

Kausaalisuus ja tavoitekokeen jäljittely

Juha Karvanen

Matematiikan ja tilastotieteen laitos
Jyväskylän yliopisto



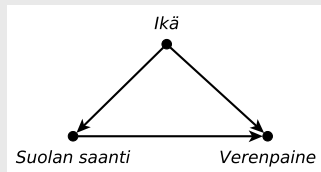
Kausaalisuuden hierarkia

Taso	Yksilökysymys	Väestökysymys
Ennuste	Mikä on riskini sairastua verenpainetautiin viiden vuoden aikana?	Kuinka suuri osuus väestöstä sairastuu verenpainetautiin seuraavan viiden vuoden aikana?
Interventio	Mikä on riskini sairastua verenpainetautiin, jos vaihdan vähäsuolaiseen ruokavalioon?	Kuinka suuri osuus väestöstä sairastuisi verenpainetautiin, jos kaikki noudattaisivat vähäsuolaista ruokavaliota?
Kontrafaktuaali	Olisinko voinut välttää verenpainetaudin, jos olisin noudattanut vähäsuolaista ruokavaliota? (Kysyjällä on verenpainetauti)	Kuinka monta verenpainetautitapausta olisi voitu välttää, jos koko väestö olisi noudattanut vähäsuolaista ruokavaliota?

J. Pearl & D. Mackenzie, *Miksi, syyn ja seurauksen uusi tiede*, Terra Cognita, 2018. Alkuteos: The Book of Why

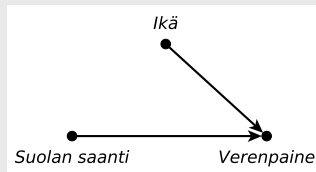
Sekoittajat (confounders)

Havainnoiva tutkimus



Ikä on sekoittava tekijä suolan saannin vaikutukselle verenpaineeseen.

Kokeellinen tutkimus



Suolan saantiin kohdistuva interventio katkaisee iän vaikutuksen suolan saantiin. Sekoittumista ei tapahdu.

Termien suomennot: <https://sanasto.tilastoseura.fi/>

Esimerkki: Finriski-laskuri

FINRISKI-laskuri ⓘ

Sukupuoli *

☒ Mies
☐ Nainen

Ikä *

70

Tupakoitko? *

☒ kyllä
☐ ei

Kolesteroli *

5.3 mmol/l

HDL-kolesteroli *

1.2 mmol/l

Systolinen verenpaine *

140 mmHg

Sairastatko diabetesta? *

☐ kyllä
☒ ei

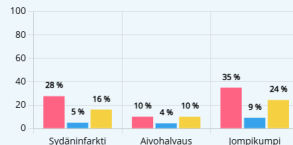
Onko jompikumpi vanhemmistasi sairastanut sydäninfarktiin? *

☒ kyllä
☐ ei

Onko jompikumpi vanhemmistasi sairastanut aivohalvauksen? *

☐ kyllä
☒ ei

Tulokset



Sinun riskisi saada sydäninfarkti/sepelvaltimotauti 28%, aivohalvaus 10% tai jompikumpi 35%

Sydäninfarktin/sepelvaltimotaudin 5%, aivohalvauksen 4% tai jommankumman 9% riski saman ikäisellä, samaa sukupuolta olevalla henkilöllä, jolla ei ole riskitekijöitä

Sydäninfarktin/sepelvaltimotaudin 16%, aivohalvauksen 10% tai jommankumman 24% riski saman ikäisellä, samaa sukupuolta olevalla henkilöllä, jolla suomalaisväestön keskimääräinen riskitekijätaso

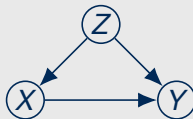
Finriski-laskuri on ennustemalli, ei kausaalimalli. Intervention (esim. tupakoinnin lopettaminen) vaikutusta riskiin ei voi luotettavasti päätellä tästä mallista.

Potentiaalisten vasteiden kehikko (potential outcomes framework)

- Tunnetaan myös nimellä Neyman-Rubinin kausaalimalli tai Rubinin malli.
- Keskeiset lähteet: Neyman (1923), Rubin (1974), Rosenbaum & Rubin (1983), Holland (1986)
- Potentiaaliset vasteet:
 - $Y^{x=1}(u)$ vaste, jos yksilö u on käsitelty,
 - $Y^{x=0}(u)$ vaste, jos sama yksilö u on käsittelemätön.
- **Kausaalipäättelyn perusongelma**: emme voi havaita molempia tuloksia.
- Havaitsematon vaste on **kontrafaktuaalinen**.
- Keskeiset oletukset:
 - Konsistenssi: Jos $X = x$ niin $Y = Y^{X=x}$.
 - Stabiili yksilövaste (stable-unit-treatment-value assumption, SUTVA)
 - Positiivisuus: $p(X = x | Z = z) > 0$ jos $p(Z = z) > 0$.

