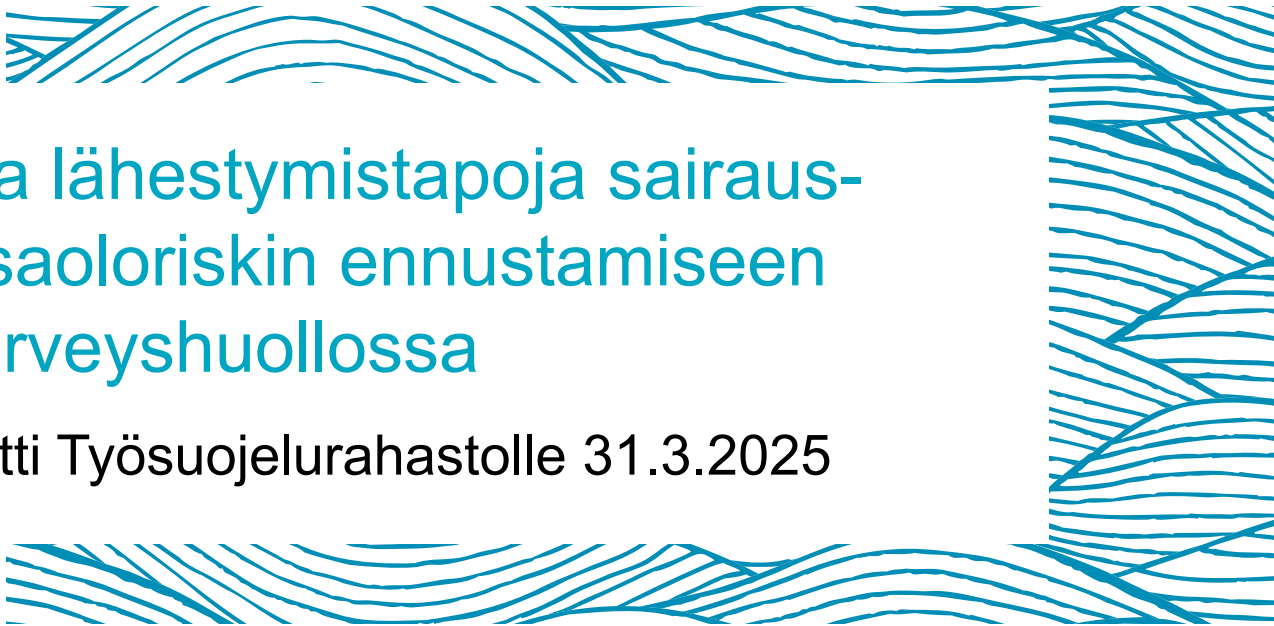




Uusia lähestymistapoja sairaus- poissaoloriskin ennustamiseen työterveyshuollossa

Raportti Työsuojelurahastolle 31.3.2025

Anniina Anttila, Mikko Nuutinen, Riikka-Leena Leskelä, Anu Pekki, Mark Van Gils,
Riitta Sauni



Työsuojelurahasto
Arbetskyddsfonden
The Finnish Work Environment Fund



Nordic
Healthcare
Group



Tampereen
yliopisto

finla
TYÖTERVEYS

ISBN 978-952-03-3873-2 (verkkojulkaisu)
Tampere 2025



Sisällysluettelo



Tiivistelmä.....	4
Johdanto.....	6
Työkyvyttömyys	7
Tavoitteet	8
Menetelmät.....	10
Aineisto	10
Muuttajat.....	12
Tulokset	13
Masennusoireet ja työn psykososiaaliset kuormitustekijät sairauspoissaoloriskin ennustajina	13
Sairauspoissaolojen ennustemallit	14
Työkykyriskiprofiilien tunnistaminen	14
Johtopäätökset ja tulosten hyödyntäminen	16
Lähteet.....	17

Tiivistelmä

Tässä tutkimushankkeessa pyrimme löytämään sairauspoissaoloriskiä ennustavia tekijöitä modernein tekoälypohjaisin analyysimenetelmin. Olimme kiinnostuneita selvittämään sitä, millä tavoin työterveyshuoltoon jatkuvasti kertyvää tietoa pystyttäisiin paremmin hyödyntämään, jotta työterveyshuollon resurssit ja toimenpiteet saataisiin kohdennettua oikea-aikaisesti niistä parhaiten hyötyviin henkilöihin. Pyrimme syventämään ymmärrystä työkykyyn vaikuttavista tekijöistä etsimällä suuresta aineistosta sellaisia yhteyksiä, joita perinteisillä analyysimenetelmillä ei ole ollut mahdollista havaita. Lisäksi pyrimme löytämään laajasta terveystieteessä käytettyjen kysymysten joukosta sellaisia keskeisiä kysymyksiä, joilla olisi parhaat edellytykset tunnistaa lisääntyneessä sairauspoissaoloriskissä olevat henkilöt. Tämä mahdollistaisi suppeamman ja siten tiheämmin toistettavissa olevan kyselyn muodostamisen.

Käytössämme oli laaja aineisto, joka sisälsi yli 18 000 terveystieteiden vastaukset suomalaiselta työterveyshuollon palveluntuottajalta. Kyselydata kattoi kymmeniä muuttujia esimerkiksi vastaajan työkyvystä, elämäntavoista kuten alkoholin ja tupakan käyttö ja liikunta- ja ravintotottumukset sekä mielenterveydestä, työolosuhteista ja työn luonteesta. Terveystieteiden vastaukset yhdistettiin sairauskertomustietoihin, jolloin saatiin yhteensä yli sadan muuttujan aineisto.

Tutkimushankkeessa havaittiin, että masennusoireiden kartoittamiseen käytettävässä DEPS-seulassa on yksittäisiä kysymyksiä, jotka ennustavat sekä toistuvia lyhyitä että pitkiä sairauspoissaoloja. Keskeisimpiä pitkiä sairauspoissaoloja ennustavista kysymyksistä olivat ”Tuntuuko tulevaisuus toivottomalta”, ”Tunnetko itsesi surumieliseksi”, ”Tuntuuko sinusta, että kaikki vaatii ponnistusta” ja ”Tunnetko itsesi arvottomaksi”. Toistuvia lyhyitä sairauspoissaoloja ennusti parhaiten työn psykososiaalisia kuormitustekijöitä kartoittaviin kysymyksiin lukeutunut ”voin vaikuttaa työni sisältöön, työtahtiini ja työaikaani”.

