

Arvokas arkaluonteinen tutkimusaineisto

Annaleena Okuloff

Terveystieteiden tieteenala-asiantuntija

3.5.2016



TIETOARKISTO





Work Disability among Employees with Diabetes: Latent Class Analysis of Risk Factors in Three Prospective Cohort Studies

Lifestyle-related risk factors and trajectories of work disability over 5 years in employees with diabetes: findings from two prospective cohort studies

Leg-length discrepancy is associated with low back pain among those who must stand while working

Sports, Smoking, and Overweight During Adolescence as Predictors of Sciatica in Adulthood: A 28-Year Follow-up Study of a Birth Cohort

Gender difference in genetic association between *IL1A* variant and early lumbar disc degeneration: a three-year follow-up

Genetic susceptibility of intervertebral disc degeneration among young Finnish adults



Lainsäädäntö

- Henkilötietolaki
- Laki lääketieteellisestä tutkimuksesta
- Laki viranomaisen toiminnan julkisuudesta
- Laki Terveyden ja hyvinvoinnin laitoksesta
- Laki terveydenhuollon valtakunnallisista henkilörekistereistä
- Biopankkilaki



Lainsäädäntö

- Muutoksia luvassa
 - EU:n tietosuoja-asetus, voimaan 2018
 - Sosiaali- ja terveydenhuollon asiakastietojen toissijaisen käytön laki
 - Edellisten aiheuttamat muutokset muuhun lainsäädäntöön
- ISAACUS palveluoperaattori, Sitra



Henkilötietolaki

- Henkilötiedolla tarkoitetaan kaikenlaista luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa elävää koskevaksi.

Arkaluonteinen tieto:

mm. terveyttä, seksuaalista suuntautumista, uskontoa ja poliittista toimintaa koskevat tiedot

- Salassa pidettävää, **tutkimuksessa käyttö sallittu**



Arkaluonteisen datan käsittely

- Henkilötietolaki edellyttää henkilötietojen käsittelyssä huolellisuutta ja suunnitelmallisuutta
- Henkilötunnukset ja koodiavain säilytettävä erillään, vähintään salasanalla suojattuna
- Niin kauan kuin koodiavain on olemassa, on kyse edelleen henkilötiedoista
- Aineiston käyttöversiossa vain tarvittavat tiedot
- Tutkija vastuussa tietoturvasta
 - Turvallinen tiedonsiirto ja aineiston säilyttäminen



Arkaluonteisen datan käyttö tutkimuksessa

- Lääketieteellinen tutkimus:
 - tutkimusta, jossa puututaan ihmisen tai ihmisen alkion taikka sikiön koskemattomuuteen ja jonka tarkoituksena on lisätä tietoa terveydestä, sairauksien syistä, oireista, diagnostiikasta, hoidosta, ehkäisystä tai tautien olemuksesta yleensä.
 - Tutkittavien suostumuksella
- Rekisteriaineistot myös ilman suostumusta viranomaisluvalla



Arkaluonteisen datan käyttö tutkimuksessa

- Lääketieteellinen tutkimus vaatii eettisen arvioinnin
 - Sairaanhoidopiirien eettiset toimikunnat
 - Lääketutkimukseen TUKIJA, valtakunnallinen lääketieteellinen tutkimuseettinen toimikunta
- Pelkästään rekisteritietoja käyttävään tutkimukseen ei laki vaadi lausuntoa, mutta sittenkin usein viranomainen, esim. THL, vaatii



Vastaavan tutkijan pätevyys

- Jos kyseessä on kliininen lääketutkimus, tutkimuksesta vastaavan henkilön täytyy olla lääkäri tai hammaslääkäri, jolla on asianmukainen ammatillinen ja tieteellinen pätevyys
- Muussa lääketieteellisessä tutkimuksessa vastaavalla henkilöllä täytyy olla kyseisen tutkimuksen edellyttämä ammatillinen ja tieteellinen pätevyys



Tutkittavan suostumus

- Henkilötietolaki ja sen tulkinta antaa tutkittavan suostumukselle tiukat reunaehdot
 - Suostumuksen on oltava yksiselitteinen
 - Arkaluonteisen tiedon osalta lisäksi nimenomainen
 - Nykyinen tulkinta sekä tuleva tietosuoja-asetus: Suostumus on annettava selkeästi suostumusta ilmaisevalla toimella, kuten kirjallisella, mukaan lukien sähköisellä, tai suullisella lausumalla



Tutkittavan informoitu suostumus

- Kirjallinen tiedote
- Suostumuksesta käytävä ilmi, mihin henkilö suostumuksensa antaa.
 - Osallistua tutkimukseen
 - Tutkimuksessa kerättävän tiedon käyttöön tutkimuksessa
 - Terveystietojen käyttöön tutkimuksessa
 - Muiden rekisteritietojen yhdistämiseen
 - Muuhunkin tutkimukseen?



