

Hoitosuosituksen näytönasteen ja vahvuuden arviointi GRADE-työryhmän tapaan

Suomalaiset lääkärit ovat edelläkävijöitä näyttöön perustuvien hoitosuosituksen tuottamisessa ja käytössä. Yleislääkärin käsikirja ja tietokanta (nykyisin Lääkärin tietokannat) sai alkunsa 1980-luvun loppupuolella, ja Käypä hoito -suosituksia on laadittu vuodesta 1994. Näyttöön perustuvissa suosituksissa on olennaista kuvata suosituksen käyttäjälle, minkätasoiseen tutkimustietoon suositus perustuu. Yhtenäiset, käytökelpoiset ja läpinäkyvät menetelmät helpottavat tutkimusten laadun ja näytönasteen arviointia. Kuvaamme tässä artikkelissa näytönasteen ja suositusten vahvuuden arviointia kansainvälisen GRADE-työryhmän tapaan. Monet hoitosuosituksen tekijät eri puolilla maailmaa ovat omaksuneet tämän arviointitavan sellaisenaan tai sitä soveltaen.

Uuden tutkimustiedon hallinta on vaikeaa kiireisille klinikoille. Näyttöön perustuvissa hoitosuosituksissa pyritään kokoamaan tutkimustietoa selkeiksi, käytännön lääkäriä hyödyttäväiksi toimintaohjeiksi. Tiedon kiteyttäminen systemoituviin kirjallisuuskatsauksiin ja hoitosuosituksiin tukee uuden tiedon soveltamista käytäntöön.

Hoitosuositukset perustuvat tutkimustiedon kattavaan kokoamiseen ja kriittiseen arviointiin. Hoitosuosituksia tukeva oleellinen tieto tiivistetään näytönastekatsauksiin, joissa kuvataan hoidon tai toimintamallin taustalla oleva tieteellinen näyttö tai sen puute. Tutkimuksen asetelman ja näytönasteen huomiotta jättäminen on monesti johtanut harhaan hoidon vaikutuksia arvioitaessa. Yksi esimerkki

tästä on menopaussin jälkeinen hormonikorvaushoito, jota vuosikymmenen ajan suositeltiin vaihdevuosi-ikäisille naisille sydän- ja verisuonisairauksien ehkäisyyn (American College of Physicians 1992). Tieto vaikuttavuudesta perustui kohorttitutkimuksiin ja kumoutui myöhemmissä satunnaistetuissa hoitokokeissa (Anderson ym. 2004). Jos näyttönastetta olisi arvioitu nykyiseen tapaan kriittisesti, hoidon mahdollisia hyötyjä ja haittoja olisi ehkä punnittu tarkemmin. Suositukset olisivat olleet varovaisempia eikä lääkityksen käyttöä olisi suositeltu oireettomille naisille sydän- ja verisuonisairauksien ehkäisyyn.

Suosituksen laatimisessa oleellista on tieto tutkimustiedon näytön laadusta. Näytön laatu tarkoittaa tarkkuutta, jolla arvio tutkittavan hoidon vaikuttavuudesta pitää paikkansa. Hoitosuosituksissa tarvitaan myös arvio suosituksen vahvuudesta. Vahvuudella kuvataan sitä, missä määrin voidaan olla varmoja, että suositusta noudatettaessa saadaan aikaan enemmän hyötyä kuin haittaa. Käytännössä suosituksen vahvuutta arvioitaessa otetaan kantaa siihen, kuinka painokkaasti suosituksen laatijat ehdottavat suosituksen mukaista toimintaa yksittäistä hoitopäätöstä tehtäessä.

Näytönasteen arvioinnin lähtökohdat GRADE-työryhmän mukaan

Hoitosuosituksen tekijät ovat ratkaisseet kysymyksen näytönasteen arvioinnista eri tavoin eri maissa (Atkins ym. 2004a). Vuonna 2000 käynnistyneen GRADE-yhteistyöryhmän (The Grading of Recommendations Assessment, Development and Evaluation) (www.gradeworkinggroup.org) (Atkins ym. 2004b)

toiminta sai alkunsa pyrkimyksestä tuoda selkeyttä luokitukseen ja yhdenmukaistaa arviointia (Guyatt ym. 2008a). Tavoitteena oli myös, että hoitosuosituksen tekijät voisivat enemmän hyödyntää samaa taustatyötä. GRADE-työryhmään osallistuvia ja sen luokittelukriteerejä käyttäviä keskeisiä hoitosuosituksen laatijoita on GRADE:n sivuilla lueteltu yhteensä 45. Näistä tärkeimpiä ovat mm. Cochrane Collaboration, NICE (National Institute for Clinical Excellence), SIGN (The Scottish Intercollegiate Guidelines Network) ja CDC (Centers for Disease Control and Prevention).

