



*Elämän
tähden*

Näkökulmia "big datasta" ja tiedon louhinnasta erikoissairaanhoidon maailmassa

Erkki Kujansuu, dosentti, eMBA
PSHP

Määritelmä

Raghupathi and Raghupathi Health Information Science and Systems 2014, 2:3

<http://www.hissjournal.com/content/2/1/3>

- By definition, big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and/or hardware; nor can they be easily managed with traditional or common data management tools and methods [7].
- Big data in healthcare is overwhelming not only because of its volume but also because of the diversity of data types and the speed at which it must be managed

Terveydenhuollon data on todella BIG

Raghupathi and Raghupathi Health Information Science and Systems 2014, 2:3

<http://www.hissjournal.com/content/2/1/3>

- Reports say data from the U.S. healthcare system alone reached, in 2011, 150 exabytes. At this rate of growth, big data for U.S. healthcare will soon reach the zettabyte (10^{21} gigabytes) scale and, not long after, the yottabyte (10^{24} gigabytes) [6].
- Kaiser Permanente, the California-based health network, which has more than 9 million members, is believed to have between 26.5 and 44 petabytes of potentially rich data from EHRs, including images and annotations [

Potilaista/asiakkaista kertyy tietoa potilasjärjestelmiin

Strukturoimattomana tekstinä

- Käsitteiden ja sanojen valinnat ovat ja pysyvät epätäsmällisinä

Eri ”kielillä”

- Kirurgiaksi/psykiatriaksi, helsingiksi/savoksi

Erilaisiin järjestelmiin niiden rakenteita noudattaen

Lisää dataa syntyy

Erilaisista lääkintälaitteista

- ”Raakadataa”, josta osaa käytetään hoidollisten johtopäätösten ja kirjauksien tekoon

Potilaiden itse keräämänä

- Erilaisista mukana kulkevista digitaalisista laitteista ja muista tiedonkeruun välineistä

Laajalti kyberavaruuteen

- Näennäisen kaoottisesti syntyvää tietoa terveyteen liittyvistä ilmiöistä

Totuttuja tutkimuksen ajattelumalleja

Lääketieteellisen tutkimuksen gold standard on etukäteen suunniteltu, kontrolloitu ja satunnaistettu vertailututkimus

Edistyksellinen lääketieteellinen tutkimus perustuu genetiikkaan ja edistyksellisiin molekyyli-tason tutkimusmenetelmiin

Palvelujärjestelmän tutkiminen on vähemmän arvostettua ja usein sosiaalitieteiden malleihin perustuvaa

Potentiaalisia Big Datan sovellusalueita

Väestön terveyden kehittäminen

- Uusien yhteyksien havaitseminen ja uusien mallien luominen
- Epidemiaseurannat ja rokotukset

Hoitomuotojen kehittäminen

- Todellisen arkivaikutavuuden selvittäminen

Geenianalytiikka

Potilasprofilointi asiakassegmentoinnin tueksi

Laadun ja kustannusten analyysi

- Auttaa käyttäjä palveluiden valinnassa ja maksajaa tuottajien valinnassa

Yksilölliset hoitovalinnat

- ”Mitä muille tilanteessasi on tapahtunut”

Onko ”big datan” analyysi ikuinen lupaus vai todellinen potentiaali?

The promise and peril of big data (D Bollier, CM Firestone)

- “Big Data presents many **formidable challenges**... because data technologies are becoming so pervasive, intrusive and difficult to understand”

Center for US Health System Reform

Business Technology Office

- Big-data initiatives have the potential to **transform healthcare**, as they have revolutionized other industries. In addition to reducing costs, they could save millions of lives and improve patient outcomes.

Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework (Nitesh ym.)

- With the increase in the use of electronic medical records, ... becomes an increasingly important possibility, bringing Big Data and data science to **proactive and personalized** patient care.



Voiko ”big data” tuoda uutta ajattelumallia

Nowcasting: Disease monitoring at Internet speed

- ...researchers found that people start searching for disease-related information on Wikipedia before they seek medical attention. Using the techniques, the team monitored influenza outbreaks in the United States, Poland, Japan and Thailand, dengue fever in Brazil and Thailand, and tuberculosis in China and Thailand.
- The team was able to **forecast** all but one of these at least 28 days in advance, the lab reported

“The very point of looking to Big Data is “to identify patterns that create answers to questions you didn’t even know to ask”

Kaksi pilottiesimerkkiä

Potilaskertomusdatan tekstin louhinta

- Tavoitteena löytää IHI:n GTT-mallia hyödyntäen turvallisuuspoikkeamiin viittaavia haittatapahtumia

Paljon palveluita käyttävän asiakkaan karakterisointi

- Tavoitteena tunnistaa tietovarastoon kertyvän datan pohjalta ”oikeita suurkuluttajia”

