, and oxaliplatin; HR, hazard ratio.

Ryhmätehtävä: edellä kerrottuun perustuen, jaetuista artikkeleista kertokaa minulle

- Mitä tutkittu?
- Onko kliinisen ongelman kannalta relevantti tutkimus?
- Vaikuttaako hoitokäytäntööni?
- onko potilasmateriaali kuvattu tarpeeksi hyvin?
- Soveltuuko omaan potilasjoukkooni?
- Tehdyn tutkimuksen laatu:
  - onko tilastolliset menetelmät kuvattu?
  - onko potilasmäärä ilmoitettu etukäteen?
  - onko lopullinen potilasmäärä ilmoitettu (onko kuvattu kaikki poisjätetyt)?
  - mikä on primaari päätetapahtuma?
    - Miten se on määritelty?
  - mikä on toissijainen päätetapahtuma (jne.jos muita)?
    - miten määritelty?
  - onko tavoiteltu ero (δ) ilmoitettu päätetapahtumien suhteen?
  - onko saavutettu ero (δ) ilmoitettu päätetapahtumien suhteen?
  - onko tavoiteltu HR ilmoitettu?
  - onko saavutettu HR ilmoitettu?

Ryhmätehtävä: edellä kerrottuun perustuen, jaetuista artikkeleista kertokaa minulle

- Tukevatko tulokset johtopäätöksiä?
- Tilastollinen merkitsevyys / kliininen merkitsevyys?
- Ovatko tutkimuksen tulokset lopulliset?
- onko haittavaikutukset raportoitu?
- mikä on NNT – luku?
- mikä on NNH – luku (jokin keskeinen haittatapahtuma)?
- Saavutetun tuloksen taloudellinen intressi
- Kirjoittajien sidonnaisuudet / riippumattomuus
- Oma mielipide tutkimuksesta
  - Riittävän hyvä
  - Tekisin paremman! Miten?

Absoluuttinen varmuus lääketieteessä on illuusio!

- Se että hoitosuosituksia on olemassa, on parempi vaihtoehto kuin se, että niitä ei olisi
- Täytyy muistaa, että kyseessä on vain suositus, joka ei missään tilanteessa pysty kokonaan korvaamaan koulutettua, kokenutta ja kriittistä kliinikon harkintaa potilaan parhaasta hoidosta
- Mielipiteelläkin on siis sijansa
- Kauhistuttavinta olisi ”keittokirjalääketiede”

Se, miten lääkäri itse sietää jatkuvaa epävarmuutta on yksi lääkärin ammattitaidon mitta

- Se, miten lääkäri itse sietää epävarmuutta heijastuu suoraan hoitotilanteeseen ja potilaan saamaan hoitoon
- Pahimmillaan lääkärin kykenemättömyys sietää epävarmuutta johtaa potilaan vahingoittamiseen – usein ylihoitamisen kautta
- Lääkäri ei pysty olemaan aktiivisesti hoitamatta tilanteissa, joissa tietoa on usein jopa liikaa, mutta aktiivisia hoitopäätöksiä tehdään lähinnä toiveajatteluun perustuen ja omaa epävarmuutta peläten ja väistellen
- Potilas kärsii

- Kaikkein luotettavimpana tukena hoitopäätöksen tekemiseen, kultaisena standardina, pidetään nykyisin tuloksia laajoista, kaksois-sokkoutetuista satunnaistetuista ja kontrolloiduista tutkimuksista

- Loppujen lopuksi myös nämä parhaimmillaankin tuottavat vastauksen ainoastaan tiettyyn kysymykseen tietyllä todennäköisyydellä ja tietyssä potilasjoukossa

- Tulokset eivät juuri ole yleistettävissä tutkimusjoukon ulkopuolelle

Lääketieteellinen hoito perustuu

- todennäköisyyksien arviointiin
- yritykseen hallita tiedon epävarmuutta erilaisin koejärjestelyin ja tilastollisin menetelmin
- mielipide-vaikuttajien mielipiteisiin
- hoitosuosituksiin
- lääkärin oman epävarmuuden sietämiseen