GRADE-työryhmä on luonut näytönasteen arviointiin prosessin, joka pyrkii olemaan mahdollisimman yksiselitteinen, selkeä ja toistettava. Työryhmä on päätenyt näytönasteen neliportaiseen luokitteluun (TAULUKKO 1). Luokkien määritelmä on kytketty arvioon todennäköisyydestä, jolla mahdollinen uusi tieto muuttaa käsitystä hoidon vaikutuksesta tai hoidon tehon arvioinnin tarkkuudesta. Luokittelua on käytetty eniten hoitojen ja hoitotutkimusten arviointiin, mutta sitä sovelletaan myös epidemiologisten havainnoivien tutkimusten näytön arvioinnissa.

Lähtökohtana selkeä kysymys. Näytönastetta arvioitaessa tulee tutkimuskysymys muotoilla selkeästi ja yksiselitteisesti. Kliinistä hoitoa koskevilla kysymyksillä on neljä keskeistä osaa: Potilaat (Patients), Interventio, (Intervention), Vertailu (Comparator) ja Tulostuuttuja (Outcome). Näitä kutsutaan PICO-muuttujiksi. Esimerkkinä voisi toimia kysymys ”Kun haimasyöpäpotilaita (P) hoidetaan oirenmukaisesti, ovatko endoskooppisesti asennetut stentit (I) avokirurgiaa (C)

parempia oireiden lievityksessä (O), kuolleisuuden vähentämisessä (O) tai kustannuksiltaan (O).

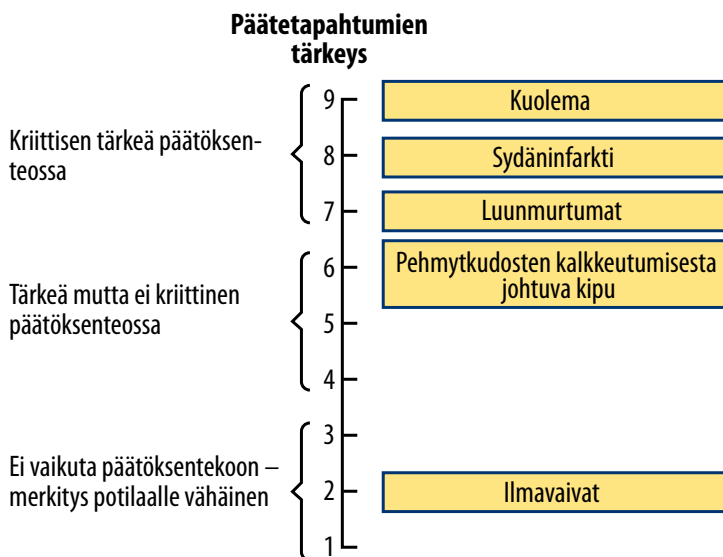
Oleelliset tulostuuttajat. GRADE-luokituksen lähtökohtana ovat potilaan kannalta oleelliset tulostuuttajat. Jo ennen kirjallisuuden läpikäymistä suositusryhmä valitsee potilaan kannalta merkittävät tulostuuttajat, joita vasten vaikuttavuutta arvioidaan. Tällaisia muuttujia ovat esimerkiksi kuoleman välttäminen, sairauden oireiden lievittyminen tai niiltä välttyminen ja elämänlaadun paraneminen (KUVA 1).

Usein potilaan kannalta äärimmäinen ja olennainen tulostuuttaja on kuoleman riski. Kuitenkin jos halutaan selvittää oireiden lievitymistä parantumattomasti sairailta potilailla, se saattaa olla potilaan lähtökohdista elinaikaa tärkeämpi tulostuuttaja. GRADE-prosessiin kuuluu sekä keskeisten tulosta mittaavien että mahdollisia haittoja kuvaavien tulostuuttajien tunnistaminen ja huomioiminen arvioinnissa.