Tutkittavan suostumus

- Tietosuoja-asetus, johdanto-osa kohta 33:
 - Usein tieteellisiä tutkimustarkoituksia varten tehtävän käsittelyn tarkoitusta ei ole mahdollista täysin määrittää siinä vaiheessa, kun henkilötietoja kerätään. Tästä syystä rekisteröityjen olisi voitava antaa suostumuksensa tietyille tieteellisen tutkimuksen aloille silloin, kun noudatetaan tieteellisen tutkimuksen tunnustettuja eettisiä standardeja. Rekisteröidyillä olisi oltava mahdollisuus antaa suostumuksensa ainoastaan tietyille tutkimusaloille tai tutkimushankkeiden osille siinä määrin kuin tarkoitus sen mahdollistaa.



Rekisterit

- Rekisterit = kansalliset hallinnolliset sekä tilasto- ja tutkimustarkoitusta varten olevat rekisterit
- Rekistereitä, mm.
 - Syntyneiden lasten rekisteri (THL)
 - Hoitoilmoitusrekisteri (HILMO) (THL)
 - Syöpärekisteri (THL)
 - Yli 9 päivän sairauslomat (KELA)
 - Potilastietojärjestelmät (mm. sairaalat)



Rekisterit

- Lisää lainsäädäntöä
 - Laki Eläketurvakeskuksesta
 - Laki potilaan asemasta ja oikeuksista
 - Laki sosiaalihuollon asiakkaan asemasta ja oikeuksista
- Salassa pidettävää tietoa, saa käyttää tutkimuksessa
- Lupa käyttöön rekisterin pitäjältä
- Suostumusta ei yleensä tarvita

- Rekisteritutkimus eli rekisterit tutkimusaineistona

Arkistointi ja jatkokäyttö

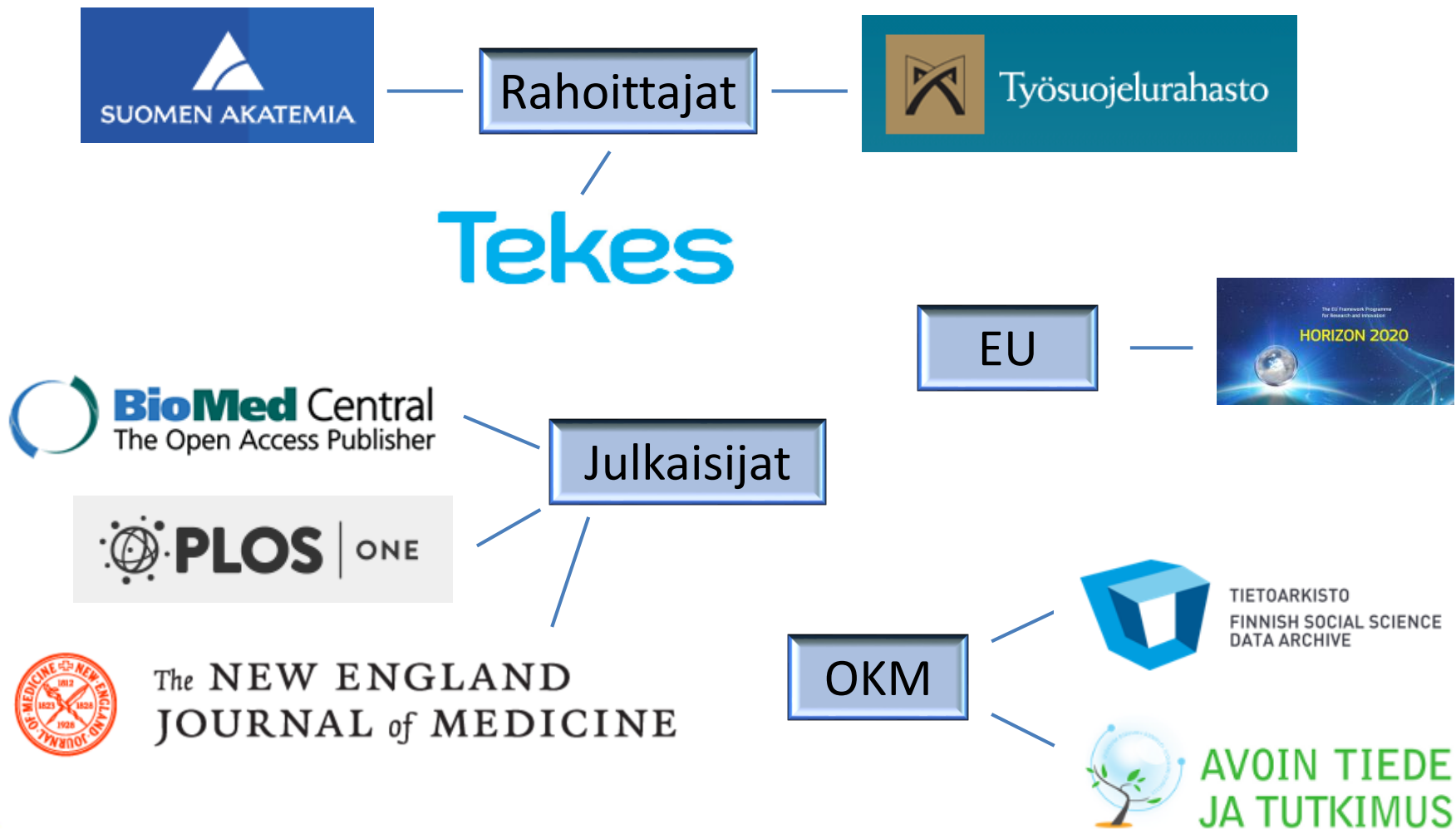
Annaleena Okuloff
Terveystieteiden tieteenala-asiantuntija
3.5.2016