Koneoppimista hyödyntäen luotiin ennustemalleja, joilla pyrittiin tunnistamaan riskiä vähintään yhteen yli 30 päivän sairauspoissaolajaksoon kahden vuoden seuranta-aikana. Varsin hyvä ennustavuus saatiin mallilla, jossa hyödynnettiin tietoja työterveyspalveluiden käytöstä ja aiemmista sairauspoissaoloista, joiden jo ennalta tiedettiin olevan keskeisiä tekijöitä sairauspoissaolojen ennustamisessa. Kun kaikki muuttujat luokiteltiin teemoittain, löydettiin kustakin teemasta tiettyjä muuttujia, jotka toimivat muita paremmin sairauspoissaoloriskin ennustajina ja joihin työterveyshuollon toimenpiteitä olisi mahdollista kohdistaa. Näiden muuttujien lisääminen ennustemalliin paransi mallin kykyä tunnistaa riskiä pitkiin sairauspoissaoloihin.

Klusterointimenetelmillä tunnistettiin kuusi työntekijäprofiilia, joilla oli erilaiset yhdistelmät

riskitekijöitä ja niihin nähden johdonmukaiset eroavaisuudet pitkien ja lyhyiden sairauspoissaolojen ilmaantuvuudessa. Näiden profiilien pohjalta on mahdollista jatkossa kehittää malleja työterveyshuollon toimenpiteiden kohdentamiseen sairauspoissaolojen ennaltaehkäisemiseksi kunkin ryhmän erityispiirteet ja tarpeet huomioiden.

Työsuojelurahasto on osallistunut hankkeen rahoittamiseen.

Johdanto

Sairauspoissaolot ovat yksi työkyvyttömyyden ilmentymä, jolla on suuri merkitys sekä yksilön, työpaikan että yhteiskunnan tasolla. Vaikka julkisuuteen harmillisen paljon nousee työelämästä irtautumisen ja ”downshiftauksen” ihannointia, työterveyshuollon arkityössä kohtaa työn monet merkitykset yksilölle. Työ tuo yksilölle taloudellista turvaa, mutta myös tärkeitä sosiaalisia suhteita ja merkityksen tunnetta arvokkaana osana yhteiskuntaa.

Työpaikan tasolla sairauspoissaolot aiheuttavat suoria kustannuksia paitsi sairausajan palkan ja sijaisjärjestelyiden muodossa, myös välillisesti esimerkiksi työn sujuvuuden ja asiakaskokemuksen haasteiden kautta. Sairauspoissaoloista aiheutuvan menetetyn työpanoksen kustannusten suomalaisilla työpaikoilla on vuoden 2014 selvityksessä arvioitu olevan noin 3,4 miljardia euroa vuodessa¹.

Yhteiskunnan tasolla työkyvyttömyyden taloudellinen merkitys on valtava väestön ikääntyessä. Sairauspäivärahojen ja työkyvyttömyyseläkkeiden ohella on huomioitava myös työkykyisten työntekijöiden merkitys veronmaksajina ja toisaalta yhteiskunnan sujuvan toiminnan mahdollistajina työtehtäviensä kautta. Työkyvyttömyyden ehkäisy ja työurien pidentäminen onkin jossain muodossa noussut kaikkiin Suomen hallitusohjelmiin ainakin viimeisten parinkymmenen vuoden aikana. Työkyvyttömyyden ehkäisy on tunnistettu keskeiseksi tavoitteeksi myös laajalti OECD-maissa².

Suomen työterveyshuoltojärjestelmä poikkeaa useista muista siinä, että valtaosa työnantajista tarjoaa työntekijöille myös sairaanhoidon palveluita lakisääteisten ennaltaehkäisevien palveluiden ohella. Työterveyshuoltolain ja asetuksen hyvästä työterveyshuoltokäytännöstä henki on ollut, että sairaanhoito tulisi järjestää samojen henkilöiden toimesta kuin ennaltaehkäisevä työ, jotta sairausvastaanotoilta kertyvää tietoa pystyttäisiin hyödyntämään esimerkiksi työkyvyn tuen tarpeessa olevien yksilöiden tunnistamiseen. Työterveyshuollon osajista on kuitenkin tällä hetkellä pula, jonka myötä sairaanhoito saattaa jäädä irralliseksi ja hajaantua suurelle joukolle sairausvastaanoton henkilöstöä. Tämä haastaa työkykyongelmien oikea-aikaista tunnistamista ja sairausvastaanotoilta kertyvän tiedon kulkua yrityksen nimetyille vastuuhenkilöille työterveyshuollossa ja tämän tiedon hyödyntämistä ennaltaehkäisevässä työssä. Uusien sähköisten työkalujen kehittäminen ja hyödyntäminen työkykyriskin tunnistamiseksi on yksi mahdollisuus vastata tähän haasteeseen.

Työterveyshuollon palveluiden kohdentaminen työkykyriskissä oleviin henkilöihin on todettu kustannusvaikuttavaksi keinoksi ennaltaehkäistä työkyvyttömyyttä ja siitä aiheutuvia kustannuksia³. Työkykyriskin tunnistamisesta on saatu lupaavaa tutkimusnäyttöä mm. terveyskyselyitä hyödyntämällä^{4,5}.

Työkyvyttömyys

Työkyvyttömyyden määritelmä vaihtelee riippuen viitekehyksestä, jossa asiaa tarkastellaan. Tässä tutkimuksessa pyrittiin tunnistamaan potilastietojärjestelmästä ja terveystietojärjestelmästä saatavia työkyvyttömyyden riskitekijöitä suhteessa työterveyshuollon tietojärjestelmiin kirjattuihin sairauspoissaolotietoihin. Näin ollen tutkimuksessa analyysien lähtökohtana on ollut sairausvakuutuslain mukainen työkyvyn määritelmä. Työkyvyttömyydellä tarkoitetaan tällöin sellaista sairaudesta johtuvaa tilaa, jonka kestäessä vakuutettu on sairauden edelleen jatkuessa kykenemätön tekemään tavallista työtään tai työtä, joka on siihen läheisesti verrattavaa⁶.