Graafisten mallien kehikko

- Tunnetaan myös nimellä rakenteellinen kausaalimalli tai Pearl'n malli.
- Keskeiset lähteet: Wright (1921), Pearl (1995), Pearl (2009)
- Rakenteellinen kausaalimalli määrittelee muuttujien väliset funktionaaliset suhteet käyttäen **rakenneyhtälöitä**.
- Kausaalimalli voidaan esittää parametrittomasti **suunnattuna silmukattomana graafia** (directed acyclic graph, DAG).



- Interventio $\text{do}(X = x)$ määritellään muodollisesti alimallin kautta, jossa X :n rakenneyhtälö asetetaan vakioksi x .
- Kausaalivaikutus saadaan interventionjälkeisestä jakaumasta $p(Y = y \mid \text{do}(X = x))$, joka usein kirjoitetaan $p(y \mid \text{do}(x))$.

Takaovikorjaus (backdoor adjustment)

Potentiaaliset vasteet

- Seuraavan takaovioletuksen (sivuutettavuusoletuksen) vallitessa

$$Y^x \perp\!\!\!\perp X \mid \mathbf{Z}, \text{ kaikilla } x,$$

missä $\mathbf{Z}$ on joukko **sekoittajia**, muuttujan X kausaalivaikutus muuttajaan Y identifioituu takaovikorjauksella

$$p(Y^x = y) = \sum_{\mathbf{z}} p(\mathbf{z})p(y \mid x, \mathbf{z}).$$

- Graafit eivät ole välttämättömiä.
- Identifiointiin ei tarvita oletuksia muista muuttujista.

Graafiset mallit

- Määrittele DAG $\mathcal{G}$.
- Etsi sekoittajien joukko $\mathbf{Z}$, joka täyttää takaovikriteerin

$$(Y \perp\!\!\!\perp X \mid \mathbf{Z})_{\mathcal{G}}$$

eli Y ja X ovat **d-separoituja** graafissa $\mathcal{G}$ ehdolla $\mathbf{Z}$.

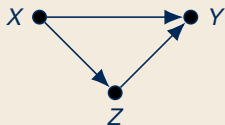
- Muuttujan X kausaalivaikutus muuttajaan Y identifioituu takaovikorjauksella

$$p(y \mid \text{do}(x)) = \sum_{\mathbf{z}} p(\mathbf{z})p(y \mid x, \mathbf{z}).$$

Identifioivan funktionaalin voi selvittää myös identifiointialgoritmeilla (R-paketit `causaleffect` ja `dosearch`).

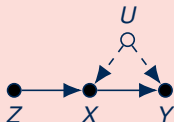
Kausaalipäätelyn menetelmiä ja kysymyksiä

Suorat ja epäsuorat vaikutukset



Suora vaikutus: $p(y \mid \text{do}(x, z))$

Instrumenttimuuttujat



Lineaarisessa mallissa

$$\beta = \frac{E(Y \mid Z = z)}{E(X \mid Z = z)} = \frac{\beta_{yz}}{\beta_{xz}},$$

Sensitiivisyysanalyysi

- Kuinka havaitsematon sekoittaja voisi vaikuttaa tuloksiin?

Laajennuksia

- Havainnoivien ja kokeellisten aineistojen yhdistäminen
- Kausaalivaikutusten siirrettävyys
- Valikoitumisharha ja puuttuva tieto
- Kontrafaktuaalipäätely
- Luonnolliset kokeet

Tavoitekokeen jäljittely (target trial emulation)

Tutkija päättää millaisen kokeen haluaisi tehdä ja jäljittelee sitä rekisteriaineistolta (tai muulla havainnoivalla aineistolla).

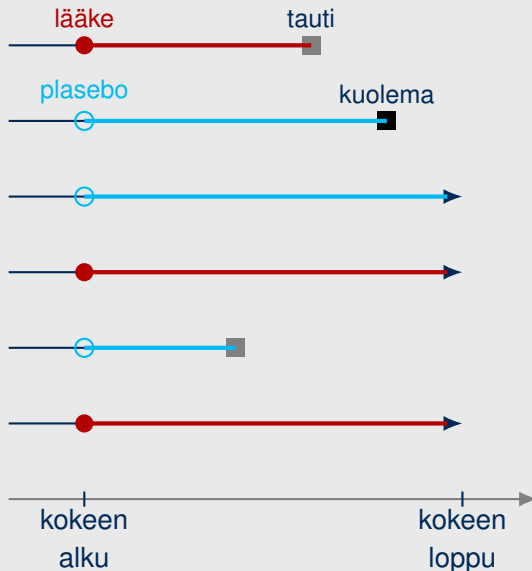
Kliininen koe

1. Mukaanotto- ja poissulkukriteerit
2. Hoito satunnaistetaan.
3. Kokeen alku ja loppu päätetty ennakkoon
4. Analysoidaan protokollan mukaan.