TIETOARKISTO



Tiedepoliittinen tilanne





Tiedepoliittinen tilanne

- The International Committee of Medical Journal Editors esittää:
 - Jatkossa ICMJE:n jäsenlehdissä hyväksyttäisiin julkaistavaksi vain sellaisten kliinisten tutkimusten artikkeleita, joiden tulosten taustalla oleva tutkimusaineisto avataan <http://www.icmje.org/news-and-editorials/M15-2928-PAP.pdf>
 - Eettinen näkökulma



Taustalla

- Korkeatasoiset rekisterit ja väestötutkimukset
- Rahoittaja haluaa rahoilleen enemmän vastinetta
- Vähenevä tutkimusrahoitus
- Pois päällekkäisestä tiedonkeruusta
- Tutkijoiden aika ja resurssit säästyvät
- Ilmiöiden monipuolisempi tutkiminen; uudenlainen osaamisen hyödyntäminen
- Genetiikassa konsortiot ovat jopa välttämättömiä



Arkaluonteisen datan arkistointi ja jatkokäyttö

- Arkaluonteista tietoa sisältävien aineistojen jatkokäyttö:
 1. Tutkittavien suostumuksella
 2. Anonymisoituna myös arkistointi
 - » Eettisistä syistä myös anonymisoituna arkistointiin pyydettyä suostumusta



Mitä on arkistointi?

= Aineisto avataan muiden tutkijoiden saataville
jatkokäyttöä varten

- Sähköiset tutkimusaineistot
- Aineisto tallennetaan data-arkistoon, jossa se kuvaillaan ja josta muut tutkijat voivat saada sen käyttöönsä
- Aineiston saatavuutta voidaan tarvittaessa rajoittaa



Aineiston kuvailu eli metadata

- Mahdollistaa aineistonhallinnan
- Säilyttää aineiston käyttökuntoisena
- Varmistaa osaltaan tutkimustulosten luotettavuuden ja tarkistettavuuden
- Edellytys aineistojen luetteloinnille ja löytämiselle sekä viittaamiselle

Ilman kuvailevaa tietoa aineisto on merkityksetön kokoelma tiedostoja, numeroita, merkkejä tai bittejä!



Anonymisoitu aineisto

- Tietosuoja-asetus:
 - Tietosuojaperiaatteita ei tämän vuoksi pitäisi soveltaa anonymeihin tietoihin ... tai henkilötietoihin, joiden tunnistettavuus on poistettu siten, ettei rekisteröidyn tunnistaminen ole tai ei ole enää mahdollista.
 - HetiL:n tulkinta sama
- Anonymisoitu aineisto voidaan arkistoida



Anonymisoinnin lähtökohdat

- Tutkittavaa ei saa tunnistaa edes välillisesti
- Niin kauan kuin aineistoon on mahdollista palauttaa tunnistetiedot, aineisto ei ole anonymi
- Edes alkuperäinen tutkimusryhmä ei saa kyetä selvittämään tutkittavien henkilöllisyyttä
- Henkilötunnusteiden korvaaminen koodilla ei ole anonymisointia
 - Koodilla suojattu aineisto on pseudonyymi



Anonymisointi

- Anonymisointiin ei ole olemassa valmista kaikkiin aineistoihin soveltuvaa menettelytapaa
- Tunnusteiden poisto tulee suunnitella aina aineistokohtaisesti
- Suorien ja vahvojen epäsuorien tunnusteiden lisäksi tulee tarkastella epäsuorien tunnusteiden poiston ja muokkaamisen tarve
- Lisäksi tulee varmistaa, ettei yksittäisiä henkilöitä voi tunnistaa muualta saatavien tietojen perusteella