Työkykyyn ei vaikuta vain yksilön terveys, vaan se on työn ja ihmisten voimavarojen välinen suhde. Terveystietojärjestelmien ohella mm. osaaminen, johtaminen ja työolot vaikuttavat työkykyyn. Työkyvyn tukemisen onnistumisen edellytyksiä ovat jaettu käsitys työkyvystä, yhteinen sopiminen omasta ja toisten rooleista sekä selkeät prosessit kohti toimijoiden yhteistä päämäärää⁷. Työkyvyn arviointi on työterveyshuollon erityisosaamista. Työn ja terveyden välistä suhdetta työterveyslääkäri arvioi terveystarkastuksissa, työpaikkaselvityksissä ja sairausvastaanotolla. Työn ja työpaikan tuntemus mahdollistaa työterveyslääkäreille työolosuhteiden muokkaamismahdollisuuksien tunnistamisen kuntoutus-, työhön paluun ja työllistymisen tukiprosesseissa⁸. Työkykyä uhkaavat tekijät on mahdollista tunnistaa näitä riskitekijöitä seulovilla tutkimuksilla⁹.

Työterveyshuolto on terveydenhuollon kentässä ainoa taho, jolla on suora yhteys työpaikoille. Tämän yhteistyön kautta työterveyshuollolla on ainutlaatuisia vaikutusmahdollisuuksia - ei vain yksilön terveyteen vaan myös työhön liittyviin tekijöihin yhteistyössä työnantajan kanssa. Tunnistamalla työkykyriskejä niin yksilö- kuin työpaikkatasolla on mahdollista edistää sekä yksilön että koko työyhteisön työkykyä työpaikkaan kohdistuvien toimenpiteiden avulla.

Jotta työkyvyn heikkenemiseen pystytään puuttumaan entistä ennaltaehkäisevämmiin ja oikea-aikaisempiin kohdennettuihin työterveyshuollon interventioihin, tulee pyrkiä entistä paremmin yhdistämään sekä sairauskertomuksesta että työterveyskyselyistä saatavaa tietoa ja hyödyntämään edistyneimpiä mahdollisia analytiikkamenetelmiä. On tuloksellisempaa kohdentaa työterveyshuollon tukitoimet henkilöihin, joilla on tunnistettu työkykyriski, kuin tarjota kohdentamattomia toimenpiteitä⁷.

Tavoitteet

Tämän hankkeen tavoitteena oli syventää ymmärrystä sairauspoissaolojen riskitekijöistä ja työkyvyttömyysriskin tunnistamisesta, jotta voitaisiin ohjata työterveyshuollon rajalliset resurssit paremmin niiden henkilöiden auttamiseen, joilla on lisääntynyt riski sairauspoissaoloille. Aiempaan tutkimukseen pohjautuen tiedetään tiettyjen sairauksien lisäävän sairauspoissaoloja^{10–12}. On myös näyttöä siitä, että terveystarkastuksia pystytään hyödyntämään työkyvyttömyyden ennustamisessa ja että työterveyshuollon toimien kohdentaminen riskihenkilöihin on kustannusvaikuttavaa¹³. Tässä tutkimuksessa pyrittiin laajan aineiston avulla kehittämään uusia aiempaa tarkempia ja arjen työssä paremmin sovellettavia menetelmiä työterveyshuollon interventioiden kohdentamisen tueksi.

Työkyvyttömyyden riskitekijöihin suunnatut tutkimukset on usein tehty tietyllä kohderyhmällä kuten yksittäisellä työpaikalla tai toimialalla, mikä haastaa tulosten yleistettävyyttä. Kun huomioidaan työterveyshuollon palvelujärjestelmän sekä sosiaalivakuutuksen erilaiset mallit eri maissa, on kansainvälisten tutkimusten tuloksia vain rajallisesti mahdollista soveltaa suomalaiseen työelämään, minkä vuoksi suomalaiset tutkimukset tuovat arvokasta lisätietoa. Tässä tutkimuksessa pyrittiin syventämään sairauspoissaolojen riskitekijöihin liittyvää ymmärrystä siten, että se olisi sovellettavissa suomalaisiin eri toimialojen yrityksiin. Keskeisenä teemana oli työterveyshuoltoon jatkuvasti kertyvän tiedon hyödyntämisen erilaiset mahdollisuudet sairauspoissaolojen ennustamiseksi.

Aiemmassa tutkimuskirjallisuudessa on esitetty tutkimuksia, joissa koneoppiva malli ennustaa potilaskohtaista riskiä pitkiin sairauspoissaoloihin, usein kohdistettuna tiettyyn sairausryhmään^{14–20} ja joissain tutkimuksissa yleisellä potilastasolla^{9,21–26}. Tutkimuksissa on pääsääntöisesti käytetty perinteisiä tilastollisia malleja, joiden avulla ei ole mahdollista löytää monimutkaisia muuttujavaikutuksia ja -kombinaatioita, joilla on riskiä kasvattava tai vähentävä vaikutus. Tässä tutkimuksessa etsittiin aiempaan tutkimukseen verrattuna kehittyneemmällä koneoppimiseen perustuvilla algoritmeilla työkykyriskiin vaikuttavien työterveyskysymysten ja sairauskertomustietojen kombinaatioita ja sopivia kynnyksarvoja.

Tarkastelua vietiin aiempaa tutkimuskirjallisuutta pidemmälle tunnistamalla korkeariskisten potilaiden joukosta ryhmiä, joissa työkykyilmiöt ja työkykyä uhkaavat tekijät ovat erilaisia. Tieto korkeariskisistä potilasryhmistä ja niiden tunnistamisesta auttaa kohdentamaan tarkemmin tarvittavia ennaltaehkäiseviä toimia.

Tutkimuksella pyrittiin vastaamaan seuraaviin kysymyksiin:

1. Millaisia potilasryhmiä tunnistetaan datalähtöisesti työterveyskyselyjen, potilastietojen ja sairauspoissaolokäyttäytymisen perusteella?
2. Miten tarkasti ja millaisella aikajänteellä työterveyskyselyvastausten ja sairauskertomuksesta kerättävän tiedon perusteella voidaan tunnistaa henkilöt, joilla on lisääntynyt riski pitkittyvään työkyvyttömyyteen?
3. Mitkä työterveyskyselyn kysymykset, sairauskertomukseen rakenteisesti kirjatut muuttujat ja näiden yhdistelmät ennustavat parhaiten työkyvyn heikkenemistä?
4. Onko työterveyskyselyvastausten tai sairauskertomukseen rakenteisesti kirjatun tiedon pohjalta tunnistettavissa tekijöitä, jotka suojaavat henkilöä pitkittyvältä työkyvyttömyydeltä?
5. Millainen on työterveyskyselyyn vastanneen joukon edustavuus verrattuna Suomen työntekijäjoukkoon?