Jäljittely

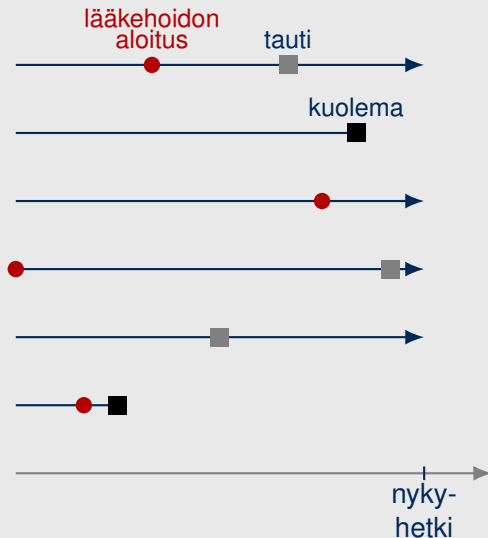
1. Samat mukaanotto- ja poissulkukriteerit
2. Sekoittajat otetaan huomioon kausaalipäätelyssä.
3. Tutkimusasetelman alku ja loppu päätetty ennakkoon
4. Analysoidaan (saman) protokollan mukaan.

Satunnaistettu kontrolloitu koe



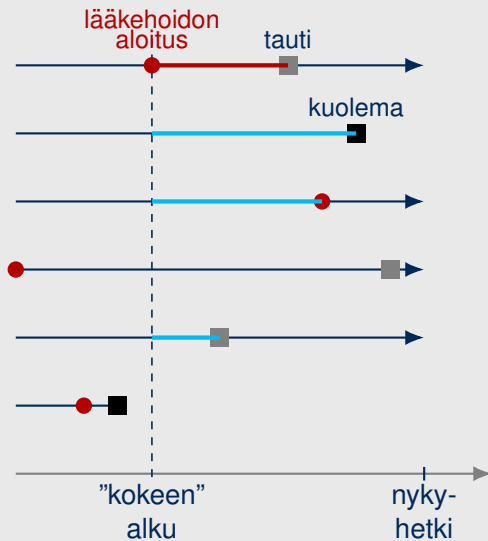
Mukaanotto- ja poissulkukriteerien tarkastelu, satunnaistus ja hoidon alku tapahtuvat kaikki samana ajanhetkenä.

Rekisteriaineisto



- Lääkehoito voi alkaa milloin tahansa.
- Lääkehoidon aloitus riippuu sekoittajista, jotka vaikuttavat myös tautiriskiin.

Tavoitekoetta jäljittelevä rekisteritutkimusasetelma

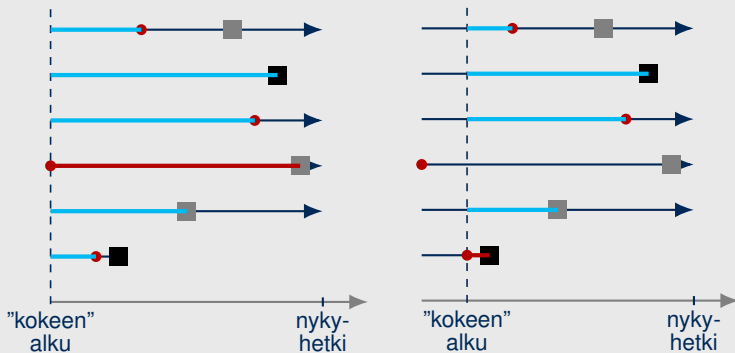


Hoitoryhmään kuuluvat ne, jotka aloittavat lääkehoidon aloitushetkellä. **Seuranta-aika** voi päättyä tautitapahtumaan, sensurointiin tai lääkehoidon lopetukseen.

Verrokkiryhmään kuuluvat ne, jotka ovat elossa eivätkä käytä lääkettä aloitushetkellä. **Seuranta-aika** voi päättyä tautitapahtumaan, sensurointiin tai lääkehoidon aloitukseen.

Estimoitaessa lääkehoidon kausaalivaikutusta tautiriskiin otetaan huomioon **sekoittajat**.

Rekisteritutkimusasetelmassa on monta aloitushetkeä



Estimoinnissa on huomioitava, että sama henkilö voi olla verrokki useita kertoja.

Keogh RH, Gran JM, Seaman SR, Davies G, Vansteelandt S. Causal inference in survival analysis using longitudinal observational data: Sequential trials and marginal structural models. *Statistics in Medicine*. 2023; 42(13): 2191-2225

Yhteenveto

- Kausaalipäättelyssä yhdistyy asiantuntijatieto ilmiöstä ja aineiston analysointi.
 - Oletukset sekoittajista ovat keskeisiä.
- Tavoitekokeen jäljittely on systemaattinen tapa kausaalivaikutuksen estimoimiseen havainnoivasta aineistosta, joka yleensä on rekisteriaineisto.
- Tavoitekokeen jäljittely vaatii suurta tarkkuutta tutkimusasetelman suunnittelussa, aineiston käsittelyssä ja tilastollisessa analyysissä.
 - Cashin et al. Transparent Reporting of Observational Studies Emulating a Target Trial—The TARGET Statement, *JAMA*, 2025.
- Ad-hoc-menetelmät voivat johtaa huomattavan suureen harhaan estimaateissa ja merkittäviin virhepäätelmiin kliinisessä käytännössä.
 - Dickerman et al. Avoidable flaws in observational analyses: an application to statins and cancer, *Nature Medicine*, 2019.