Anonymisointi

- Suoria tunnisteita
 - Koko nimi, henkilötunnus, sähköpostiosoite, valokuva
- Epäsuoria tunnisteita
 - Puhelinnumero, syntymäaika, postinumero, ikä, sukupuoli, siviilisääty...
 - Harvinainen sairaus



Anonymisointi

- Anonymisointikeinoja:

Kaikenlaisissa aineistoissa:

- Muuttujan poistaminen – työpaikan nimi
- Muuttujan arvojen uudelleen luokittelu –
Sysmä -> maaseutumainen kunta

Kvantitatiivisissa aineistoissa:

- Muuttujan arvojen harkinnanvarainen luokittelu –
ikäjakauma 18-109 -> 18-100 vuotiaat ikä sellaisenaan,
lopun luokassa yli 100



Anonymisointi

- Kvantitatiivisen aineiston anonymisointikeinoja:
 - Tunnisteiden poistaminen avointen kysymysten vastauksista
 - Otoksen käyttäminen koko aineiston sijasta – käytetään usein tilastokeskuksessa. Muussa kuin rekisteritutkimuksessa useimmiten kyseessä otos jo valmiiksi.
 - Satunnaisvaihtelun lisääminen – pituus ilmoitetaan senttimetrin tarkkuudella millimetrien sijaan
 - Lisätietoja Tietoarkiston käsikirjassa:
<http://www.fsd.uta.fi/aineistohallinta/fi/tunnisteellisuus-ja-anonymisointi.html#anonymisoinnin-lahtokohtat>



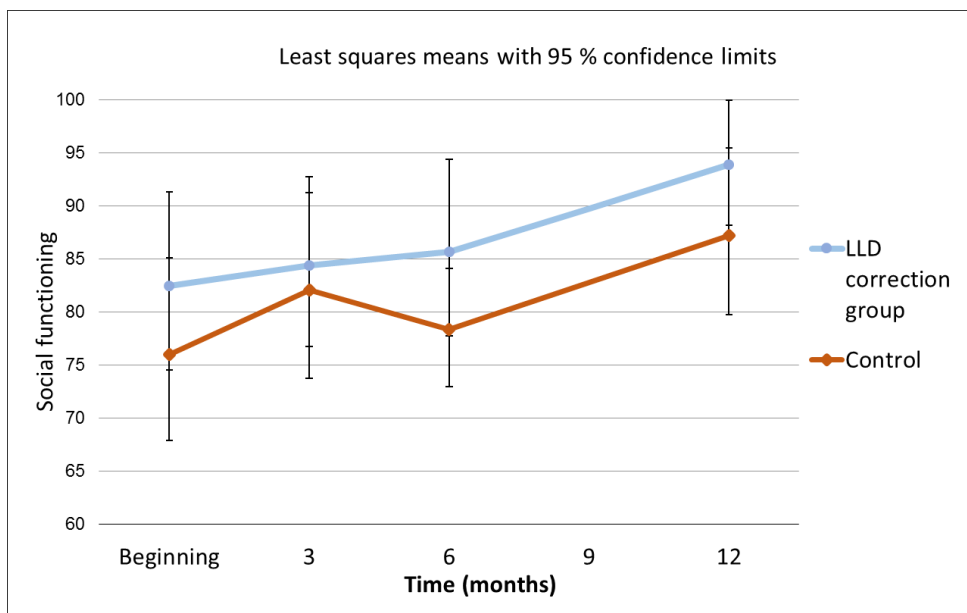
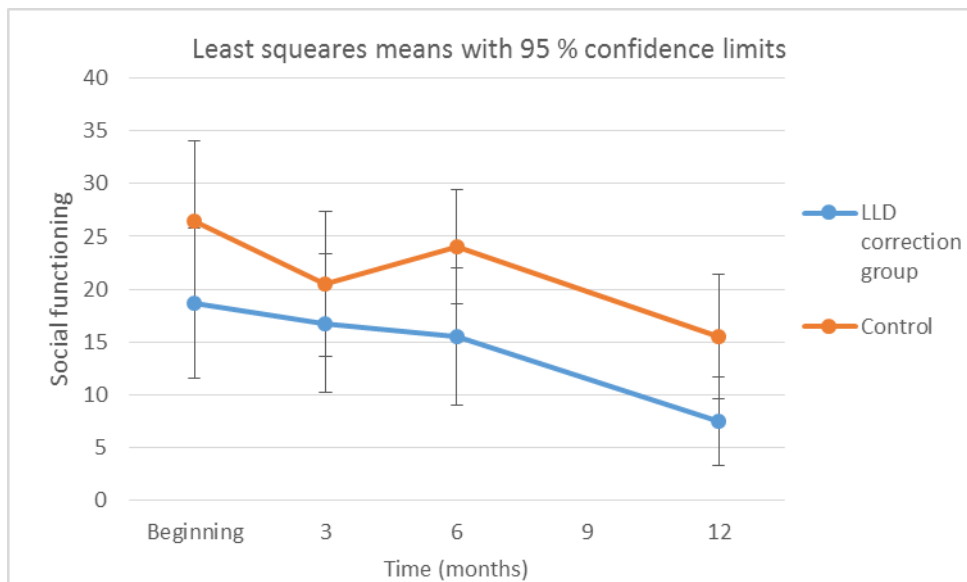
Arkistoinnin käytännön vaatimukset