Ennustemalleja rakennettaessa päädyttiin otoksen jakamiseen sairauspoissaolojen määrän mukaan kahteen joukkoon, 30 vuosittaisen sairauspoissaolopäivän toimiessa raja-arvona. Toisaalta kun jo 15 vuosittaisen sairauspoissaolopäivän on todettu ennustavan pysyvää työkyvyttömyyttä, haluttiin testata myös muiden sairauspoissaolorajojen toimivuutta. Arviomme oli, että toistuviin lyhyisiin poissaoloihin vaikuttavat tekijät ainakin jossain määrin eroaisivat pitkien sairauspoissaolojen riskitekijöistä.

Työkyvyttömyyden kustannuksista valtaosa liittyy tuki- ja liikuntaelin- sekä mielenterveyden sairauksien pohjalta maksettaviin etuuksiin. Näiden diagnoosiryhmien sisällä on kuitenkin hyvin heterogeeninen joukko henkilöitä, joiden tuen tarve vaihtelee. Laajasta työterveyskyselyaineistosta pyrittiin löytämään sellaiset olennaiset kysymykset, jotka sairauskertomustietoon yhdistettynä auttaisivat jatkossa sairauspoissaoloennusteen ja muiden tekijöiden osalta erilaisiin ryhmiin kuuluvien henkilöiden tunnistamisessa. Tämä mahdollistaisi tulevaisuudessa suppeamman kyselyn muokkaamisen, joka olisi helposti toistettava ja johon voisi vastata esimerkiksi työterveyshuollon matkapuhelinsovelluksella.

Näin kertyvä tieto auttaisi työterveystiimiä tunnistamaan erityisessä riskissä olevat henkilöt ja toisaalta suojaavat tekijät.

Tässä raportissa kuvattu tutkimushanke on osa suurempaa kokonaisuutta. Pitemmän aikavälin tavoitteena on arvioida tunnistettujen riskiryhmien palvelupolkuja ja niiden kehittämistarvetta, ennaltaehkäistä sairastumista ja työkyvyttömyyttä sekä toisaalta pyrkiä vähentämään työelämään liittyvien ilmiöiden medikalisoimista. Tämä avaa myös raportointimahdollisuuksia yksittäisen asiakasyrityksen henkilöstön jakautumisesta tunnistettuihin riskiryhmiin silloin, kun yritys on riittävän suuri raportointiin yksilönsuojaa vaarantamatta. Tietoa pystyttäisiin hyödyntämään yrityksen keskeisten työkykyriskien hallinnassa, toimenpiteiden suunnittelemisessa ja niiden vaikuttavuuden arvioimisessa.

Menetelmät

Tutkimus pohjautui kehittyneeseen data-analyysiin sekä koneoppiviin algoritmeihin.

Koneoppivia malleja käytettiin löytämään merkitsevimmät riskitekijät suuresta joukosta muuttujia ja ennustamaan riskitekijöiden avulla potilaskohtaisen riskin halutulle päätemuuttujalle (kuten riski pitkään sairauspoissaoloon). Tutkimuksessa sovellettiin koneoppivia menetelmiä, kuten random forest, tukivektorkone (support vector machine, SVM), gradient boosting sekä neuroverkoja. Perinteisempiä lineaarisia malleja, kuten logistinen regressio, käytettiin verrokkituloksena.

Mallien toimintaa ja yksittäisten muuttujien sekä muuttujakombinaatioiden vaikutusta ennustuksissa tutkittiin XAI-menetelmillä (explainable machine learning) sekä herkkyysanalyysillä (partial dependence plots, PDP). XAI-menetelmillä, kuten Shapley values^{27,28}, määritettiin mitkä tekijät ovat tärkeitä ennustuksissa erityisesti yksittäisten potilaiden osalta. XAI-menetelmien avulla vastattiin esimerkiksi kysymyksiin siitä, mitkä tekijät vaikuttivat yksittäisen potilaan matalaan tai korkeaan ennustettuun riskiin. Herkkyysanalyysin²⁹ avulla karakterisoitiin monimutkaisien epälineaaristen mallien toimintaa koko potilasjoukon tasolla. Toisin sanoen, miten esimerkiksi muutos tietyn muuttujan arvossa vaikutti keskimäärin riskiin. Esimerkiksi miten BMI-arvon lasku luvusta 30 lukuun 25 vaikutti keskimäärin ennustettuun riskiin.

Erilaisten potilasryhmien eli ns. potilasprofiilien tunnistamisessa hyödynnettiin pääkomponenttianalyysiä (principal component analysis, PCA) sekä klusterointimenetelmiä (esimerkiksi K-means clustering). Pääkomponenttianalyysi etsi usean kymmenen muuttujan aineistosta latentit dimensiot, jotka kuvasivat potilasjoukkoa mahdollisimman tarkasti. Toisin sanoen, pääkomponenttianalyysi yhdisti samaa asiaa mittaavat muuttujat yksittäisiksi uusiksi muuttujiksi, jotka kuvasivat tiettyä potilasominaisuutta. Klusterointi toteutettiin pääkomponenttianalyysin määrittämille latenteille dimensioille. Tutkimuksessa analysoitiin optimaalisen pääkomponenttianalyysin ja klusteroinnin lopputuloksena löytyviä potilassegmenttejä ja niiden ominaisuuksia liittyen esimerkiksi tuleviin sairauspoissaoloihin.