Tutkimuksen tyyppi. Laadukkaat satunnaistetut, vertailevat hoitotutkimukset tuottavat GRADE-ryhmän mukaan vahvaa astetta A näyttöä hoidon vaikuttavuudesta. Havainnoivista tutkimuksista saatu tieto on lähtökohtaisesti astetta C. Mahdollisten haittojen kuten harvinaisten haittavaikutusten arvioinnissa satunnaistetut tutkimukset eivät ole yhtä vahvoja – potilaita on usein liian vähän, jotta erot haittataapatumissa saataisiin luotettavasti esiin. Myös hoidon tehoa arvioidessa monet tutkimuksen menetelmiin liittyvät seikat saattavat heikentää käsitystä näytönasteesta. Näytönasteen huonontaminen perustuu harhan todennäköisyyteen / olemassaoloon tutkimuksessa (KUVA 2).

TAULUKKO 1. Suomalaisissa hoitosuosituksissa käytössä oleva GRADEn mukainen näytönasteen luokittelu ja määritelmät.

Aste	Määritelmä
A = vahva	Uusi tutkimus tuskin muuttaa käsitystä vaikutuksen suuruudesta
B = kohtalainen	Uusi tutkimus saattaa muuttaa käsitystä vaikutuksen tarkkuudesta ja suuruudesta
C = heikko	Uusi tutkimus hyvin todennäköisesti muuttaa merkittävästi käsitystä vaikutuksen tarkkuudesta ja todennäköisesti myös suuruudesta
D = hyvin heikko	Arvio vaikutuksesta on hyvin epävarma



KUVA 1. Hoitotulosten hierarkia tärkeysjärjestyksessä potilaan kannalta. Esimerkkinä fosfaattipitoisuutta pienentävien lääkkeiden vaikutusten arviointi potilailla, joilla on munuaisten vajaatoimintaan liittyvä hyperfosfatemia (Guyatt ym. 2008b).

Näytön asteen laskeminen ja nostaminen

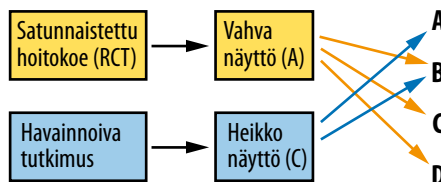
Näytönastetta laskevia tekijöitä on viisi: tutkimuksen laatu, tulosten epäyhtenäisyys, tulosten epäsuoruus, tulosten epätarkkuus sekä julkaisuharha.

Tutkimuksen laatu. Tavallisin harhaa tuottava tekijä on tutkimuksen toteutuksessa: esimerkiksi satunnaistetussa tutkimuksessa potilaiden jakoa ryhmiin ei ole salattu hoidon aloituksesta päättäviltä (allocation concealment), sokkoutus puuttuu tai on epäonnistunut, lopputulosmittarit eivät ole kliinisesti merkittäviä, seuranta on liian lyhyt tai tutkimuksesta poissääjääneiden määrä on suuri. Myös tutkimuksen lopettaminen suunniteltua aikaisemmin hoidon hyödyn osoittauduttua väliarviossa merkitseväksi voi johtaa harhaan.

Tulosten epäyhtenäisyys. Näytönastetta laskee tulosten epäyhtenäisyys, esimerkiksi se, että systemoidussa katsauksessa eri osatutkimukset tuottavat hyvin erilaisia tuloksia. Tämä voi johtua sattumasta tai toistaiseksi tuntemattomasta syystä tai esimerkiksi siitä, että tutkittavat potilasryhmät, tehdyt interventiot tai

tulosmuuttajat poikkesivat eri tutkimuksissa toisistaan. Joissakin tutkimuksissa potilaat saattoivat olla sairaampia kuin toisissa, lääkkeiden annokset ja hoidon kesto vaihtelivat tai paranemisen seurannassa käytettiin erilaisia mittareita.