Potilasturvallisuuden mittaaminen Global Trigger Toolilla (GTT)

Lähtökohta: Potilasturvallisuutta tulisi mitata systemaattisesti

Suomessa ei ole implementoituna valtakunnan tason mittausmenettelyä

- Todellisuudessa emme tiedä missä olemme ja miten turvallisuus kehittyy

Institute for Healthcare Improvement (IHI) on kehittänyt validoidun mittarin (GTT)

- Perustuu potilaskertomusten määrämuotoiseen lukemiseen ja määriteltyjen haittatapahtumiin viittaavien liipaisuhavaintojen (trigger) rekisteröintiin

GTT:n rajoitteita

Perustuu otantaan, antaa yleisellä tasolla kuvan sairaalan suoriutumisesta

- Ei mahdollista fokusoitua johtamista tai kehittämistä
- Ei ole ajantasainen
- Vaatii standardoidun kouluttamisen, omistautuneita henkilöitä ja työpanosta

Vaihtoehtona sähköiseen potilastietojärjestelmään pohjautuva automaattinen analyysi kaikista potilaista

Tekstin louhinta työkaluna

Liipaisujen havaitseminen tekstin automaattisella analysoinnilla

- Perustuu järjestelmän opettamiseen tunnistamaan liipaisuiksi määritetyjä sanoja ja yhdistelmiä

Pilotoitiin neurokirurgiassa ja vastasyntyneiden teholla

- Neurokirurgiassa osa GTT-triggereistä, vastasyntyneillä oma sanastomäärittely
- Validiteetti molemmissa osoittautui kohtuullisen hyväksi, sensitiivisyys 57-100 %, spesifisyys 71-95 %
- Neonatologiassa kyettiin välittömään hoitoprosessin parantamiseen

Rutiinikäyttöön tarvitaan kuitenkin jatkuvatoiminen työkalu

- Toimiva alusta kehitetty Tays neurokirurgian erikoisalalle yhteistyössä CGI:n ja SAS Instituutin kanssa

Kehitetyn järjestelmän toiminnallisuus

- **Integraatio** potilaskertomusjärjestelmään sekä muihin lähdejärjestelmiin
- **Malli** triggereiden automatisoituun hakuun ja analysointiin
- **Analysikäyttöliittymä (portaali)** haittapahtumien syvempään tutkimiseen ja arviointiin
- **Yhteenvetoraportit** jatkuvaan seurantaan

1 Data integraatio

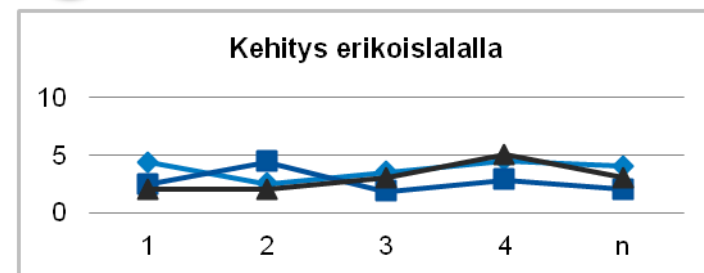
2 Mallinnus

3 Portaali / Selain



Trigger tulokset				Riskiprofiili	
C1	✓	M1	✗	X %	
C2	✗	M2	✗	Riskimuuttujat	
C3	✓	M3	✓	Var 1	✗
C4	✗	M4	✗	Var 2	✓
...

4 Mittarointi / Seurantaraportit



Paljon palveluita käyttävän asiakkaan karakterisointi

Lähtömateriaalina potilashallinnon järjestelmästä PSHP:n tietovarastoon kertyvä potilaskohtainen data

- Mm. Ikä, sukupuoli, hoitopaikka, hoitoaika, diagnoosit, toimenpiteet etc.
- 3,5 vuoden ajalta 330878 potilasta ja 2550333 hoitojaksoa

Yhteistyökumppanina Quva Oy

Menetelmänä ohjattuun oppimiseen perustuva koneoppiminen

- Vrt. Amazonin, Netflixin ym. oppivat asiakasjärjestelmät, jotka osaavat todeta ”kiinnostaisi ehkä sinuakin”