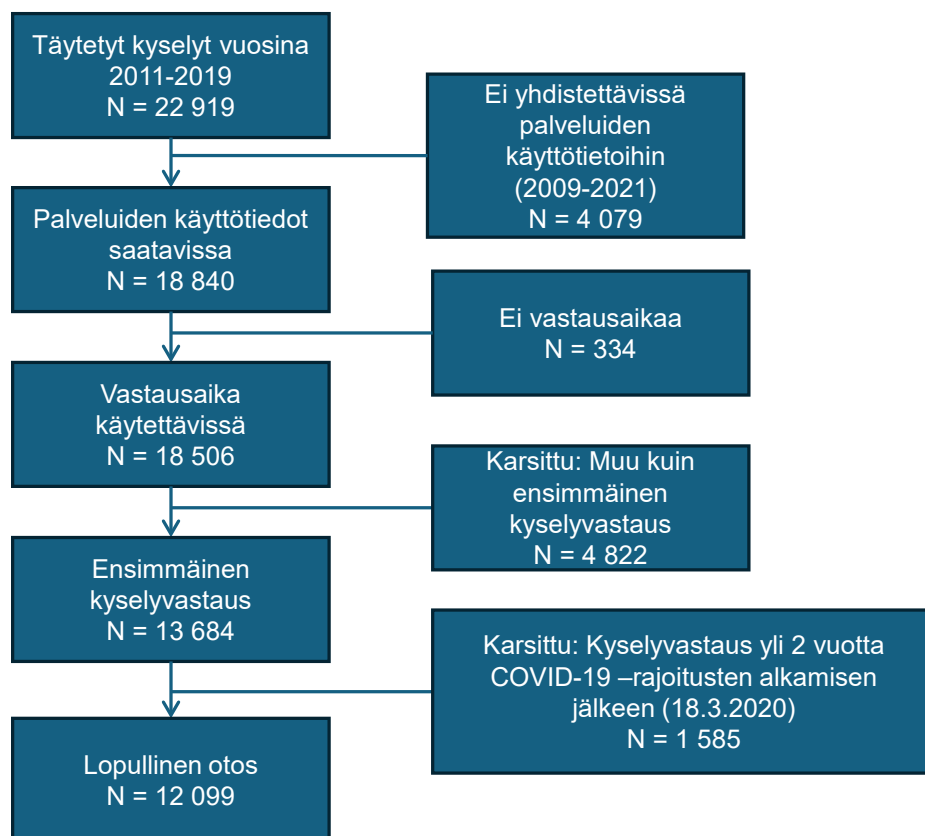
Aineisto

Tutkimusaineistona oli Finlan hoidossa olevien työntekijöiden työterveyskyselyitä vuosilta 2011–2019 sekä sairauskertomustiedot Finlan hoidossa tällä ajanjaksolla olleilta työntekijöiltä. Tutkimusotoksen muodostivat ne noin 18000 henkilöä, joilta löytyi sekä terveyskyselyvastaukset että sairauskertomustiedot Finlan potilastietojärjestelmästä. Sairauskertomustietoja oli käytettävissä myös kahden vuoden ajalta ennen kyselyajankohtaa sekä kaksi vuotta kyselyajankohdan jälkeen, jotta pystyttiin hyödyntämään kyselyä edeltäviä sairauspoissaolotietoja ennustemalleissa sekä käyttämään kyselyn jälkeisten kahden vuo-

den sairauspoissaolotietoja päätemuuttujina. Analyyseistä päädyttiin lopulta rajaamaan pois 18.3.2018 jälkeen täytetyt kyselyt, sillä niiden osalta COVID-19 olisi oleellisesti muuttanut sairauspoissaolokäyttäytymistä kahden vuoden seuranta-aikana ja vaikeuttanut tulosten tulkintaa.

Tutkimusaineistona toiminut terveystarkastus on ollut käytössä sekä lakisääteisissä että vapaaehtoisissa terveystarkastuksissa sekä työsuhteen alkaessa että terveydenseuranta- tai määräaikaistarkastuksissa. Kysely on ollut mahdollista täyttää sekä sähköisenä että paperisena ja paperilla täytettyjen kyselyiden vastaukset on siirretty sähköiseen muotoon ja ovat osana aineistoa.

Hankkeelle haettiin tutkimuslupa sosiaali- ja terveysalan tietolupaviranomaiselta Findatalta. Aineistoa käsiteltiin anonymisoituna Findatan tietoturvalisissä Kapseli-ympäristössä, johon oli pääsy vain rajatulla joukolla tutkimusryhmän jäseniä.



Tutkimuksen kulku. Tutkimukseen valittiin työntekijöitä, jotka olivat täyttäneet työterveyskyselyn vähintään kerran vuosina 2011-2019 ja joiden kyselyt pystyttiin yhdistämään palveluiden käyttötietoihin työterveyshuollon sähköisessä järjestelmässä. Henkilöt, jotka olivat vastanneet kyselyyn korkeintaan kaksi vuotta ennen COVID-19-rajoitusten alkamista karsittiin tutkimuksesta. Myös työntekijät, joiden kyselyvastaukseen ei ollut kirjattu täyttöajankohtaa rajattiin pois. Lopullinen tutkimusotos sisälsi 12 099 työntekijän tiedot.

Muuttujat

Aineistona toiminut kysely sisältää laajasti kysymyksiä eri aiheista, joista analyysija varten käytettiin sellaisia, joilla arveltiin olevan mahdollisia yhteyksiä sairauspoissaoloihin. Näitä olivat esimerkiksi henkilön itse ilmoittamat sairaudet ja oireet sekä oma arvio näiden vaikutuksesta hänen työkykyynsä. Terveystottumuksista käytettävissä oli tietoja mm. henkilön ruokavaliosta, liikuntatottumuksista, alkoholin ja tupakan käytöstä. Työhön liittyvistä kysymyksistä poimittiin tiedot mahdollisista epäsäännöllisistä työajoista (osa-aikainen työ, vuorotyö, yötyö), kuuluminen toimihenkilöihin tai työntekijöihin sekä mahdollinen esimiesasema.

Käytettävissä oli myös tietoja työn psykososiaalisista kuormitustekijöistä sekä vastaajan arvio työpaikkansa esihenkilötyöstä ja ilmapiiristä. Mielialaoireita kartoitettiin paitsi DEPS-kyselyllä, joka on validoitu menetelmä masennuksen seulonnassa, myös kahden kysymyksen seulalla, joka on laajasti terveydenhuollossa käytetty suppeampi matalan kynnyksen työkalu mielialaoireiden tunnistamiseen. Keskeisistä työkykyyn vaikuttavista tekijöistä poimittiin kyselystä myös mm. henkilön oma arvio työkyvystään ja sen mahdollisista muutoksista, työstä palautumisesta sekä siitä, arveleeko vastaaja olevansa työkykyinen nykyiseen työhönsä kahden vuoden kuluttua

Potilastietojärjestelmästä saatavassa aineistossa pystyttiin hyödyntämään mm. sairauspoissaolopäivien, sairauspoissaolajaksojen ja vastaanottokäyntien lukumäärät, vastaanottavat ammattiryhmät ja diagnoositiedot. Sairauspoissaolotiedot kattoivat myös ensimmäiset 1–9 poissaolopäivää ennen Kelan korvaamia sairauspoissaolopäiviä. Sairausvakuutuslain 13 luvun 5 § muutoksessa joulukuussa 2010 ennaltaehkäisevän työn korkeampi Kela-korvaus sidottiin työkyvyn seurannan ja varhaisen tuen malleihin, minkä myötä myös muualla kuin työterveyshuollossa kirjatut sairauspoissaolot on tutkimusjaksolla jo kattavasti toimitettu työterveyshuollon tietojärjestelmiin. Tämä parantaa tiedon luotettavuutta ja tekee sen vertailukelpoisemmaksi työterveyshuollossa tyypillisesti käytettävissä olevien sairauspoissaolotietojen kanssa.

