

# Nettikyselyt ovat nykyaikaa

Seppo Laaksonen

FT

Nettikysely tarkoittaa kyselyä, tiedustelua, haastattelua tai yleisemmin surveytä, jossa vastaaja käyttää apuna internetiä, verkkoa, älypuhelinia tai muuta modernia sähköistä välinettä. Englanninkieliseksi termiksi on vakiintunut *web survey*. Tätä termiä käytetään myös Bethlehemin ja Biffignandin (2012) aihetta käsittelevässä tuoreessa kirjassa.

Nettikyselyt ovat nopeaa vauhtia yleisty-mässä. Oikein käytettynä niistä voi olla suurta apua otospohjaisten tilastojen laatimisessa. Nettikysely on vaihtoehtoisia menetelmiä, kuten puhelin-, posti- ja käyntikyselyjä, halvempi ja nopeampi (ks. menetelmistä laajemmin, Laaksonen 2013). Siksi nettikyselyn käyttö todennäköisesti laajenee varsin nopeasti. Tämä koskee myös taloudellisia tilastoja ja indikaattoreita.

Nettikyselyjen käyttöön liittyy myös monia ongelmia, joista on hyvä olla tietoinen. Markkinatutkimuksessa yleistyneet nettipaneelit perustuvat ei-todennäköisyyspohjaisiin aineistoihin, mistä syystä tulosten tilastollista luotettavuutta on vaikea arvioida. Pahimmillaan netti-

kyselyjä käytetään holtittomasti. Ongelmana ei ole kysely sinänsä, vaan niiden holtiton käyttö.

Tässä artikkelissa esitellään nettikyselyjen käyttöä ja niihin liittyviä ongelmia. Artikkelin johtopäätös on se, että nettikyselyt ovat nykyaikaa, mutta hyvät käytännöt ovat vielä hakuksessa.

## Nettikyselyn ominaisuuksia

Parhaimmillaan nettikyselytiedon laatu on yhtä hyvä kuin postikyselyn. Netti on kuitenkin postikyselyä halvempi ja nopeampi sekä yleensä myös laadukkaampi. Molempien kyselyjen piire on se, että vastaaja toimii yksin eikä kukaan pääse vaikuttamaan vastauksiin, ainakaan periaatteessa. Toki on mahdollista, että vastauksen voi antaa väärä henkilö tai joku voi vaikuttaa vastauksiin, mutta tämä ongelma on onneksi vähäinen. Muutamia sellaisia kriminalisoitavissa olevia tapauksia esiintyy sekä posti- että nettikyselyissä. Puhelinhaastattelussakin tämä voi olla mahdollista, mutta ei suomalaisessa käyntihaastattelussa.

Nettikyselyssä käytetään itsetäytettävää lomaketta kuten postitiedustelussa. Sellainen on hyödyllinen myös käyntihaastattelussa sellaisen kysymysten osalta, jotka koetaan herkiksi. Nykyään tämä tapahtuu kannettavalla tietokoneella, jonka haastattelija antaa tarvittaessa vastaajan käyttöön eikä katsele vastauksia. Aikaisemmin käytettiin puoluekannatusta kysytessä uurnaa, johon vastaaja tiputti vastauslipun. Nykyään kannatusmittaukset toteutetaan puhelimitse, jolloin anonyymiutta ei ole. Tämä tieto jääkin tyhjäksi noin 40 prosentilla vastaajista.

Joissakin kysymysaiheissa on merkitystä sillä, että haastattelija on läsnä ja tietää siis vastaukset merkitsemällä ne talteen. On havaittu, että osa haastateltavista antaa yleisesti hyväksyttävämpiä vastauksia haastattelijalle kuin vastatessaan yksin.