- Jälkikäteen tehtynä vaatii jonkin verran aikaa ja vaivaa ja voi olla huomattavan hankalaa jos dokumentointia ei ole
- Käytännössä kuitenkin kuvailun tietoja tarvitaan
 - aineistosta tehtyihin julkaisuihin
 - perehdyttäessä uutta tutkijaa aineistoon
 - omaan jatkotutkimukseen aineistolla

Arkistointia varten tehty kuvailu auttaa
jo projektin aikana



Ilman hyvää dokumentointia voi käydä vaikka näin:





Arkistoinnin käytännön vaatimukset

Tarvittavat tiedot aineiston kuvailuun:

- Projektin kuvaus ja aineiston alkuperäinen käyttötarkoitus
- Tutkimusasetelma
- Tutkimuspopulaation valinta ja aineiston keräys
 - Poissulkukriteerit
 - Flowchart



Arkistoinnin käytännön vaatimukset

Tarvittavat tiedot aineiston kuvailuun:

- Mitatut muuttujat, niihin tehdyt muutokset; käytetyt mittarit
- Vastausten koodaus:
Sukup-muuttuja: 1=nainen, 2=mies
- Kyselylomake
- Rekisterinumero interventioiden ja kliinisten kokeiden osalta



Study Populations

FPSS concerns workers of 10 towns and 21 hospitals [10, 15]. For the present study, the base cohort was participants who responded to the survey in 2004 ($n = 48,076$, response rate 66%). Of the respondents, 1,359 were identified as having diabetes. After exclusion of people who died or retired during one year after baseline, the analytic sample was 1,324 (368 men, 956 women).

The GAZEL cohort study, established in 1989, is comprised of employees from the French national gas and electricity company: Electricité de France-Gaz de France (EDF-GDF) [16]. At baseline, 20,625 employees (73% men), aged 35–50 years, participated. Of the 20,625 participants who responded to at least one survey between the years 1989 and 2003, 914 were identified as having diabetes. After exclusion of people who died or retired during one year after baseline, the analytic sample was 842 (678 men, 164 women).

The Whitehall II study is a prospective cohort study of British civil servants (government employees) [17]. The target population was all London-based office staff, aged 35–55 years, working in 20 civil service departments on recruitment to the study in 1985–88. At baseline 73% (10,308 employees) participated. Since then, 8 follow-up examinations have taken place approximately every 2 to 3 years. Diabetes cases were identified from 1985 to the end of 1997 (study phases 1–5), for which data on sickness absence for follow-up were available. We detected 279 cases of diabetes who had at least one year of follow-up time (196 men, 83 women).

Virtanen et al. Work Disability among Employees with Diabetes: Latent Class Analysis of Risk Factors in Three Prospective Cohort Studies.

[PLoS One](#). 2015 Nov 16;10(11)



Ascertainment of Diabetes

In FPSS, identification of diabetes cases was based on three sources: (1) national registers of purchased diabetes medicines (oral medication or insulin); (2) entitlements to special reimbursements for the costs of medication by the Social Insurance Institution of Finland; and 3) from responses to a survey question on doctor-diagnosed diabetes. In GAZEL, diabetes was identified from responses to a checklist of over 50 doctor-diagnosed chronic conditions in annual surveys. In Whitehall II diabetes was assessed by self-reported doctor-diagnosed diabetes, or use of diabetes medication (all phases) and this was supplemented by the results of fasting glucose and a two-hour oral glucose tolerance tests in study phases 3 and 5. Diabetes was diagnosed by fasting glucose of ≥ 7.0 mmol/L or oral glucose tolerance test (2-hour post-load glucose ≥ 11.1 mmol/L) [18].

Risk Factors

In all cohorts, comorbid conditions included cardiovascular disease, hypertension, asthma, rheumatoid arthritis and depression or psychological symptoms. In FPSS, data on comorbid physical diseases were based on entitlements to special reimbursement, and data on psychological symptoms was based on the General Health Questionnaire (GHQ) 12-item version [19]. In GAZEL, data for all comorbid diseases were derived from survey responses to a checklist of doctor-diagnosed chronic conditions. In Whitehall II, information on coronary heart disease was based on clinical examination, medical records, and hospital records. Information on comorbid physical diseases was based on self-reported longstanding illnesses, and information on psychological symptoms was based on the GHQ 30-item version [19, 20]. In FPSS and GAZEL, self-reported height and weight and in Whitehall II, height and weight measured at the study clinic were used to calculate body mass index (BMI) from which overweight

Virtanen et al. Work Disability among Employees with Diabetes: Latent Class Analysis of Risk Factors in Three Prospective Cohort Studies.