Masennusoireet ja työn psykososiaaliset kuormitustekijät sairauspoissaoloriskin ennustajina

Ensimmäisessä jo julkaistussa artikkelissa keskityimme masennuksen diagnostiikassa käytettyyn DEPS-seulontakyselyyn sekä esihenkilötyötä ja työn psykososiaalisia kuormitustekijöitä ja esihenkilötyötä kartoittaviin kysymyksiin³⁰. Päätemuuttujina käytimme kahden vuoden seuranta-ajalla kyselyn täyttämisen jälkeen toteutuneita pitkiä yli 30 päivän sairauspoissaolojaksoja sekä vähintään kuudesti kahden vuoden seurantajaksolla toistuneita lyhyitä 1–10 päivän sairauspoissaolojaksoja.

Masennusoireista keskeisimmät sairauspoissaoloja ennustavat kysymykset olivat ”Tuntuuko tulevaisuus toivottomalta”, ”Tunnetko itsesi surumieliseksi”, ”Tuntuuko sinusta, että kaikki vaatii ponnistusta” ja ”Tunnetko itsesi arvottomaksi”. Vastaukset ”melko paljon” ja ”erittäin paljon” lisäsivät kaikkien näiden kysymysten kohdalla sekä toistuvien lyhyiden että pitkien sairauspoissaolojen todennäköisyyttä kahden vuoden seuranta-aikana. Työn psykososiaaliin kuormitustekijöihin ja esihenkilötyöhön liittyvistä kysymyksistä sairauspoissaoloja ennustivat huonot vaikutusmahdollisuudet omaan työhön, huonot mahdollisuudet kehittyä työssään, työn vähäinen vaihtelevuus sekä kokemus siitä, että vastaajan työtä ei arvosteta.

DEPS-kysymykset ennustivat paremmin pitkiä sairauspoissaoloja kuin esihenkilötyöhön ja työn psykososiaaliin kuormitustekijöihin liittyvät kysymykset. Myös toistuvien lyhyiden poissaolojen ennustamisessa DEPS-kysymykset olivat valtaosin työn psykososiaaliin kuormitustekijöihin liittyviä kysymyksiä parempia. Lyhyitä, toistuvia sairauspoissaoloja ennusti kuitenkin parhaiten vastaus ”osittain eri mieltä” tai ”eri mieltä” väitteeseen ”voin vaikuttaa työni sisältöön, työtahtiini ja työaikaani”, joka oli osa psykososiaalisen kuormituksen kyselykokonaisuutta.

DEPS-seula puoltaa joka tapauksessa paikkansa validoituna masennusoireiden seulontatyökaluna, mutta tutkimustulostemme perusteella osalla kysymyksistä on arvoa myös sairauspoissaoloriskin ennakoimisessa. Työn psykososiaalisten kuormitustekijöiden kysymykset voivat edelleen toimia hyvinä keskustelunavaajina vastaanotolla, mutta niiden yhteys sairauspoissaoloihin oli vähäisempi lukuun ottamatta kysymystä koskien vaikutusmahdollisuuksia omaan työhön.

Sairauspoissaolojen ennustemallit

Toisessa tutkimusjulkaisussa opetettiin koneoppimisen menetelmillä sairauspoissaolojen ennustemalleja. Aineistosta prosessoitiin 129 muuttujaa, jotka jaettiin teemoittain kolmeen kategoriaan, joiden alle muodostettiin yhteensä 12 eri teemoja käsittelevää alamallia: Demografiset tekijät, työnkuva, mittaukset, terveystottumukset, sairaudet ja oireet, mieliala, psykososiaaliset tekijät, työkyky, esihenkilötyö ja työilmapiiri sekä käyntidiagnoosit, palveluiden käyttö ja sairauspoissaolot. Aiempaan tutkimukseen perustuen oli odotettavissa, että käyntidiagnoosit, palveluiden käyttö ja sairauspoissaolot ennustaisivat parhaiten tulevia sairauspoissaoloja, minkä vuoksi nämä muodostivat referenssiaineiston kehitettävien ennustemallien suorituskyvyn arviointiin. Päätemuuttujana käytettiin pitkiä, yli 30 päivän sairauspoissaoloja kahden vuoden seuranta-aikana terveystarkastuksen täyttämishetkestä laskien. Alamallien toimintaa karakterisoitiin XAI menetelmillä.

XAI-analyysin perusteella potilaat, joilla oli kipuja, yli kuusi oiretta tai 2–3 sairautta, painoindeksi (BMI) ≥ 27 , nukkuivat 6 tuntia tai vähemmän tai joiden työkykypistemäärä oli 7 tai vähemmän, oli suurentunut riski pitkiin sairauspoissaoloihin, etenkin jos he työskentelivät työntekijäasemassa, osa-aikatyössä tai vuorotyössä. Lisäksi useat työhön ja mielenterveyteen liittyvät muuttujat lisäsivät pitkien sairauspoissaolojen riskiä. Osamalleista korkein suorituskky oli ”Palvelun käyttö” -alimallille (AUROC=0.74). Ennustustarkkuus parani sisällyttämällä muuttujia muista alamalleista, jolloin AUROC-arvo nousi tasolle 0,79.

Työkykyriskiprofiilien tunnistaminen

Kolmannessa osatutkimuksessa hyödynnettiin datapohjaista klusterointia tunnistamaan työntekijöiden joukosta alaryhmiä, joilla olisi keskenään samankaltaiset riskitekijäprofiilit. Tavoitteena oli ymmärtää, millaisia erilaisia ryhmiä työntekijöiden joukosta löytyy, millaiset ovat näiden ryhmien toteutuneet sairauspoissaolot ja näiden pohjalta tunnistaa, millaisista ennaltaehkäisevistä toimenpiteistä eri ryhmät voisivat hyötyä.