Esimerkiksi Eurostatin rahoittamassa suomalaisten turvallisuustutkimuksessa oli mahdollisuus verrata kolmella eri keruutavalla, netti-, käynti- ja puhelinkyselyllä, saatuja tuloksia (Aromaa ym. 2010). Etukäteen tiedettiin, että uhriksi joutumista koskevat kysymykset (vaikkapa puolison tai entisen puolison aiheuttama väkivalta) arvioitiin aroiksi aiheiksi. Siksi käyntihaastattelussa niihin vastattiin anonyymisti, periaatteessa samalla tavalla kuin nettikyselyssä. Puhelintiedustelussa samaa anonyymiutta ei voinut olla. Tutkimus näyttääkin, että puhelintiedustelussa uhriksi joutuminen oli vähäisempää kuin muissa kahdessa, joissa tulokset olivat herkkien kysymysten osalta melko samoja.

Vastaavanlaisia eroja on havaittu muissakin tutkimuksissa. Taloustieteitä lähellä oleva kysely on kuluttajabarometri. Siinä Suomessa on kaksi kertaa (vuosina 2011 ja 2012) pilotoitu nettikyselyä ja verrattu sen antamia tuloksia

puhelinhaastattelulla saatuihin tuloksiin. Vuoden 2011 pilotoinnissa kokeiltiin yhdistekyselyä (*mixed mode*) siten, että otokseen tulleita houkuteltiin ensiksi vastaamaan netissä, mutta viikon kuluttua puhelinhaastattelijat alkoivat soittaa niille, jotka eivät olleet heti vastanneet nettikyselyyn (Laaksonen 2011). Tämä jälkimmäinen vastasi perinteistä kuluttajabarometrin kyselyä, joka tulosten vertailemiseksi toteutettiin samaan aikaan myös eri otokselle.

Kuluttajabarometrin pääindikaattori koostuu neljänlaisista kysymyksistä, jotka kaikki koskevat ihmisten suhtautumista tulevaisuuteen. Yhdessä pyydetään arvioimaan, onko maan talous tulevien 12 kuukauden aikana menossa parempaan tai huonompaan suuntaan tai pysyvän ennallaan. Vastaava kysymys on muotoiltu työttömyydelle, kotitalouden omalle taloustilanteelle ja kotitalouden kyvyille säästää. Kun näistä arvioista lasketaan keskiarvo siten, että minimi voi olla -100 ja maksimi +100, saadaan hyödyllinen indikaattori ihmisten lähiajan taloustilannetta koskevista subjektiivisista mielialoista. Ei liene vaikea arvata, että kyselytapa vaikuttaa arvioon. Vuonna 2011 toteutetussa tutkimuksessa nettivastaajat suhtautuivat pessimistisemmin kuin puhelinvastaajat (indikaattorien arvot olivat positiivisia luvuin 11 ja 17). Kun molemmat vastaukset lasketaan yhteen, pessimistisyys näkyy yhä arvona 14. Ero perinteiseen puhelinkyselyyn on lievästi merkitsevä.

Voimme uskoa, että nettivastaajat ovat rehellisempiä, koska ei ole miellyttävää kertoa pessimistisiä mielialojaan toiselle henkilölle. Mutta voimmeko julkaista sellaista tietoa? Tilastokeskuksen syyskuun 2013 Findikaattorissa todetaan: ”Kuluttajien luottamus talouteen parani hieman syyskuussa elokuuhun verrattuna. Kuluttajien luottamusindikaattori oli syyskuussa 6,3, kun se oli elokuussa ja heinäkuussa

5,0. Luottamus talouteen oli hieman vahvempi kuin vuosi sitten (3,4), mutta selvästi alhaisempi kuin pitkällä ajalla keskimäärin (12,3).”

Havaitsemme, että indikaattori on laskenut vuoden 2011 pilotin ajoilta. Näemme myös, että pienetkin erot julkistamisessa katsotaan merkittäviksi, vaikka ne eivät ole merkitseviä. Ei siis ole ihme, että Tilastokeskus ei ole innostunut käyttämään yhdistekyselyllä saatuja tuloksia virallisina lukuina. Näin siksi, että ilman uudelleen laskentaa aikasarjaan tulisi piikki alaspäin. Uusi laskenta yhden aikasarjan tapauksessa on helppoa. Sitä tehdään tavan takaa tilastoinnissa ja usein paljon monimutkaisemmissakin tilanteissa. Vuoden 2012 pilotissa tuloksen perussävy on sama eli nettiin vastanneet ovat pessimistisempiä (Oinonen 2013). Tilastokeskus tiettävästi edelleen epäroii ottaa käyttöön tämän yhdistepohjaisen kyselyn tuloksia.