Tulosten epäsuoruus. Myös tulosten tai vaikutusten epäsuoruus ja yleistettävyyden ongelmat voivat heikentää arviota näytön-



KUVA 2. Näytönasteen arviointi sekä laskeminen ja nostaminen hoitosuosituksissa. Satunnaistetut hoitokokeet tuottavat yleensä astetta A-näyttöä hoitosuosituksia varten. Harhan mahdollisuus saattaa kuitenkin heikentää näytönastetta (keltaiset nuolet). Havainnoivien tutkimusten näytönaste on GRADEn mukaan yleensä heikkoa astetta C. Sitä voidaan nostaa, jos tulos on vakuuttava, toistettu ja harhan mahdollisuus pieni (siniset nuolet).

asteesta, esimerkiksi se, että tutkimuksen kohderyhmä poikkeaa hoitosuosituksen kohderyhmästä. Toisinaan kahta lääkettä on verrattu erikseen lumehoitoon mutta ei keskenään. Jos toinen lääkkeistä on osoittautunut merkittävästi lumetta paremmaksi ja toinen vain hieman paremmaksi, voitaisiin ajatella, että selvästi paremmaksi osoittautunut lääke on toista parempi. Näyttö on kuitenkin epäsuoraa suoran vertailun puuttuessa kahden lääkkeen väliltä, jolloin ei ole kunnollista näyttöä hoitojen välisestä tehoerosta.

Tulosten epätarkkuus. Tulokset voivat olla epätarkkoja pienen potilas- tai tapahtumamäärän johdosta. Tämä näkyy tulosten suurina luottamusväleinä.

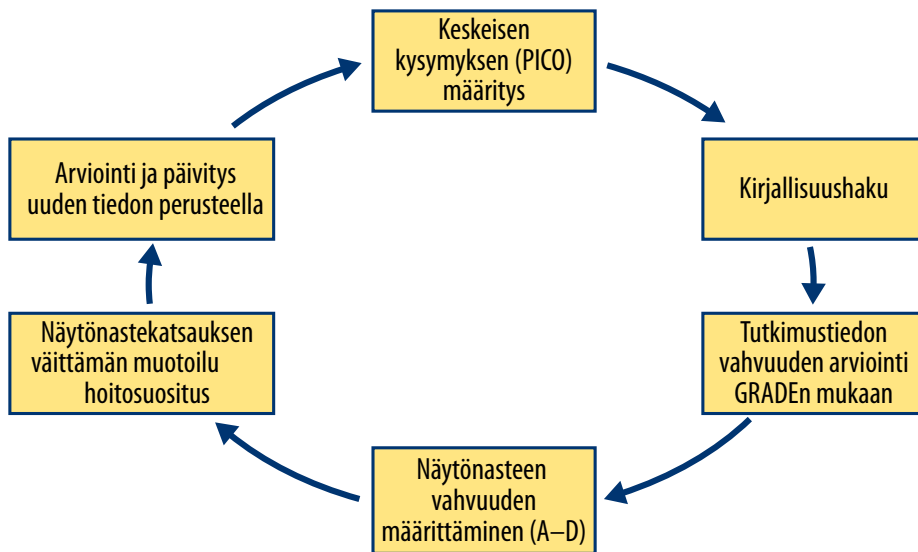
Julkaisuharha. Tutkimukset, joissa ei saada tulosta jonkin hoidon paremmuudesta vertailukohteeseen nähden, jäävät useammin julkaisematta kuin sellaiset, joista on saatu selvä tulos. Tämä vääristää julkaisujen perusteella tehtyä arviota hoidon tehosta niin, että hoidon vaikutus näyttää suuremmalta kuin se todellisuudessa on. Tätä harhaa on pyritty ehkäisemään vaatimalla kaikkien tutkimusten rekis-

teröimistä (Controlled Clinical Trials 2010) jo niiden suunnitteluvaiheessa (DeAngelis ym. 2004), jolloin voidaan paremmin valvoa, että tulokset raportoidaan niiden valmistuttua.

Näytön asteen nostaminen. Havainnoivien tutkimusten näyttö on GRADE-luokittelun mukaan lähtökohtaisesti heikkoa (aste C) (KUVA 2). Kuitenkin näytönastetta voidaan nostaa, jos tulos on hoito- ja verrokkiryhmissä toistetusti huomattavan erilainen. Esimerkiksi tupakan ja keuhkosyövän yhteydestä on olemassa ainoastaan havainnoivia tutkimuksia, mutta yhteys on toistetusti osoitettu huomattavaksi ja epäilemättä astetta A. Havainnoivan tutkimuksen näytönastetta voidaan nostaa, jos on näyttöä annos-vastesuhteesta tai kaikki ajateltavissa olevat harhat heikentäisivät tulosta.