[PLoS One](#). 2015 Nov 16;10(11)



Covariates

Socio-demographic baseline covariates were age, sex, occupational grade (based on occupational grade and categorised as high, intermediate and low), and marital status (married or cohabiting versus single, divorced or widowed). Data on occupations in each cohort were based on company records (FPSS and GAZEL) or self-reports (Whitehall II). In FPSS, high occupational grade included e.g., teachers and physicians, intermediate grade e.g., registered nurses and technicians, and low grade included e.g., cleaners and maintenance workers [15]. The GAZEL study comprises employees in executive and middle management positions (high grade), other non-manual workers (intermediate grade) and manual workers (low grade) [16]. In Whitehall II, the corresponding occupational groups were administrative (high grade), professional or executive (intermediate grade), and clerical or support (low grade) [23]. These indicators of socioeconomic status have previously been used in the Individual-Participant-Data Meta-analysis in Working Populations project [24]. For analyses, occupational grade was dichotomised as intermediate/low versus high.

Virtanen et al. Work Disability among Employees with Diabetes: Latent Class Analysis of Risk Factors in Three Prospective Cohort Studies.

[PLoS One](#). 2015 Nov 16;10(11)

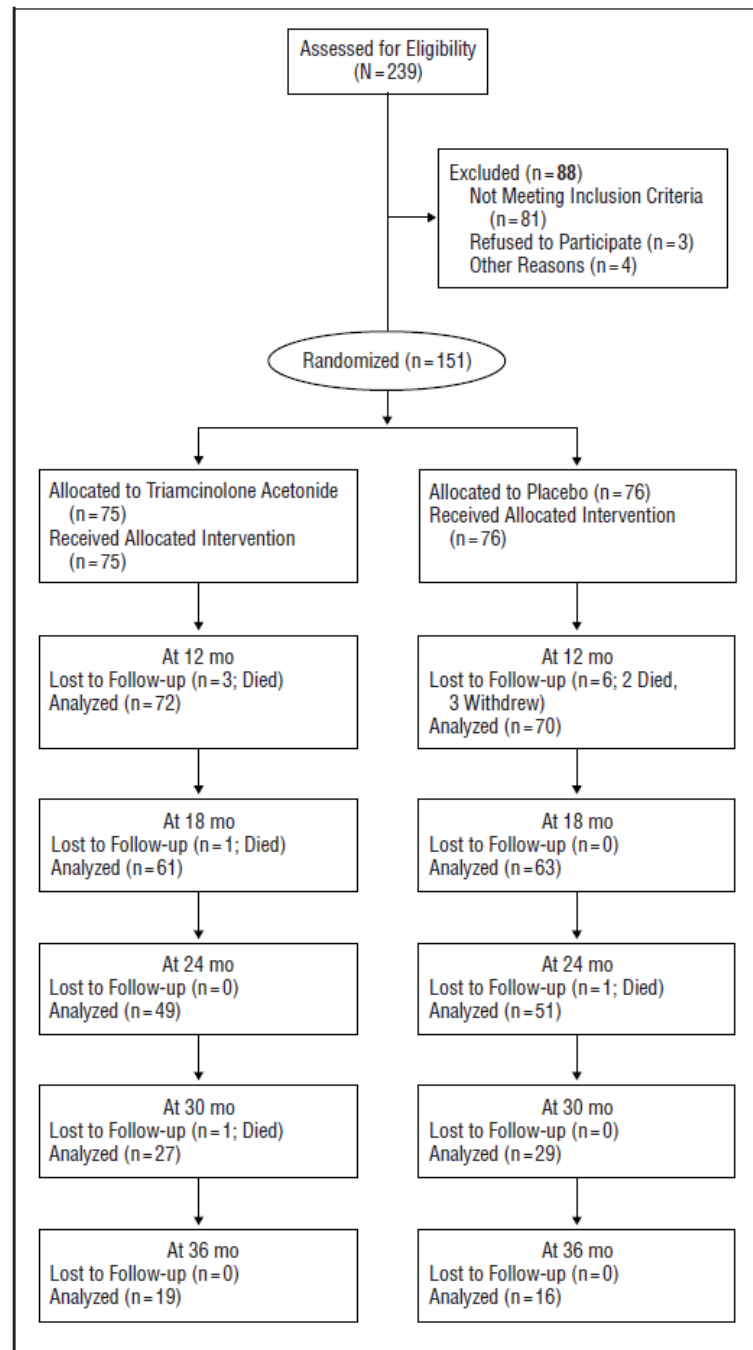


Figure 1. Study flowchart.