Yhdistekyselyjen tulevaisuus on kuitenkin valoisaa, koska se on halvempi ja se parantaa kokonaisvastausastetta. Jälkimmäinen seikka on havaittu myös ns. eteläisen Suomen metropolitutkimuksessa, jossa netin sijasta saattoi vastata postitse.<sup>1</sup> Vastausaste ei ollut erityisen korkea, mutta oli selvästi havaittavissa, että tiettyt nettiä käyttävät väestöryhmät eivät olisi ilman tätä vaihtoehtoa osallistuneet tutkimukseen.

Sama tulos saatiin vuonna 2012 Virossa toteutetussa European Social Survey -pilottitutkimuksessa (Lumiste ym. 2013).<sup>2</sup> Netissä vastasi jopa 60 prosenttia kaikista, loput 40 prosenttia vastasi haastattelijalle. Suomen yhdiste-

kyselyissä ei näin korkeita vastausasteita ole netin puolella saatu. Viro onkin tavallaan edelläkävijä uuden tietotekniikan alalla.

Keruutavan vaikutus tulee esiin tässäkin tutkimuksessa. Esimerkiksi nettivastajat suhtautuivat maahanmuuttoon kielteisemmin kuin haastattelijalle ilman anonymiteettia vastanneet. Tämä on samanhenkinen tulos kuin edellä esitellyissä esimerkeissä.<sup>3</sup>

### Oikein tehtynä nettikysely on hyvä väline

Nettikyselyä voi pitää esimerkkien valossa hyvänä menetelmänä. Se on parhaimmillaan luotettava. Tiedonkeruun kustannusten arvioidaan olevan noin viidesosa puhelinhaastattelun ja noin kymmenesosa käyntihaastattelun kustannuksista (Aromaa ym. 2010). Postikyselyyn verrattuna laatu on samaa tasoa, mutta postikysely on kalliimpi. Paperi- ja postituskulujen lisäksi tiedot pitää erikseen tallentaa ja aikaa menee myös tarkistuksiin.

Muutama vuosi sitten nettikyselyjen pullonkaula oli huonosti tehty lomake. Nykyään tekniikka ei enää aseta esteitä hyvän lomakkeen tekemiselle. Esimerkiksi nettilomakkeen ponnahdusikkuna (pop-up) auttaa vastaajaa ymmärtämään kysymyksen oikein, ehkä yhtä hyvin kuin haastattelija olisi läsnä. Kysymys on siis vaivannäöstä ja osaamisesta, jolla lomake tehdään. Silloin vastaajalta tuhrautuu nettivastaukseen vähemmän aikaa kuin muilla tavoilla tehtyihin kyselyihin (Aromaa ym. 2010).

<sup>1</sup> *Helsingin yliopiston professorien Mari Vaattovaaran ja Matti Kortteisen hanke, jossa ovat mukana myös mm. Teemu Kemppainen ja Seppo Laaksonen.*

<sup>2</sup> *Europeansocialsurvey.org-sivuilta löytyy aineistojen lisäksi paljon hyödyllistä tietoa, myös kyselyvälineistä ml. artikkeleita yhdistekyselyistä.*

<sup>3</sup> *Tutkimuksen Minskissä kesäkuussa 2013 esitellyt Kaur Lumiste hyväksyi näkemykseni että nettivastaukset ovat luultavasti todellisempia, mutta nettikyselyjen käyttöön subtaudutaan epäroiden, koska se voisi antaa Virossa liian huonon kuvan.*

Lisää hyviä puolia ilmenee, kun nettiä käytetään muihin keruutapoihin yhdistettynä. Kyse on näin ollen mitä mainioimmasta välineestä edellyttäen, että toimitaan hyvien survey-perinteiden mukaan. Näin on menetelty edellä esitetyissä esimerkeissä. Niissä otos on toteutettu oikein siten, että tavoiteperusjoukko on määritelty tarkasti ja tavoiteperusjoukon tavoittamiseksi käytetään hyviä kehikkoja. Suomessa näitä ovat väestörekisteri ja yritysrekisteri. Toki niissäkin on puutteita, mutta samat puutteet kohdataan muitakin kyselytapoja käytettäessä.