Analyyseissä hyödynnettiin samoja muuttujia kuin aiemmissakin tutkimuksissa. Koska muuttujia oli paljon (104) ja niistä monet korreloivat keskenään, tehtiin ensin pääkomponenttianalyysi muuttujien määrän vähentämiseksi. Pääkomponenttianalyysin ominaisarvojen (eigenvalues) perusteella valittiin viisi komponenttia. Pääkomponentit nimettiin sen mukaan, mitkä muuttujat latautuivat millekin pääkomponentille: PK1: mieliala ja masennusoireet, PK2: esihenkilötyö ja työilmapiiri, PK3: työkyky, PK4: huimaus ja aistioireet ja PK5: sydän- ja verisuonisairaudet.

Tämän jälkeen toteutettiin klusterointi K-means-algoritmilla, johon syötettiin alkuperäisten muuttujien sijaan pääkomponenttien arvot. Silhoutte-arvon perusteella analyysistä valikoitui

kuusi klusteria, joiden koot vaihtelivat välillä 540–5500 työntekijää. Klusteriin 1 kuuluivat ne, joilla oli mieliala- ja masennusoireita (PK1 koholla), klusteriin 2 puolestaan kuuluivat ne, joilla oli sydän- ja verisuonisairauksia (PK5). Klusterissa 3 olevilla ei ollut mainittavia riskitekijöitä ja lisäksi he kokivat esihenkilötyön ja työpaikan ilmapiirin positiivisena (PK2). Klusteriin 4 kuuluivat ne, jotka kokivat työkykynsä huonoksi (PK3) ja klusteriin 5 ne, jotka raportoivat erilaisia aistioireita (PK4). Klusteriin 6 kuuluivat ne, joilla ei ollut sairauksia tai oireita, mutta jotka olivat tyytymättömiä esihenkilötyöhön tai kokivat työilmapiirin huonoksi (PK2).

Klusterien ominaisuuksia analysoitiin vielä erillisellä XGBoost-luokittelumallilla, joka pyrki ennustamaan jokaisen työntekijän kuulumista eri klustereihin. Malli hyödynsi kaikkia alkupe räisiä 104 muuttujaa, jotta pystyttiin kuvaamaan, mitkä yksittäiset muuttujat ennustivat parhaiten klustereihin kuulumista. Shapley-arvoja käytettiin muuttujien ennustusvaikutuksen kuvaamisessa. Pääkomponenttien kuvaamien muuttujien lisäksi havaittiin, että ikä vaikutti todennäköisyyteen kuulua tiettyyn klusteriin: klusteriin 1 kuuluvat olivat tyypillisesti nuoria, kun taas klusteriin 2 kuuluivat todennäköisemmin vanhemmat työntekijät.

Työntekijöiden sairauspoissaolojen määrät ja kestot eivät olleet mukana klusterointia tehtäessä, vaan klusterit muodostettiin puhtaasti taustatietojen ja riskitekijöiden pohjalta. Klusterien muodostamisen jälkeen analysoitiin, miten klusterit assosioituvat joko useisiin lyhyisiin tai pitkiin sairauspoissaoloihin. Kun muita klustereita verrattiin klusteriin 3 (terveet), huomattiin, että klusterissa 4 (koettu huono työkyky) oli korkein riski pitkille sairauspoissaoloille ja klusterissa 2 riski oli myös kohonnut. Klustereissa 5 (aistioireet) ja 6 (tyytymättömiä esihenkilötyöhön ja työilmapiiriin) oli puolestaan kohonnut riski lyhyille sairauspoissaoloille.

Johtopäätökset ja tulosten hyödyntäminen

Tutkimushankkeessa havaittiin, että tietyillä masennusoireiden kartoittamiseen käytettävän DEPS-seulan kysymyksillä pystytään ennakoimaan sekä toistuvia lyhyitä että pitkiä sairauspoissaoloja. Kokemukset vaikutusmahdollisuuksista omaan työhön vaikuttavat toistuvien lyhyiden sairauspoissaolojen esiintyvyyteen – hallinnan tunne työssä on yksilölle tärkeää.

Hyödyntämällä terveystarkastusta ja työterveyshuollon potilastietojärjestelmästä poimittuja tietoja saatiin luotua ennustemalleja, joilla pystyttiin varsin hyvin ennustamaan riskiä pitkään sairauspoissaolajaksoon kahden vuoden seuranta-aikana. Aiemmat sairauspoissaolot, palveluiden käyttö ja diagnoosit olivat odotetusti keskeisiä muuttujia mallin rakentamisessa, mutta näiden ohella löydettiin useasta datakategorista tiettyjä muuttujia, jotka toimivat muita paremmin sairauspoissaoloriskin ennustajina ja joihin työterveyshuollon toimenpiteitä olisi mahdollista kohdistaa. Näiden muuttujien lisääminen ennustemalliin paransi mallin kykyä tunnistaa riskiä pitkiin sairauspoissaoloihin.

Aineisto kerättiin yhdeltä suomalaiselta työterveyshuollon palveluntuottajalta, jonka asiakasyritysten toimialajakauma oli aineiston keruuajana varsin samankaltainen kuin suomalaisten yritysten toimialajakauma. Tämä parantaa tutkimuksen tulosten yleistettävyyttä laajemmin suomalaiseen työelämään. Sairauspoissaolotiedot kattoivat myös ensimmäiset 1–9 poissaolopäivää ennen Kelan korvaamia sairauspoissaolopäiviä, mikä on yksi tutkimuksen vahvuuksista.

Modernien työterveyspalveluiden tulee kyetä sähköisten työvälineiden avulla tunnistamaan tukea tarvitsevat asiakasorganisaatiot, henkilöryhmät ja yksilöt sekä tarjoamaan näille vaikuttavia palveluita. Tämä tutkimus pyrki syventämään tietoa tukea tarvitsevien yksilöiden tunnistamisesta yhdistämällä sairauskertomusten ja terveystarkastusvastausten tietoa. Lisäksi kehittyneemmät koneoppimiseen perustuvat klusterointimenetelmät mahdollistivat erilaisen riskimuuttujien yhdistelmien suhteen samankaltaisten potilasjoukkojen tunnistamisen. Tunnistettujen potilasjoukkojen pohjalta voidaan jatkossa arvioida tarkemmin kunkin joukon tarpeita erilaisille toimintamalleille ja palveluille sekä tunnistaa työkykyriskiä paitsi yksilö- myös työpaikkatasolla. Tämä mahdollistaa tulevaisuudessa myös toimenpiteiden vaikuttavuuden seurannan kertyvän potilas- ja organisaatiodatan jatkuvan analysoinnin perusteella.

Työsuojelurahasto on osallistunut tämän hankkeen rahoittamiseen.