Näytönasteen arviointi suomalaisissa hoitosuosituksissa

Hoitosuositusten laatiminen aloitetaan keskeisten vastattavien kysymysten määrittämisellä ja kirjallisuushaulla (KUVA 3). Käypä hoito-suosituksessa keskeisen ongelman määrittämi-



KUVA 3. Hoitosuosituksen laatimisprosessi.

sen jälkeen tehdään laaja kirjallisuushaku, jonka tulokset jaetaan vastuualueittain ryhmän jäsenille koottaviksi näytönastekatsauksiin. Ennen näytönastekatsauksen laatimista määritetään PICO-periaatteita noudattaen kysymys, johon kirjallisuuden perusteella halutaan löytää vastaus, ja tehdään tarvittaessa uusia kirjallisuushakua. Lääkärin tietokannoissa päätyönä on jo olemassa olevien suositusten päivittäminen uudella tutkimustiedolla. Tärkeimpiä lähteitä tässä työssä ovat Cochrane-katsaukset sekä muut systemoidut kirjallisuuskatsaukset (Varonen ym. 2005). Kun kirjallisuus on koottuna, näytönastekatsauksen kirjoittaja muotoilee keskeisen tutkimuksessa tuotetun vastauksen.

Näytönastekatsauksen alussa oleva väittämä on katsauksen kiteytys. Siinä pyritään kuvaamaan yksi tai useampia potilaan kannalta keskeisiä tulomuuttujia PICO-rakenteen mukaisesti. Väittämän sanamuoto heijastaa näytönastetta:

A = on / lisää / vähentää / ei ole vaikutusta, B = ilmeisesti on / ilmeisesti lisää / ilmeisesti vähentää / ilmeisesti ei ole vaikutusta, C = lie-

nee / saattaa lisätä / saattaa vähentää / ei liene vaikutusta, D = saattaa lisätä / vähentää mutta luotettava näyttö puuttuu, ei ole näyttöä.

Vahvaa näyttöä voi olla myös negatiivisesta vaikutuksesta, eli esimerkiksi on asteen A näyttöä siitä, että antibiooteista ei ole hyötyä vilustumisen hoidossa.

Myös kotimaisten hoitosuosittelujen laatijoilla on joitakin keskinäisiä eroja näytönasteen arvioinnissa, joskin niiden käytännön merkitys suositusten käyttäjän kannalta lienee vähäinen. Käypä hoito -suosituksissa arvioidaan näytön sovellettavuutta erityisesti suomalaisen väestöön, kun taas Lääkärin tietokantojen näytönastekatsaukset on tarkoitettu myös kansainväliseen käyttöön, ja siten niissä korostuu vähemmän suomalainen näkökulma tai Suomelle ominaiset aihevalinnat. Käypä hoito -suosituksessa on asteen A näyttöön vaadittu yleensä ainakin kaksi erillistä tutkimusta, joiden tulokset ovat yhdenmukaiset, mutta Lääkärin tietokannoissa on katsottu yhdenkin suuren, laadukkaan tutkimuksen todistusvoiman riittävän.

TAULUKKO 2. Suosituksen vahvuuden arviointiin vaikuttavia tekijöitä.