Nykyään yleisin keruuväline on puhelin. Sillä ei kuitenkaan kyetä saavuttamaan täyttä tavoiteperusjoukkoa. Suomessa tavoitetaan pienellä työmäärällä ehkä 65 prosenttia, ja isolakin työmäärällä jäädyään alle 90 prosentin. Käyntihaastattelulla voitaisiin kaikki tavoittaa, mutta työmäärä olisi silloin suunnaton. Postiosoitteet ovat rekistereissä kaikille mutta läheskään kaikkia ei voida kenttätöaikana kotoa tavoittaa.

Useimmissa muissa maissa rekisterit ovat huonompia kuin Suomessa. Esimerkiksi Virossa 17 prosenttia ja Liettuassa 25 prosenttia rekisterin ihmisistä ei asu sen ilmaisemassa paikassa<sup>4</sup>. Kaikkialla ollaan yhä useammin pitkiä aikoja poissa kotoa, joten vastauskato on väistämätön. Vastauskato onkin isoin riesa nykypäivän kyselyissä.<sup>5</sup> Nettikyselyä varten voidaan Suomessa ja monissa muissa maissa saada siis yhtä hyvät tiedot otoshenkilöistä kuin muitakin

tiedusteluja varten. Tämän jälkeen ongelmaksi jää vastaajien houkuttelu kyselyyn osallistumiseen. Muutama vuosi sitten vain ani harva olisi ollut halukas vastaamaan nettikyselyssä. Kuten havaitsimme, Virossa vastaushalukkuus on jo varsin hyvä. Jos kenttätöaika on pitkä, aikaa houkutteluun on. Kuluttajabarometrin ongelma on lyhyt kenttätöaika, koska ihmisten näkemyksiä kysytään joka kuukausi. On luultavaa, että ihmiset voidaan pikku hiljaa opettaa nettikyselyiden kautta kertomaan elinoloistaan ja mielipiteistään sellaisia tietoja, jotka nähdään yhteiskunnallisesti tärkeiksi.

Suurin ongelma onkin siinä, että nettikyselyjä ei läheskään aina tehdä hyvien käytäntöjen mukaan, vaan laadusta ollaan valmiita eri syistä tinkimään. Useimmiten syynä on helppous ja säästäminen. Seuraavassa esitellään muutamia yleisessä käytössä olevia nettikyselymenetelmiä, joiden tilastollinen luotettavuus on kyseenalainen.

## Nettipaneeli

Markkinatutkimuslaitokset ovat jo jonkin aikaa houkuttelleet jonkinlaisilla otantapohjaisilla ratkaisuilla valittuja ihmisiä tulemaan mukaan omiin verkostoihinsa, ns. nettipaneeleihin (*opt-in panels*). Yleensä mukaan on otettu sellaisia henkilöitä, joilla jo on nettiyhteydet. Joissakin tapauksessa laitokset antavat paneelin jäsenille käyttöön tiedonkeruulaitteen, jolla voi olla nettiyhteydessä laitokseen. Toisinaan osallistumisesta maksetaan pienehkö palkkio<sup>6</sup>. Markkinatutkimuslaitos, joka on saanut rekrytoiduksi 50

<sup>4</sup> Tämä on peräisin tiedoista, jotka olen saanut toimiessani European Social Surveyn otantaa ohjaavassa ryhmässä.

<sup>5</sup> Konsultointimatkoillani vuosina 2010-2011 havaitsin tästä kiinnostavan poikkeuksen. Se on Etiopia, jossa väki ei lähde kauas kotoa ja naapuri tietää, missä tämä on, jos osaa oikein kysyä.