Paljon palveluita käyttävän asiakkaan karakterisointi

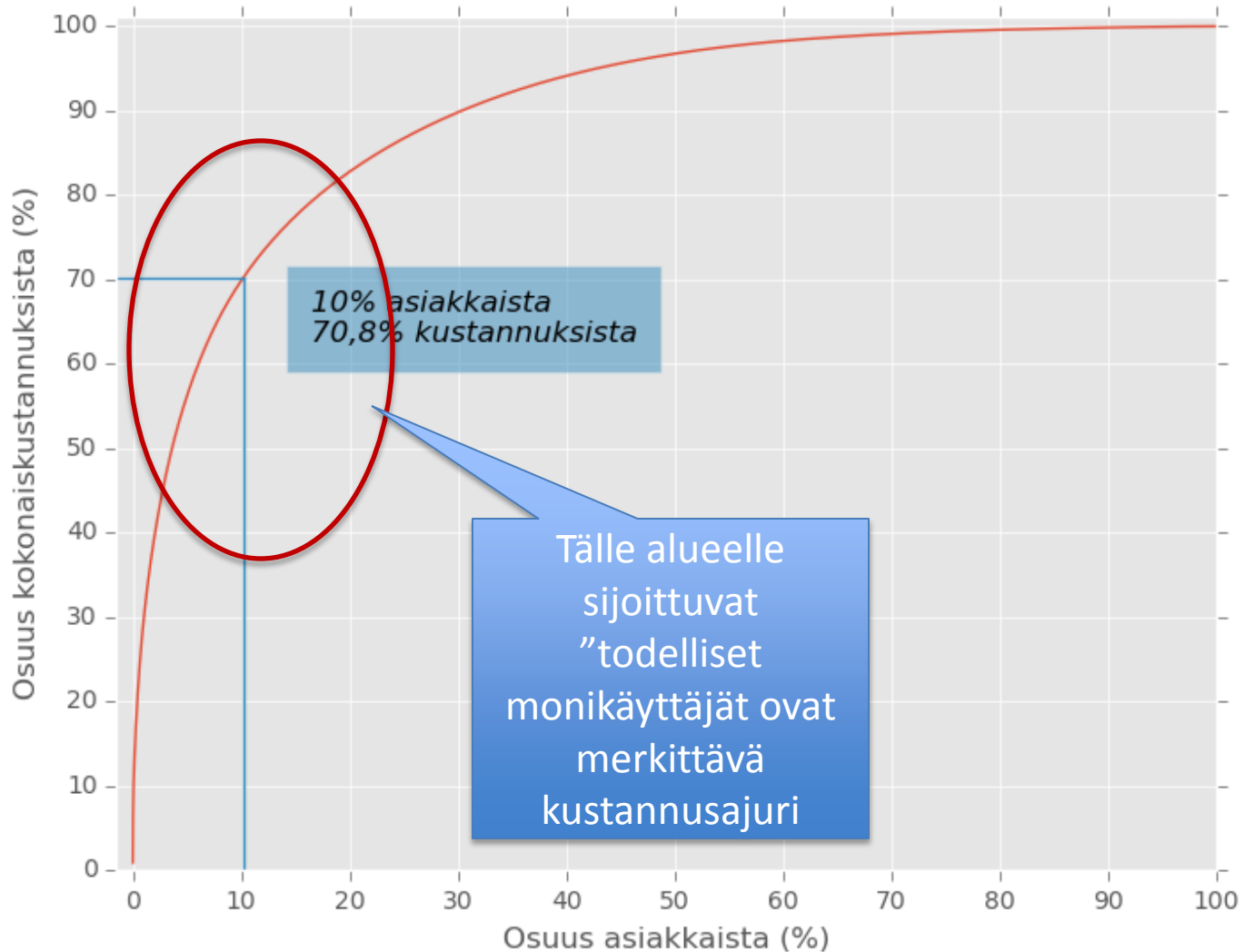
Opetusmateriaalina useilta erikoisaloilta saadut ”positiiviset” referenssit, joita asiantuntija piti ”todellisina monikäyttäjinä”

- Ei haluta tunnistaa potilaita, jotka vakavan sairauden vuoksi tarvitsevat toistuvaa/intensiivistä hoitoa

Pilottihankkeessa saavutettiin oppimistaso, jossa esiin saaduista 100 potilaasta 62 oli ”todellisia positiivisia”

Tavoitteena on kyetä opettamalla järjestelmä ennustamaan potilaan piirteiden perusteella tuleva ”todellinen positiivinen”