Tekijä	Esimerkki
Epävarmuus arviossa mahdollisista hyödyistä, riskeistä, epämuksuista ja kustannuksista	Inhaloitavien steroidien hyöty astman hoidossa on osoitettu selvästi, kun taas teofylliinihoitoon kroonisen bronkiitin pahenemisvaiheessa liittyy hyötyyn nähden paljon riskejä
Hoidolla estettävän tapahtuman kliininen merkitys	On merkittävämpää estää syvän laskimotrombin hoidolla tappava keuhkoembolia kuin posttromboottinen oireyhtymä
Hoidon vaikutuksen suuruus	Klopidogreeli antaa asetyylisalisyylihappoon verrattuna pienemmän tehon aivohalvauksen ehkäisyssä TIA-potilaalla kuin varfariini luumeeseen verrattuna eteisvärinäpotilaalla
Arvio hoidon vaikutuksen tarkkuudesta	Asetyylisalisyylihapon teholla on lumelääkkeeseen verrattuna aivohalvauksen estossa suurempi luottamusväli eteisvärinäpotilailla kuin TIA-potilailla
Hoitoon liittyvät riskit	Asetyylisalisyylihapo ja klopidogreeli aiheuttavat akuutin koronaarisyndrooman yhteydessä suuremman vuotoriskin kuin asetyylisalisyylihapo yksin
Hoidon aiheuttama haitta	Varfariinihoidon toteuttaminen on työläämpää kuin asetyylisalisyylihapollääkitys tiheän seurannan tarpeen vuoksi
Päätetapahtuman riski	Kirurgisten potilaiden tromboembolisten komplikaatioiden riski vaihtelee suuresti leikkauksen ja potilaan mukaan
Hinta	Klopidogreeli on huomattavasti kalliimpi lääke kuin asetyylisalisyylihapo
Potilaiden arvot	Nuoret ihmiset saattavat arvostaa elämän pidentämistä kärsimyksistä huolimatta enemmän kuin vanhat

Suosituksen vahvuuden arvioiminen

GRADE-työryhmä on pitänyt tärkeänä arvioida paitsi näytönastetta myös suositusten vahvuutta. Se suosittelee, että näytönasteen lisäksi hoitosuositusryhmät ottavat kantaa myös suositusten vahvuuteen ja ehdottaa tähän kahta luokkaa, vahvoja ja heikkoja suosituksia. Vahvan suosituksen tärkein ominaisuus on, että suositusryhmä on vakuuttunut suosituksen noudattamisen tuottavan hyötyjä, jotka ylittävät mahdolliset haitat. Heikon suosituksen tärkein ominaisuus on, että suositusryhmä arvioi suosituksen noudattamiseen liittyvien hyötyjen ylittävän haitat mutta hyödyn olevan vähäinen tai vielä epävarma. Tekijöitä, jotka vaikuttavat suosituksen vahvuuden arviointiin, on esitetty **TAULUKOSSA 2**.

GRADE-työryhmä esittää eri osapuolten arvojen ja näkökulmien huomioon ottamista suositusten vahvuutta arvioitaessa. Tällöin huomioidaan niin lääkärin, potilaiden kuin päättäjienkin arviot hoidon hyödyistä ja haitoista.

Vahvalle suositukselle tunnusomaisia arvioita ovat: 1) Lähes kaikki potilaat valitsisivat kyseisen hoidon ja vain hyvin harvat kieltäytyisivät. Hoidosta pidättäytyminen vaatii aina keskustelua potilaan kanssa. 2) Kliinikoiden mielestä useimpien potilaiden tulisi saada kyseinen hoito. 3) Päättäjien näkökulmasta hoidon tulisi kuulua hoitosuosituksien piiriin.

Heikolle suositukselle tunnusomaisia arvioita ovat: 1) Useimmat potilaat valitsisivat kyseisen hoidon, mutta monet myös kieltäytyisivät siitä. 2) Kliinikoiden näkökulmasta heikon suosituksen toteuttaminen vaatii yksilöllistä harkintaa, ja hoidosta tulisi päättää potilaan kanssa käytävän keskustelun perusteella hänen arvojensa ja toiveittensa mukaisesti. 3) Päättäjien näkökulmasta suosituksen omaksuminen vaatii keskustelua ja kaikkien asianosaisten näkökulman huomioon ottamista konsensuksen saavuttamiseksi.

ESIMERKKI 1. On kohtalaista näyttöä (B-taso), että rilutsoli lisää ALS-potilaan elinaikaa keskimäärin noin 2–3 kuukaudella. Tämä antaa kuitenkin aihetta vain heikkoon suositukseen, ja viime kädessä lääkkeen käytöstä tai siitä pidättäytymisestä tulee päättää potilaan tahtoa kuunnellen.