Lähteet

1. Rissanen M, Kaseva E. Menetetyn työpanoksen kustannus. Sosiaali- ja terveysministeriön työsuojeluosasto, toimintapolitiikkayksikkö; 2014.
2. OECD. Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries. OECD. Epub ahead of print November 24, 2010. DOI: 10.1787/9789264088856-en.
3. Taimela S, Aronen P, Malmivaara A, Sintonen H, Tiekso J, Aro T. Effectiveness of a Targeted Occupational Health Intervention in Workers with High Risk of Sickness Absence: Baseline Characteristics and Adherence as Effect Modifying Factors in a Randomized Controlled Trial. *J Occup Rehabil.* 2010;20:14–20.
4. Lundin A, Kjellberg K, Leijon O, Punnett L, Hemmingsson T. The Association Between Self-Assessed Future Work Ability and Long-Term Sickness Absence, Disability Pension and Unemployment in a General Working Population: A 7-Year Follow-Up Study. *J Occup Rehabil.* 2016;26:195–203.
5. Pihlajamäki M, Uitti J, Arola H, et al. Self-reported health problems and obesity predict sickness absence during a 12-month follow-up: a prospective cohort study in 21 608 employees from different industries. *BMJ Open.* 2019;9:e025967.
6. Sairausvakuutuslaki 1224/2004. 1224/2004; Oikeusministeriö, Edita Lakitieto Oy Available from: <https://www.finlex.fi/fi/laki/ajantasa/2004/20041224#O3L8P4>.
7. Juvonen-Posti P, Tarvainen K, Helin-Salmivaara A, Räsänen K, Liira J. Lääkäri työkykyä tukemassa. *Lääketieteellinen Aikakauskirja Duodecim.*;2021:293–300.
8. Sauni R, Uitti J. Sairaanhoido. Työstä terveyttä. *Duodecim*; 2018. p. 49–56.
9. Roelen CA, Bültmann U, Van Rhenen W, Van Der Klink JJ, Twisk JW, Heymans MW. External validation of two prediction models identifying employees at risk of high sickness absence: cohort study with 1-year follow-up. *BMC Public Health.* 2013;13:105.
10. de Vroome EMM, Uegaki K, van der Ploeg CPB, et al. Burden of Sickness Absence Due to Chronic Disease in the Dutch Workforce from 2007 to 2011. *J Occup Rehabil.* 2015;25:675–684.
11. Claessen H, Arndt V, Drath C, Brenner H. Overweight, obesity and risk of work disability: a cohort study of construction workers in Germany. *Occup Environ Med.* 2009;66:402–409.
12. Ahola K, Virtanen M, Honkonen T, Isometsä E, Aromaa A, Lönnqvist J. Common mental

disorders and subsequent work disability: a population-based Health 2000 Study. *J Affect Disord.* 2011;134:365–372.

13. Taimela S, Justen S, Aronen P, et al. An occupational health intervention programme for workers at high risk for sickness absence. Cost effectiveness analysis based on a randomised controlled trial. *Occup Environ Med.* 2008;65:242–248.
14. Holm J, Frumento P, Almondo G, et al. Predicting the duration of sickness absence due to knee osteoarthritis: a prognostic model developed in a population-based cohort in Sweden. *BMC Musculoskelet Disord.* 2021;22:603.
15. van Hoffen MFA, Norder G, Twisk JWR, Roelen CAM. Development of Prediction Models for Sickness Absence Due to Mental Disorders in the General Working Population. *J Occup Rehabil.* 2020;30:308–317.
16. Notenbomer A, Van Rhenen W, Groothoff JW, Roelen CAM. Predicting long-term sickness absence among employees with frequent sickness absence. *Int Arch Occup Environ Health.* 2019;92:501–511.
17. Bosman LC, Roelen CAM, Twisk JWR, Eekhout I, Heymans MW. Development of Prediction Models for Sick Leave Due to Musculoskeletal Disorders. *J Occup Rehabil.* 2019;29:617–624.
18. Norder G, Roelen CAM, Van Der Klink JJJ, Bültmann U, Sluiter JK, Nieuwenhuijsen K. External Validation and Update of a Prediction Rule for the Duration of Sickness Absence Due to Common Mental Disorders. *J Occup Rehabil.* 2017;27:202–209.
19. Roelen CAM, Van Hoffen MFA, Waage S, et al. Psychosocial work environment and mental health-related long-term sickness absence among nurses. *Int Arch Occup Environ Health.* 2018;91:195–203.
20. Real E, Jover L, Verdaguer R, et al. Factors Associated with Long-Term Sickness Absence Due to Mental Disorders: A Cohort Study of 7.112 Patients during the Spanish Economic Crisis. *PLOS ONE.* 2016;11:e0146382.
21. van der Burg LRA, van Kuijk SMJ, ter Wee MM, et al. Long-term sickness absence in a working population: development and validation of a risk prediction model in a large Dutch prospective cohort. *BMC Public Health.* 2020;20:699.
22. Airaksinen J, Jokela M, Virtanen M, et al. Prediction of long-term absence due to sickness in employees: development and validation of a multifactorial risk score in two cohort studies. *Scand J Work Environ Health.* 2018;44:274–282.
23. Schouten LS, Joling CI, van der Gulden JW, Heymans MW, Bültmann U, Roelen CA. Sc-

reening manual and office workers for risk of long-term sickness absence: cut-off points for the Work Ability Index. *Scand J Work Environ Health*. 2015;41:36–42.

24. Roelen CAM, Bultmann U, Stapelfeldt CM, Jensen C, Heymans MW. Multicentre validation of frequent sickness absence predictions: Table 1. *Occup Med*. 2016;66:69–71.
25. Roelen CAM, Heymans MW, Twisk JWR, et al. Health measures in prediction models for high sickness absence: single-item self-rated health versus multi-item SF-12. *Eur J Public Health*. 2015;25:668–672.
26. Roelen CA, Van Rhenen W, Groothoff JW, Van Der Klink JJ, Bultmann U, Heymans MW. The development and validation of two prediction models to identify employees at risk of high sickness absence. *Eur J Public Health*. 2013;23:128–133.
27. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56–67.
28. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available from: <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>. 2017.
29. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat*. 2001;29:1189–1232.
30. Anttila A, Nuutinen M, Van Gils M, Pekki A, Sauni R. Associations of depressive symptoms and psychosocial working conditions with sickness absences in a Finnish cohort of 11,495 employees. *Prev Med Rep*. 2024;47:102899.