Sivutuotteena potilaiden hoidon kustannuskäyrä



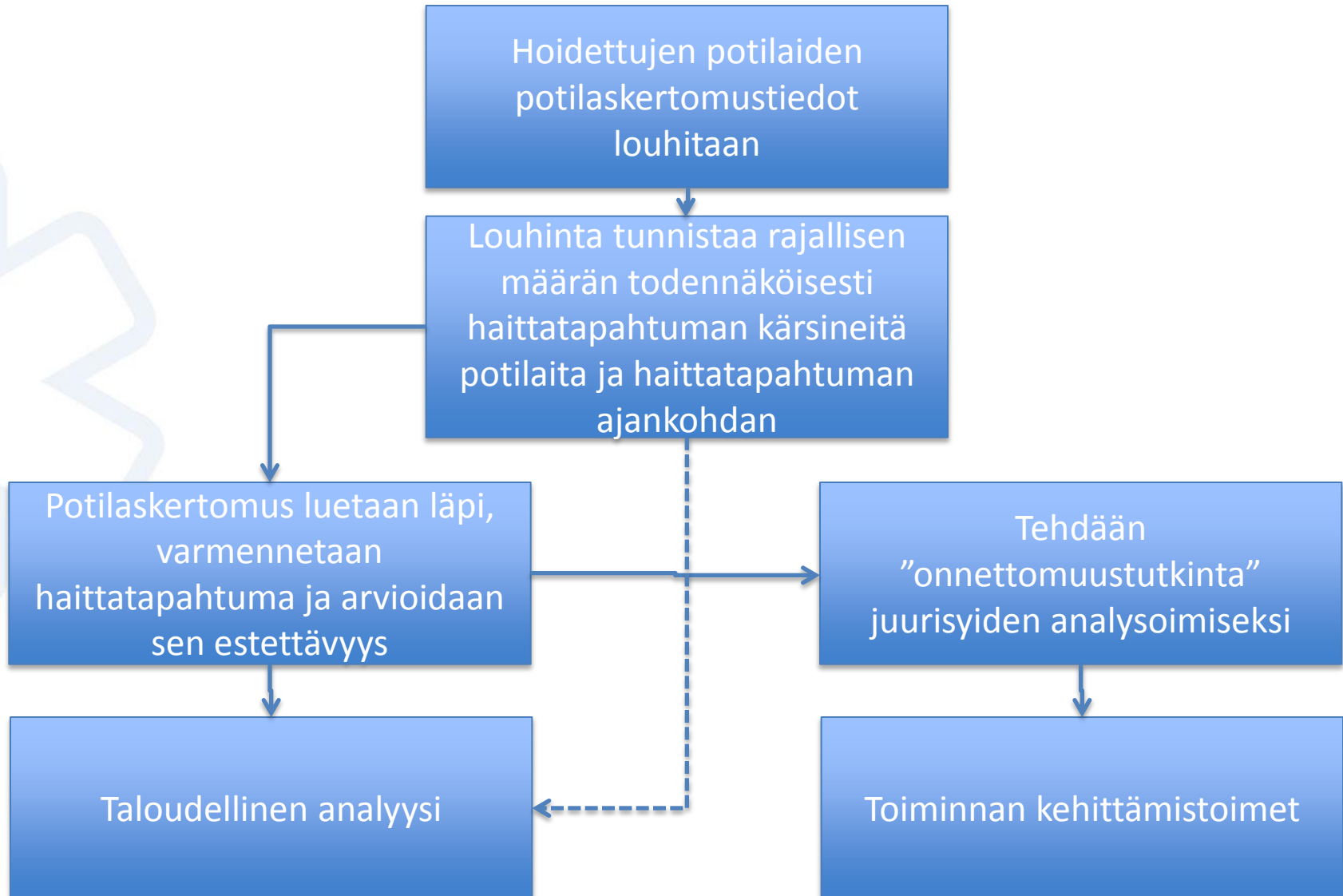
Täytyy kuitenkin tietää, mitä tuloksilla tekee

Tieto sinänsä on usein mielenkiintoista

Mutta:

- Pitää olla käsitys siitä, mihin ja miten tietoa aikoo käyttää
- Pitää olla varma että tiedon hankkimiskustannukset ovat katettavissa tuottavuuden/laadun/vaikuttavuuden parantamisella

”Tavoitearkkitehtuuri” potilasturvallisuuden kehittämiseksi



Case managerointi monisairaalla (tai sellaiseksi tulevalla)

Vaatii

- Erikoissairaanhoidon sisällä uudenlaista työtettä ja vastuun määrittelyä
- Perusterveydenhuollon ja erikoissairaanhoidon syventyvää yhteistyötä
- Rajat ylittävää prosessien määrittelyä
- Asiakastarpeen tarkempaa määrittelyä ja ratkaisukeinojen valintaa
- Asiakkaan ja ympäristön osallisuutta ratkaisujen haussa

Ellei edellä kuvatun kaltaista toimintamallia toteuteta, karakterisoinnilla on (vain) akateemista mielenkiintoa

Lopuksi

Todella suurten datamäärien analysointi on varmasti tulevaisuudessa terveydenhuollon (järjestäjän ja tuottajan) arsenaalissa

Todella suurten datamäärien analysoinnilla tulee olemaan merkittävä osa tieteellisen tutkimuksen keinovalikoimassa

Todella suurten datamäärien analyysi terveydenhuollossa on kuitenkin vasta lupaus, alustoja ja koeteltuja toimintamalleja sekä keinoja datan luokse pääsyyn puuttuu vielä