ESIMERKKI 2. Näyttö ylipainehappihoidon hyödyistä on erittäin niukkaa useimmissa indikaatioissa, joihin tätä hoitoa on kokeiltu. Myös vakavan sukeltajataudin hoidossa näyttö ylipainehappihoidon hyödyistä on astetta C. Sukeltajatauti on kuitenkin potentiaalisesti hengenvaarallinen tila, eikä ylipainehappihoidolle ole vaihtoehtoja. Näin ollen kyseisessä tilanteessa ylipainehappi on kuitenkin käytössä siellä, missä resursseja hoidon toteuttamiseen on olemassa, ja sen käytöstä voidaan antaa vahva suositus puutteellisesta näytöstä huolimatta.

GRADE-työryhmä ehdottaa, että hoitosuositusryhmät ohjeistetaan ottamaan kantaa suosituksen vahvuuteen ja tuottamaan suosituksia käytäntöjen puolesta tai niitä vastaan, joten ”nollavaihtoehto” ei periaatteessa ole. Käytännössä kuitenkin on tilanteita, joissa näyttö on heikkoa ja ryhmä voi suositella menetelmän käyttöä vain tutkimuksen yhteydessä.

Hoitosuosituksien näytönasteen ja vahvuuden määrittämiseen liittyy eroja eri kulttuurien välillä. Esimerkiksi yhdysvaltalaisissa hoitosuosituksissa pidetään tärkeänä, että hoitosuositusryhmä ottaa selkeän kannan eri hoitojen puolesta tai vastaan ja määrittää suosituksen vahvuuden.

Suomalaisissa hoitosuosituksissa on kuvattu näytönastetta ja jätetty tilaa potilaskohtaiselle harkinnalle. Suosituksen vahvuuden määrittäminen voi kuitenkin auttaa nostamaan esille suosituksen pääkohtia, ja sillä saattaa olla merkitystä hoitosuosituksen käytäntöön viemisen kannalta. Eri hoitosuosituksien tekijät voivat käyttää työn perustana samoja systemoituja katsauksia ja päätyä samaan näytönasteeseen. Koska suosituksen vahvuus on puhdas arvo päätös, se voi olla erilainen eri tekijöiden arvioimana. GRADE-työryhmän mukaan tämä ei vähennä suosituksen vahvuuden määrittämisen merkitystä. Tärkeintä on kuitenkin kertoa avoimesti, mihin suosituksen antaminen perustuu. Tämä toteutuu parhaiten, kun sekä näytönaste että suosituksen vahvuus on ilmoitettu ja on myös kerrottu perusteet sille, miksi näytönastetta on laskettu tai nostettu, sekä perusteet, joiden vuoksi yksittäisen suosituksen vahvuus on määritetty.

Pohdinta

Parhaimmillaan näytönasteen arviointi auttaa hoitosuosituksen tekijöitä pääsemään perusteltuihin päätelmiin suosituksia laadittaessa. Näytönasteen arvioinnin pitää olla systemaattinen tai toistettavissa oleva prosessi, johon GRADE-työryhmän menetelmä antaa avaimia.

Hoidon hyvä laatu, alueellinen tasa-arvoisuus ja potilasturvallisuus saavat yhä enemmän painoa terveydenhuollossa. Hoitosuosituksia ei voida aina sellaisenaan käyttää hoidon laadun kriteereinä, koska hoidon soveltamisessa käytännön potilastyöhön on otettava huomioon potilaan arvot, joiden perusteella tehtäviä päätöksiä hoitava lääkäri ohjaa ja auttaa.

GRADE-työryhmän periaatteet ovat selkeyttäneet hoitosuosituksen laadintaa. Työryhmä on onnistunut yhdistämään tutkimusten laadulliseen arviointiin satunnaistetut tutkimusasetelmat ja seurantatutkimuksen tulokset. Sen työn tuloksena useimmat hoitosuosituksen tekijät ovat seuranneet tutkimusten laadullisessa arvioinnissa työryhmän luomia kriteerejä näytönasteen arvioinnissa.

Suosituksen vahvuuden arviointiin työryhmä on yhdistänyt hoidon hyödyt ja haitat, tutkimusnäytön asteen, osapuolten arvot ja näkökulmat sekä hoidon kustannukset. Suosituksen vahvuuden arviointiin ei tule olemaan yhtä selviä pelisääntöjä kuin näytönasteen osalta, koska kyseessä on arvopäätös, joka vaihtelee yhteiskunnan eettisten arvojen, taloustilanteen, terveydenhoidon resurssien yms. mukaan. Nämä ovat tekijöitä, jotka eivät ole suoraan arvioitavissa näytönasteen perusteella ja voivat aiheuttaa näkemyseroja terveydenhuollon hallinnon, hoitavan lääkärin ja

potilaiden kesken. Toivottavaa on, että yhteiskunnan kannalta keskeisimpien suositusten vahvuuksista voitaisiin käydä avointa keskustelua kaikkien toimijoiden keskuudessa.