<sup>6</sup> Olin itse koemielessä muutaman kuukauden Nielsenin verkostossa. Sieltä sain käyttööni tiedonkeruulaitteen ostostietojen keräämiseksi. Osallistumisesta maksettiin vaatimaton palkkio.

000 ihmistä rekisteriinsä, voi kulloisenkin tarpeensa mukaan lähettää heistä otetulle otokselle lyhyempiä tai pidempiä kyselyjä. Jos nämä periaatteessa vastaamaan suostuneet henkilöt antavat vastauksensa heti, tulokset ovat välittömästi käytettävissä.

Suomessa tällaista menetelmää käytettiin viime eduskuntavaalien yhteydessä laskettaessa arvioita ehdokkaiden henkilökohtaisista äänimääristä Helsingissä ja Espoossa (*Helsingin Sanomat* 2.4.2011 ja 3.4.2011). Pääosin tulokset olivat aika odotettuja. Lukujen virhemarginaalit olivat kuitenkin isoja, eivätkä kärkisijojen ulkopuolelle jääneet tuntuneet pitävän tulosten julkistamisesta.

Nettipaneelit ovat esimerkki ei-todennäköisyysperiaatteille sovitetusta aineistosta. Keskustelua on käyty siitä, voidaanko esittää kaava, jonka avulla ei-todennäköisyysperiaatteille perustuva ”otos” saataisiin estimoiduksi.<sup>7</sup> Sellaista kaavaa ei ole löydetty. Tämän pitäisi olla selvä kaikille, koska kaava vaatisi tietoa siitä, miten aineisto on koottu eli tilastotermein tietoa sisällyttämistodennäköisyyksistä. Tämän pohjaksi taas vaadittaisiin muun muassa tavoiteperusjoukon tarkka määrittely sekä hyvä kehikko, josta tietoa pyritään keräämään. Vielä isompi ongelma on vastauskato ja muu puuttuneisuus. Eli jos kaavan aikoiisi muodostaa, olisi kehitettävä nuo tiedot. Liian yleinen tapa on olettaa, että aineisto vastaa tavoiteperusjoukkoa edustavasti. Tämä oletus voi johtaa pahasti harhaan.

Baker ym. (2013) käyvät läpi ei-todennäköisyysotantaan liittyviä ongelmia ja pyrkimyksiä ratkaista niitä. Nettipaneelit ovat yksi osa näitä. He toteavat, että niillä on arvoa jossakin tutki-

muksessa mutta varoittavat samalla niiden vääroista. Yksi apu olisi saada oikeata tai ainakin olennaisesti oikeampaa tietoa edes joistakin samoista muuttujista, joita nettipaneelissa on mukana. Tällainen tieto auttaisi kalibroimaan tuloksia parempaan suuntaan, mutta se ei poistaisi harhaa.

Hollantilaisten kehittämä LISS (*Panel - listening to people*) on tavanomaisia nettipaneelija huolellisemmin toteutettu nettikysely<sup>8</sup>. Scherpenzeel (2009) pitää sitä todennäköisyysperiaatteille pohjautuvana nettipaneelina. Sitä se ei kuitenkaan ole. Myös siinä on panelistien rekrytointi ja pyrkimyksenä on saada vastaajajoukko mahdollisimman edustavaksi. Osalle vastaajista lainataan tietokone vastaamisen helpottamiseksi, ja palkkioksi nämä voivat käyttää konetta myös omiin tarkoituksiinsa. Tällaisen paneelin kokoaminen ei ole tarkoituksenmukaista yksittäistä tiedonkeruuta varten, koska perustamiskustannus on korkea. LISS:iä käytetäänkin kuukausittaisena paneelina eli mukana olijoilta pyritään saamaan vastaukset pidemmän aikaa. Tällaisen strategian voisi uskoa toimivan muuallakin, vaikkapa Tilastokeskuksessa. Pidempiaikaisessa mukanaolossa vastausrasite yhtä kyselyä kohti vähenee. Jossain vaiheessa toki vastaajat saattavat kyllästyä, jolloin laatu heikkenee.