**Gillies et al. Safety of an
Intravitreal Injection of
Triamcinolone**

Arch Ophthalmol.
2004;122(3):336-340.



Julkaisudatan arkistointi

- Kun tutkimus on vielä käynnissä
 - Anonymisoidaan jos mahdollista
- Entä jos anonymisointi on liikaa?
- Kansallinen linjaus puuttuu
 - Julkaiseminen vs. tietosuoja vs. tietosuojatulkinta



Rekisteritietojen arkistointi

- Aineistoon yhdistetyt rekisteritiedot?
 - Rekisteritietoja hakiessa ilmoitus/lupahakemus arkistointiaikeista
 - Rekisteriviranomaisilla ei toistaiseksi virallista kantaa miten näiden suhteen tulee menetellä
 - Jo mainittu SOTE-tiedon toissijaisen käytön laki tullee antamaan rajat



Mikä arkisto aineistolle?

- Tietoarkisto
 - Digitaaliset aineistot
 - Ei tarkkoja tieteenalarajoja
- Biopankit
 - Biologiset näytteet + niihin liittyvä muu tutkimusaineisto
- Alan omat, vakiintuneet arkistot
 - Ongelmana anonymisointi
- Kielipankki
 - Haastattelunauhat, videotallenteet



Tietoarkisto

- Tutkijan kanssa tehdään sopimus
 - Julkaisuaikankohta
 - Käyttöoikeudet
- Tietoarkistossa tutkimusaineisto tallennetaan muotoon, jossa se säilyy käyttökelpoisena
- Aineisto kuvaillaan ja kuvailu julkaistaan Ailassa
- Auttaa anonymisoinnissa ja tarpeen vaatiessa anonymisoi aineiston
- Tietopalvelu neuvoo ja auttaa



Tietoarkisto

Aila

Aineistot Haku Ohjeet

- Ailassa tutkijat voivat tarkastella aineistoja ja ladata aineiston käyttöönsä suoraan tai hakea lupaa aineiston käyttöön
- Aineistohallinnan käsikirja, jossa löytyy tietoa kaikkiin tutkimusaineiston elinkaaren vaiheisiin:

<http://www.fsd.uta.fi/aineistohallinta/fi/>

[Yhteenveto](#)[Koko kuvailu](#)[Muuttujat](#)[Julkaisuja aineistosta](#)[Lataa aineisto](#)

Aineiston nimi

Lapsesta aikuiseksi: 50-vuotiaiden terveystutkimus 2009

Aineistonumero

FSD2619

Aineiston laatu

Kvantitatiivinen aineisto

Tekijät

- Pulkkinen, Lea (Jyväskylän yliopisto. Psykologian laitos)
- Kinnunen, Marja-Liisa (Jyväskylän yliopisto. Psykologian laitos)
- Kokko, Katja (Jyväskylän yliopisto. Psykologian laitos)

Muut tekijät

- Terveystarkastuksen toteuttamisesta vastasivat pääosin kaksi terveydenhoitajakoulutuksen saanutta sairaanhoitajaa.

Sisällön kuvaus

Aineisto on osa Lapsesta aikuiseksi -tutkimusta, joka on jatkunut samojen henkilöiden seurantaan vuodesta 1968. Tässä 50-vuotiaille vastaajille vuonna 2009 toteutetussa tutkimusvaiheessa kerättiin tietoa elämäntilannekyselyllä, haastattelulla, persoonallisuustestillä, kyselylomakkein sekä terveystarkastuksella. Tämä aineisto sisältää terveystarkastusaineiston.

110 miestä ja 114 naista osallistui terveystarkastukseen. Aluksi vastaajat arvioivat omaa terveydentilaansa ja täyttivät oirelomakkeen. Sen jälkeen heiltä mitattiin verenpaine, syke, pituus, paino, vyötärön ympäryys ja lantion ympäryys. Lisäksi tehtiin kädenpuristusvoimatesti. Vastaajat tekivät muistiin ja sen toimintaan liittyviä testejä, joilla arvioitiin kielellistä sujuvuutta, sanalistan oppimista ja viivästettyä mieleenpalautusta. Uutena osiona muistin mittaamisessa oli testi, jossa vastaajat luettelivat numerosarjoja käänteisessä järjestyksessä.

Aineisto on käytettävissä (D) **vain luovuttajan luvalla.**

[Lataa aineisto täältä](#)

Muunkieliset kuvailuversiot

- [englanninkielinen](#)

Aineistoon liittyvät tiedostot

- [Kyselylomake \(PDF-tiedosto, suomenkielinen\)](#)

Ajallinen kattavuus

2002

Aineistonkeruun ajankohta

2002-09-09 – 2002-12-10

Maa

Suomi

Kohdealue

Suomi

Havaintoyksikkötyyppi

Henkilö

Perusjoukko/otos

15 vuotta täyttänyt väestö

Tutkimuksen aikaulottuvuus

Pitkittäisaineisto: trendi/toistuva poikkileikkausaineisto

Otantamenetelmä

Yksiasteinen systemaattinen alkiotason todennäköisyysotanta (ryvästystä ei käytetty). Otokseen poimittiin Tilastokeskuksen Suomen väestöä koskevasta tietokannasta 2766 Suomessa asuvaa 15 vuotta täyttäneitä henkilöä. Otos ei kattanut laitoksissa eikä ulkomailla asuvaa väestöä. Perusjoukko lajiteltiin asuinpaikan, sukupuolen ja iän suhteen.