Vahvoilla suosituksilla saattaa olla uusia interventiotutkimuksia ehkäisevä vaikutus. Menopausin aikainen hormonihoito on tästä esimerkkinä. Tämän vuoksi suositusten vahvuus ei saa olla hoidon laadun kriteeri eikä suositusta tule käyttää estämään uusien interventiotutkimusten käynnistämistä.

GRADE-työryhmän työ kokoaa tutkijat ja kansallisten hoitosuosituksen laatijat parantamaan yhdessä suositusten osuvuutta ja kehittämään laadinnan menetelmiä. Tähän mennessä tehty työ on parantanut päätelmien selkeyttä ja prosessin läpinäkyvyyttä. ■

JUKKAPEKKA JOUSIMAA, LT, päätoimittaja
Kustannus Oy Duodecim
PL 874, 00101 Helsinki

HELENA LIIRA, dosentti, ylilääkäri
HUS, Yleislääketieteen yksikkö ja Kirkkonummen terveyskeskus

JUHA LIIRA, dosentti, ylilääkäri
Turun yliopisto ja Työterveyslaitos

JORMA KOMULAINEN, LT, ylilääkäri
Terveyden ja hyvinvoinnin laitos

SIDONNAISUUDET

JUKKAPEKKA JOUSIMAA: Lääkärin tietokantojen päätoimittaja.

HELENA LIIRA: Osallistuu hoitosuositustyöhön Duodecimissa.

JUHA LIIRA: Osallistuu hoitosuositustyöhön Työterveyslaitoksella.

JORMA KOMULAINEN: Käypä hoito -kehittämispäällikkö.

Summary

Judgment of the level of evidence and strength of recommendations in healthcare according to the GRADE working group

Finnish doctors are pioneers in the production and utilization of evidence-based recommendations in healthcare. Doctor's databases were established by the end of the 1980s and Current Care guidelines have been devised since 1994. In evidence-based recommendations it is essential to provide the user with a description of the quality of research data used as the basis for the recommendation. Consistent, applicable and transparent methods are helpful in the judgment of the quality of studies and strength of evidence. In this article we describe the judgment of the level of evidence and the strength of recommendations according to the international GRADE working group.

KIRJALLISUUTTA

- American College of Physicians. Guidelines for counseling postmenopausal women about preventive hormone therapy. *Ann Intern Med* 1992;117:1038–41.
- Anderson GL, Limacher M, Assaf AR, ym. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA* 2004; 291:1701–12.
- Atkins D, Eccles M, Flottorp S, ym. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004(a);4:38.
- Atkins D, Best D, Briss PA, ym. Grading quality of evidence and strength of recommendations. *BMJ* 2004(b);328:1490.
- Controlled Clinical Trials. 2010. www.controlled-trials.com/isrctn. [Siteerattu 15.1.2010].
- De Angelis C, Drazen JM, Frizelle FA, ym. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *Lancet* 2004;364:911–2.
- GRADE working group. www.gradeworkinggroup.org. [Siteerattu 18.2.2010].
- Guyatt GH, Oxman AD, Vist G, ym. for the GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008(a);336:924–6.
- Guyatt GH, Oxman AD, Kunz R, ym. for the GRADE Working Group. Rating quality of evidence and strength of recommendations. What is 'quality of evidence' and why is it important to clinicians? *BMJ* 2008(b);336:995–8.
- Guyatt GH, Oxman AD, Kunz R, ym. for the GRADE Working Group. Rating quality of evidence and strength of recommendations. Going from evidence to recommendations. *BMJ* 2008(c);336:1049–51.
- Varonen H, Jousimaa J, Helin-Salmivaara A, Kunnamo I. Electronic primary care guidelines with links to Cochrane reviews – EBM Guidelines. *Fam Pract* 2005;22(4):465–9.