LISS –paneelissa vastausasteet ovat melko korkeita, ensimmäisinä kuukausina jopa lähellä 80 prosenttia, mutta alenevat myöhemmin noin 60 prosentin pintaan. Nämäkin luvut ovat niin suuria, että voi epäillä, että aineistossa on samoja ei-todennäköisyysotannan piirteitä kuin tavallisissa nettipaneelissa.

On oletettavaa, että tulevaisuudessa nettikyselyiden otos otetaan enenevässä määrin so-

<sup>7</sup> Äskettäin seurasin, itsekin osallistuen, nettikeskustelua juuri näistä otannoista. Yksikään yli 50:stä keskustelijasta ei esittänyt mitään kaavaa, jolla ’otos’ olisi voitu estimoida.

<sup>8</sup> <http://www.lissdata.nl/lissdata/>

siaalisista verkostoista (Facebookista, Twitteristä, LinkedInistä). Baker ym (2013) esittävät näkökohtia myös tästä uudesta ilmiöstä (*online network sampling*). Heidän mukaansa näitä otantamenetelmiä ovat kehitelleet elektroniikkainsinöörit ja tietotekniikan osaajat, joilla tuskin on erityistä tilastollisen otannan tuntemusta.<sup>9</sup>

### Itsevalikoituminen

Mikäli nettikyselyä ei voida suorittaa todennäköisyysperiaatteisiin nojautuen, voidaan turvautua itsevalikoituneeseen verkkotiedonkeruuseen. Itsevalikoitumisella tarkoitetaan sitä, että verkkokyselyä mainostetaan avoimesti tutkimuksen kohderyhmälle ja vastaajat itse päättävät osallistuvatko kyselyyn vai eivät.

Miettinen (2011) on pro gradu -työssään testannut itsevalikoivuutta Kelan tarpeita silmällä pitäen. Kelan tutkimusosasto keräsi kahdet tiedot vuoden 2010 syyskuussa. Molemmissa pyrittiin kartoittamaan Suomen 16–80 -vuotiaiden kansalaisten yksityisten terveys- ja hammashoitopalvelujen hintatietoja ja tarpeita. Kyselyt tehtiin netissä, joista toinen oli itsevalikoitunut siten, että mahdollisuutta vastata mainostettiin haluttuja vastaajaryhmiä lähellä olevissa piireissä. Samalla tehtiin otosperusteinen kysely Kelan työntekijöille. Kolmanneksi tutkimuksessa tarkasteltiin TNS Gallupin Suomen kansalaisilta keräämää suppeampaa puhelinhaastattelukyselyä, jota käytettiin myös itsevalikoituneen aineiston kalibrointiaineistona.

<sup>9</sup> Cressie ym. (2013) kertovat kokemuksista aluetietojen estimoimiseksi sosiaalisen median tietojen avulla. He viittaavat mm. Google Trends -aikasarjoihin. Näiden ja netistä löytyvien karttatietojen alueella tulee lähivuosina tapahtumaan paljon.

Tutkimuksen keskeisin tulos oli, että vaikka verkkokysely on halvempi ja vähemmän työläs kuin otosperusteinen kyselytutkimus, verkkokyselytutkimuksen suunnitteluun on panostettava entistä enemmän. Itsevalikoitunutta kyselyä on mainostettava hyvin vastaajaehdokkaille, jotta valikoitumisharhaa voitaisiin hallita. Tutkimuksessa kävi myös ilmi, että kyselytutkimuksiin on tärkeää ottaa mukaan asenteita ja käyttäytymistä kuvaavia kysymyksiä, koska niiden avulla pystytään selittämään valikoitumisesta johtuvaa harhaa paremmin kuin pelkästään sosiodemografisten muuttujien avulla.