Keruumenetelmä

Käyntikysely (tietokoneavusteinen kysely CAPI) ja lomakekysely. Kohdehenkilöille lähetettiin etukäteen kirje ja tutkimuksesta kertova esite ja heihin otettiin yhteyttä puhelimitse. Suppeampi aineisto kerättiin nk. perävaunulomakkeilla.

Keruväline tai -ohje

Strukturoitu kyselylomake

Vastausprosentti

Haastatteluja saatiin 2000 joten vastausosuudeksi tuli noin 76 %. Suurin katoryhmä olivat kieltäytyneet. Perävaunukyselyyn

1 Miksi aineistonhallintaa ja jatkokäyttöä?

- Osa hyvää tieteellistä käytäntöä
- Rahoittajat edellyttävät aineistonhallintaa
- Jatkokäytön ja avoimen saatavuuden hyödyt
- Tutkimusaineiston elinkaari

2 Aineistonhallinnan suunnittelu

- Aineistonhallintasuunnitelma
- 1. Aineisto
- 2. Oikeudet
- 3. Tietosuoja ja tietoturva
- 4. Tiedostoformaatit ja ohjelmistot
- 5. Aineiston käsittelyn ja sisällön kuvaus
- 6. Elinkaari

3 Sopimukset ja oikeudet

- Tekijänoikeus
- Tutkijoiden välisistä sopimuksista
- Oikeuksien siirtymisestä sopiminen
- Tekijäyys tieteellisissä julkaisuissa ja tutkimusaineistoissa

4 Tutkittavien informointi

- OSA I: Informoinnin merkitys
- OSA II: Informoinnin sisältö, muoto ja tutkittavan suostumus
- OSA III: Informoinnin osa-alueet
- Osa IV: Esimerkkejä informoinneista

5 Tiedostoformaatit ja ohjelmistot

- Tiedostojen konvertointi
- Tutkimusaineiston digitointi
- Havaintomatriisit ja tilasto-ohjelmistot
- Tekstiaineistot
- Kuva-aineistot
- Ääni- ja äänikuvatallenteet

6 Kvantitatiivisen datan käsittely

- Havaintomatriisin tallennus
- Muuttujien nimet ja selitteet
- Muuttujan arvot ja niiden selitteet
- Muuttujien muunnokset
- Puuttuva tieto
- Painomuuttujat
- Syntaksin käyttö

7 Kvalitatiivisen datan käsittely

- Litterointi
- Aineiston sisältämien datatiedostojen organisointi
- Datatiedostojen nimeäminen
- Taustatietojen kirjaamistavat
- Lehtiaineistot

8 Tunnisteellisuus ja anonymisointi

- Mitä on henkilötieto?
- Milloin tieto on tunnistettava?
- Milloin tieto on anonymiä?
- Tunnisteellisten aineistojen käsittely
- Anonymisoinnin lähtökohdat
- Kvantitatiivisen aineiston anonymisointi
- Kvalitatiivisen aineiston anonymisointi

9 Aineiston kuvailu ja metadata

- Metadataan tallennus
- Tutkimuksen toteutuksen kuvailu
- Aineistonkeruuinstrumentti
- Tiedostojen kuvailu
- Muuttujien kuvailu
- Saatavuustiedot
- Kontekstitiedot ja paradata
- Metadataformaattit

10 Fyysinen säilytys

- Tiedostojen nimeäminen ja hallinta
- Varmuskopiointi
- Migraatio ja virkistäminen
- Tallennusmediat
- Tietoturva
- Aineiston hävittäminen
- Paperiaineiston säilytys

11 Malliesimerkit ja sanastot

Sisällys





Tietoarkistossa tapahtuu

- Tietoarkisto järjestää historian toisen Aila ja aineistojen jatkokäyttö -seminaarin Tampereella 15. syyskuuta klo 12-15.
 - Tietoarkistolaisten lisäksi aineistojen arkistoinnista ja jatkokäytöstä tulevat kertomaan tutkija **Outi Jolanki** Tampereen yliopiston terveystieteiden yksiköstä ja yliopistonlehtori **Elina Kestilä-Kekkonen** Tampereen yliopiston johtamiskorkeakoulusta.

Kiitos!

Kysymyksiä, kommentteja?



TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE