

Vesa Kautto

Aikakauslehtiartikkeleiden arviointi

Weller, Ann C. Editorial peer review: Its strengths and weaknesses. Medford, NJ: Information Today 2001. 342 s. ASIST Monograph Series. \$ 44.50. ISBN 1-57387-100-1

Tieteellisten aikakauslehtien historian katsotaan alkaneen 1600-luvulla, kun Philosophical Transactions of the Royal Society ryhtyi ilmestymään. Yhtä vanha on julkaisujen asiantuntija-arviointi: seuran sihteeri painatti käsikirjoitukset vasta, kun jotkut seuran jäsenet olivat ne tarkistaneet. Kuitenkin on todettu, että yleiseksi vertaisarviointi – peer review – tuli lehdissä vasta toisen maailmansodan jälkeen. Sitä ennen toimittaja tai toimituskunnan jäsenet päättivät hyväksymisestä tai hylkäämisestä yleensä omin avuin. Yleistyessäänkään vertaisarvioinnista ei muodostunut standardisoitua käytäntöä. Niinpä ei läheskään kaikissa lehdissä ole esimerkiksi kirjallisia ohjeita käytettävistä kriteereistä.

Normaalisti arviointiprosessi on seuraava. Artikkelin käsikirjoituksen saavuttua lehden toimitukseen se kirjataan ja valitaan kaksi tai kolme arvioijaa. Näitä pyydetään tarkastamaan käsikirjoitus ja ehdottamaan sen hyväksymistä, hyväksymistä korjattuna tai hylkäämistä. Lausuntojen saavuttua toimittaja/toimittajat päättävät, noudattavatko he arvioijien suositusta. Jos arvioijat ovat erimielisiä, toimittaja voi lähettää käsikirjoituksen toisille arvioijille tai tehdä itse päätöksen käsikirjoituksen hyväksymisestä tai hylkäämisestä.

Menettelyissä esiintyy vaihtelevia käytäntöjä. Toimittaja voi esimerkiksi lähettää arvioijille lomakkeen täytettäväksi tai laajat arviointiohjeet tai yksinkertaisesti pyytää kommentteja. Arvioijien kommentit voidaan antaa kirjoittajille sellaisinaan tai toimittaja voi tehdä niistä erityyppisiä yhteenvetoja.

Kymmenisen vuotta sitten arvioitiin, että maailmassa tuotetaan 6 000 – 7000 tieteellistä artikkelia päivässä eli kolmatta miljoonaa vuodessa. Kun useimmat julkaistut ja hylätyt artikkelit ovat käyneet läpi asiantuntijatarkastuksen, arviointiin käytetty työmäärä on valtava.

On siis erittäin aiheellista arvioida arviointia, sen menettelyjä, kannattavuutta ja luotettavuutta, takaako tarkastus laadun, julkais-taanko parhaat ja hylätäänkö huonot ja pääsevätkö radikaalit uudet ajatukset läpi.

Illinoisin yliopiston terveystieteiden apulais-professori ja kirjaston varajohtaja Ann C. Weller on koonnut vertaisarviointiin kohdistuneiden tutkimusten tuloksia laajaan katsaukseen (Weller, Ann C. Editorial peer review: Its strengths and weaknesses. Medford, NJ: Information Today 2001. Weller kävi läpi 1500 tutkimusta vuosilta 1945–1997. Näistä 300 osoittautui ei-relevantiksi. Ennen vuotta 1945 ei hänen mukaansa tutkimuksia ole tehty.

Weller ei ole analysoinut vertaisarviointitutkimusten aloittaista sijoittumista. Voitaneen olettaa, että johdantoluvun noin kolmensadan lähteen jakauma on tyypillinen: noin 62 % oli ilmestynyt lääketieteellisissä aikakauslehdissä, vain 8 % informaatiotutkimuksen lehdissä ja kokousjulkaisuissa ja 30 % muiden alojen lehdissä.

Weller tarkastelee tutkimusten valossa tieteellisten aikakauslehtien toimituksellista työskentelyä varsin laajalti. Selväksi käy, että lehden ilmoittautuminen asiantuntija-arviointia käyttäväksi lehdeksi (refereed journal, peer-reviewed journal) ei ole sisällöltään tarkan termin käyttämistä, lehti sijaitsee jossakin kohtaa vertaisarvioinnin jatkumolla.

Arvioinnin tekniset menettelyt vaihtelevat eri lehdissä esimerkiksi nimettömyyden osalta. Joissakin lehdissä arvioija ei saa tietoonsa artikkelin kirjoittajaa / kirjoittajia eivätkä kirjoittajat arvioijia. Joissakin julkaisuissa artikkelin arvioijille ilmoitetaan kirjoittajat, mutta arvioijien nimiä ei kerrota tekijöille. Toisinaan käytetään avointa menettelyä, jossa sekä kirjoittajat että arvioijat tietävät toisensa. Wellerin läpikäymän 15

selvityksen mukaan arvioijien nimettömyys oli vallalla 60-prosenttisesti ja kirjoittajien 34-prosenttisesti.

Arvioijien nimien paljastamista kirjoittajille on keskustelussa vaadittu ja vastustettu. Paljastamista on puollettu esimerkiksi sillä, että se lyhentäisi arviointiaikojä. Vastustusta on perusteltu muun muassa siten, että nuoret tutkijat eivät haluaisi silloin toimia arvioijina, koska pelkäisivät kosto-toimenpiteitä. Vertaisarviointien rinnastus nimellä varustettuihin kirja-arviointeihin ei tietenkään ole relevantti: kirjathan ovat jo ilmestyneet ennen arvostelun laatimista.

Useissa tutkimuksissa on selvitetty artikkeleiden hylkäämisen syitä ja hylkäämisprosentteja. Hylkäämisprosentit vaihtelevat aloittain (esimerkiksi fysiikka 19 %, historia 90 %) ja saman alan lehtien kesken. Hylkäämisperusteluina on tuotu esiin lukuisia syitä: ei sopiva lehden kannalta, ei uutta tietoa, ei tärkeästä kysymyksestä, teoreettiset ja käsitteelliset heikkoudet, metodiset ongelmat, havaintojen ja tulosten löysä tulkinta, puutteellinen kirjallisuuskatsaus, tilastollisen käsittelyn virheet jne.

Tilastollisen käsittelyn laatua on tutkittu erityisesti lääketieteen alueella. Eräissä tutkimuksissa tilastoasiantuntijat ovat tarkastelleet jo hyväksytyjen käsikirjoitusten tasoa omalta kannaltaan. British Medical Journalin 200 hyväksytystä käsikirjoituksesta tilastomiehet olisivat hylänneet 37 % ja Lancetin 191 käsikirjoituksesta 14 %, minkä lisäksi 32 prosenttiin Lancetin käsikirjoituksista oli tehtävä korjauksia.

Arvioijien yksimielisyyden astetta on tutkittu vähän. Asetelmaltaan mielenkiintoinen oli Ernstin ym. (1993) selvitys, jossa asiantuntijat arvioivat samaa lääketieteellistä käsikirjoitusta. Teksti lähetettiin 45:lle, 31 vastasi. Noin 2/3 piti käsikirjoitusta hyvänä, 20 % erinomaisena, 12 katsoi käsikirjoituksessa olevan tilastokäsittelyn ei-hyväksyttäväksi ja kymmenisen prosenttia ehdotti käsikirjoituksen hylkäämistä. Asiantuntijat oli valittu käsikirjoituksen alaa lähellä olevista lehdistä.

Weller analysoi yli 30 lehteen ja lisäksi American Psychological Associationin lehtiryhmään kohdistuvia tutkimuksia arviointien yhtenäisyydestä. Weller katsoo, että arvioijien keskinäinen yksimielisyys ei ollut suurta. Arvioijat ovat yksi-mielisempiä, kun on kyseessä käsikirjoituksen hylkääminen kuin sen hyväksyminen. Tuntuu siltä, että arvioijat ovat varsin samaa mieltä

huomattavan hyvistä ja huonoista käsikirjoituksista, mielipiteiden vaihtelu kohdistuu välillä oleviin. Luonnontieteissä on ilmeisesti helpompi saavuttaa yksimielisyys kuin luonteeltaan diffuusimmissa aineissa.

Joittenkin toimittajien mielestä ristiriitaiset arviot eivät ole haitaksi, jos ne johtuvat siitä, että arvioijat kiinnittävät huomiota eri asioihin, lähestyivät eri näkökulmista.

Weller sanoo, että merkittäväntä ei ole arvioijien erimielisyys vaan erimielisyyden syyt. Jos taustalla ovat ideologiset näkökulmaerot (esimerkiksi teoreetikko vs. käytännön edustaja), ei ole samalla tavalla oikeaa ja väärää ratkaisua kuin asiakysymyksissä.

Varsin suurta osaa hyljätystä käsikirjoituksista tarjotaan muuhun lehteen, ehkä muokattuna arvioijien kommenttien mukaan. Esim. Rotton (1995) sai tulokseksi edelleen tarjoamisesta 83,3 % ja Yankauer (1985) 61,5 %. Kuinka paljon muualle tarjotuista hyväksytään, ei ole juurikaan tutkittu.

Ensimmäiset elektroniset lehdet syntyivät parikymmentä vuotta sitten. Artikkelien arvioinnissa elektroniset lehdet ovat synnyttäneet erilaisia kokeiluja ja ehdotuksia. Alallaan erittäin keskeinen on American Physical Societyn aloitteesta syntynyt suurenergiatutkimuksen elektroninen arkisto Los Alamosin National Laboratoryn yhteydessä (1991–). Vuonna 1996 Los Alamosin arkistossa oli aineistoa 17:lta eri fysiikan erikoisalalta ja se oli käsitelty 13 000 lähetettyä tekstiä vuosittain. Varsinaista asiantuntija-arviointia ei ole, mutta toimituksella on oikeus hylätä tarjottuja raportteja. Haastattelemani fysiikan professorin mukaan aineisto on korkeatasoista, roskaa ei joko lähetetä tai se karsitaan. Arkisto on perusluonteelta pre print-tietokanta, siihen lähetettyjä tekstejä voidaan tarjota fysiikan lehtiin, ja jos ne julkaistaan, tiedot julkaisemisesta liitetään arkistossa oleviin tietoihin. Menettely on siis aivan toinen kuin lääketieteessä, jonka lehdet katsoisivat tällä tavalla julkistetut artikkelit jo julkaistuiksi eivätkä ottaisi niitä julkaistavaksi.

Elektronisten lehtien tarjoamiin mahdollisuuksiin liittyvät ehdotukset sisältävät esimerkiksi malleja, joissa alkuperäiskäsikirjoitukset ja niitä koskevat arviot olisivat näkyvillä tietokannoissa ainakin sitten kun artikkeli on hyväksytty. Myös myöhemmät kommentit asetettaisiin näkyviin.

Yhteenvedossa Weller kannattaa asiantuntija-arviointia. Järjestelmässä on puutteensa, mutta se on välttämätön. Perusteluna hän viittaa mm.

Lancetin “fast-track” -kokeiluun (1997). Toimitajat lupasivat tärkeille tutkimuksille nopean käsittelyn. McNamee (1998) kertoo, että 110 kirjoittajaa pyysi pikakäsittelyä. 18 torjuttiin suoralta kädeltä. Loput 92 käsikirjoitusta tutkittiin ja näistä vain 14 hyväksyttiin asiantuntija-arviointiin ja näistä taas 7 pikajulkaisemiseen. Kirjoittajien on vaikea arvioida aikaansaannostensa merkittävyyttä objektiivisesti.

Informaatioalan keskustelussa on tähdennetty laadun arvioinnin tärkeyttä valittaessa aineistoa verkkojulkaisujen kirjavasta joukosta. Wellerin kirja osoittaa, ettei perinteisten tieteellisten aikakauslehtienkään laadunvarmistus aukotonta ole. Kemian nobelisti John Cornworth kertoo asennoituneensa 16-vuotiaasta saakka hyvin epäluuloisesti julkaistuihin tuloksiin ja teorioihin. Kriittisyys on aiheellista.