Tutkijalla oli siis käytössään kaksi vertailuaineistoa. Ei-todennäköisyyspohjaisissa aineistoissa se on elintärkeä asia, koska silloin voidaan jossain määrin katsoa kuinka hyvä ”otos” on ja harhaa voidaan myös yrittää vähentää, mikäli vertailuaineisto katsotaan riittävän luotettavaksi. Miettinen teki parhaansa mutta toteaa, että ”itsevalikoituneen aineiston harhaa ei saatu painottamalla täysin korjattua. Yksi syy on, että verkkoaineisto poikkosi tavoiteperusjoukosta kaikkien käytettävissä olevien taustamuuttujien myötä. Tämän vuoksi tulosmuuttujien jakaumat olivat harhaisia. Jos valikoitumismekanismeja pystyttäisiin täydellisesti selittämään ja aineiston koko olisi tarpeeksi suuri, harha saataisiin korjattua ja jakaumatiedot vastaisivat tavoiteperusjoukkoa.”

Kuitenkin esimerkiksi yksityisten ja julkisten terveyspalvelujen käyttömääriä onnistuttiin painottamalla korjaamaan, mutta samalla Kelakorvauksen estimaatit harhaantuivat entistä enemmän. Tämä johtuu ensiksikin siitä, että valikoitumismekanismeja ei pystytty selittämään ja toiseksi siitä, että aineiston koko oli liian pieni kuvastamaan tavoiteperusjoukon vaihtelua.

Miettisen työ on pioneerimainen ja siksi kunnioitettava. Herää selvästikin tarve yrittää

kehittää valikoitumismekanismeja entistä paremmaksi, mutta miten se kussakin tilanteessa tapahtuu, on hankala kysymys. Lisää tutkimusta toisin sanoen tarvitaan. Itsevalikoitumispohjaisissa nettikyselyissä on myös mahdollista, että innokas houkuteltu vastaa useita kertoja kenties vaikuttaakseen tutkimuksen tulokseen. Apuna tähän voi olla evästeiden käyttö, jolloin samasta palvelimesta lähetetyt vastaukset voivat paljastua, ellei vastaaja ole ottanut sen ilmaisinta pois päältä.

Baker ym. (2013) pohtivat myös itsevalikoitumista tai vapaaehtoisuuskyselyjä. He toteavat, että niistä suljetaan pois ne, joilla ei ole pääsyä nettiin tai se on liian vaikeata. Tämä siis väistämättä aiheuttaa harhaa eikä sen suuruutta voida ilman hyvää vertailuaineistoa oikoa. Jos vertailuaineisto on olemassa, on mahdollista painottaa itsevalikoitunut aineisto sen tiedolla. Esimerkiksi vastaustodennäköisyyksiin perustuva painotus voi tässä olla hyödyksi (Laaksonen 2007) samoin kuin ratkaisut, joita Valliant ja Dever (2011) ovat esittäneet. Jos tehtävä onnistuu hyvin, voidaan molemmat aineistot yhdistää ja saada näin käyttöön laajempi kokonaisaineisto ainakin sellaisista väestöryhmistä, jotka käyttävät nettiä. Tämä tarkentaa tuloksia vastaavasti.

### **Holtittomat nettikyselyt**

Lopuksi kannattaa mainita ns. holtittomat nettikyselyt. Tämä ei tarkoita asiakaskyselyjä, joihin itse kutakin tavan takaa pyydetään vastaamaan. Ne saattavat täyttää tehtävänsä, vaikka moni vastaakin mieluummin silloin, kun palvelusta on jotakin valittamista.

Holtittomista nettikyselyistä voidaan puhua tapauksissa, joissa on jokin aito tutkimustavoite,

mutta itse työ tehdään holtittomasti. Kyselyn tekeminenhän on helppoa myös siksi, että netistä löytyvät ilmaiset välineet. Esimerkkinä seuraava ote SurveyMonkeysta: ”SurveyMonkey on maailman suosituin verkkotutkimustyökalu. Ilmaisten kyselytutkimusten, mielipidetutkimusten, kyselylomakkeiden, asiakaspalautteiden ja markkinatutkimusten lähettäminen on helpompaa kuin koskaan. Saat lisäksi käyttöösi ammattilaisten laatimia tutkimuskysymyksiä ja lomakkeita.” SurveyMonkey on varmasti ihan hyvä, jos sitä osaa oikein käyttää.

Monet holtittomat nettikyselyt ovat kuin nuo itsevalikoituneet kyselyt. Tavoitteena on vain saada riittävä määrä vastauksia ilman, että mielessä on mitään laatutavoitetta tai vertailumahdollisuutta hyvään aineistoon. Tutkimuksen julkistamisessa kerrotaankin lähinnä vastaajien määrästä, ei laadusta. Joku voikin uskoa 10000 vastaajan aineistoon enemmän kuin edustavaan 2000 vastaajan aineistoon. Löysemmin toteutettuja nettikyselyitä toki voi tehdä esimerkiksi pienille joukoille (työyhteisöt, opiskelija- ja harrasteryhmät), joiden kokonaistiedot ovat tiedossa eli vertailumahdollisuus on olemassa. Kyselyt sopivat myös kvalitatiivisiin tarkoituksiin siten, että tuloksia ei pyritäkään yleistämään mihinkään tavoiteperusjoukkoon. Kyse voi silloin olla perusteellisen kyselyn pilotoinnista. Jos oppi käytetään kunnon tutkimukseen, hyöty on saavutettu. Mutta aivan liian paljon tehdään johonkin epämääräiseen joukkoon kohdistuvia kyselyjä vain sen vuoksi, että tehtävä on helppo. Sopii olettaa, että Kansantaloudellisen aikakauskirjan lukijat eivät lähde sellaiseen mukaan. □

## Kirjallisuus

- Aromaa, K., Heiskanen, M., Laaksonen, S., Nikula, J. ja Ruuskanen, E. (2010), ”Finnish results from the EU pilot survey”, Final report delivered to the European Commission on 22 February 2010, julkaisematon (saatavissa tekijöiltä).
- Baker, R., Brick, J.M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J. ja Tourangeau, R. (2013), *Report of the AAPOR task force on non-probability sampling*, <http://www.aapor.org/AM/Template.cfm?Section=Reports1&Template=/CM/ContentDisplay.cfm&ContentID=5963> (viitattu 21.8.2013).
- Bethlehem, J. ja Biffignandi, S. (2012), *Handbook of Web Surveys*, Hoboken, New Jersey: John Wiley & Sons Inc.
- Cressie, N., Holan, S.H. ja Wikle, C.K. (2013), ”How can survey estimates of small areas be improved by leveraging socialmedia data?”, *Survey Statistician*, 14-15, <http://isi.cbs.nl/iass/index.htm> (viitattu 7.8.2013).
- Laaksonen, S. (2007), ”Weighting for Two-phase Surveyed Data”, *Survey Methodology* 33: 121-130.
- Laaksonen, S. (2011), ”A Single Mode vs a Mixed-Mode Design”, *Proceedings of International Conference of the European Survey Research Association*, July 2011, Lausanne.
- Laaksonen, S. (2013), *”Surveymetodiikka. Aineiston kokoamisesta pubdistamisen kautta analyysiin”*, Ventus Publishing, <http://bookboon.com/fi/surveymetodiikka-ebook>.
- Lumiste, K., Ainsaar, M., Lilleoja, L. ja Roots, A. (2013), ”European Social Survey Mixed Mode Experiment in Estonia: CAWI and CAPI Sequential Design”, *Baltic-Nordic-Ukrainian Summer School of Survey Statistics*, June 13-18, Minsk.
- Miettinen, J. (2011). *Itsevalikoituminen verkkokyselytutkimuksen haasteena*. Helsingin yliopisto, Valtiotieteellinen tiedekunta, Tilastotiede, Pro gradu –tutkielma.
- Oinonen, S. (2013), ”Mixed mode data collection pilot survey on consumer survey: Results on response,” *Baltic-Nordic-Ukrainian Summer School of Survey Statistics*, June 13-18, Minsk.
- Scherpenzeel, A. (2009), ”Start of the LISS panel: Sample and recruitment of a probability-based Internet panel”, *CentERdata*, Tilburg.
- Valliant, R. ja Dever, J.A. (2011), ”Estimating Propensity Adjustments for Volunteer Web Surveys”, *Sociological Methods & Research* 40:105–137.