



Om begreppet evidens i evidensbaserad medicin

OSKAR LINDHOLM & PEKKA LOUHIALA

Evidensbaserad medicin (Evidence-Based Medicine, EBM) har alltsedan sin uppkomst för trettio år sedan genererat såväl inflytande som uppmärksamhet. Att en rörelse vars mest citerade tes yrkar på "den samvetsgranna, explicita och omdömesgilla användningen av den nuvarande bästa evidensen vid vården av individuella patienter"¹ (Sackett et al 1996, 71) har uppbådat en hel del debatt kan te sig överraskande. Vilken läkare skulle inte utöva – eller åtminstone vilja utöva – en dylik medicin? Och vem skulle inte underskriva det förnuftiga i att "integrera klinisk erfarenhet med den bästa tillgängliga evidensen, som baserar sig på systematisk forskning" (ibid)?

Kruxet ligger, naturligtvis, i detaljerna. Dispyterna under coronaviruspandemin – då medicinsk forskning hamnat i blickfånget av såväl offentlig som akademisk diskussion världen runt – har inte handlat så mycket om huruvida våra beslut ska stödas av god evidens eller inte, utan snarare om precis vad som utgör god evidens i vilken kontext. Därtill kvarstår frågan om hur medicin eller policybeslut exakt skall "basera" sig på denna evidens.

¹ Alla översättningar från engelskan är artikelförfattarnas egna.

Det är angående frågor som dessa som EBM enligt vetenskapsfilosofen John Worrall (2010a, 356) hamnat i "total oreda". Ett av EBM-rörelsens obestridliga framsteg, den systematiska insamlingen och summeringen av ett stort antal primärundersökningar med hjälp av strikta metodologiska kriterier, har med tiden visat sig vara problematiskt. Volymen evidens är numera "ohanterlig" och "outgrundlig", och resultaten ur olika studier motsäger varandra (Greenhalgh et al 2014, 2). Att påstå att ens egen vård eller handlingsprogram baserar sig på evidens kan grunda sig mer i väl genomförd selektivitet än den underliggande evidensens kvalitet, speciellt i de fall ordet används vagt eller ospecificerat. Även om EBM numera omfattas av så gott som alla – läkarorganisationer, regeringar, försäkringsbolag, farmaceutiska bolag, patientorganisationer – har de olika intressenterna inte alltid haft samma skäl att göra det, och orden "evidens" och "evidensbaserat" används idag ställvis som rena marknadsföringsknep.

I denna artikel granskar vi EBM:s syn på evidens. Utgående från de ställningstaganden EBM gör berörande medicinsk teori markerar vi särskilt punkter där EBM:s evidenssyn tycks skapa en spänning mellan EBM och medicinsk praktik. I det första avsnittet presenterar vi EBM och de väsentliga särdragen i de för artikeln relevanta metoder som används i medicinsk forskning. I det andra avsnittet diskuterar vi EBM:s syn på evidensens natur, och evidensens förhållande till hypoteser. I det tredje avsnittet diskuterar vi förhållandet mellan evidens och praktik. Kan en praktik verkligen "basera" sig på evidens? Hur går man från evidens till riktlinjer till praktik? I det fjärde avsnittet slår vi fast att det främsta intresset i EBM ligger på vad som "fungerar" – det vill säga evidens om kausala samband – men att kausala hypoteser i EBM konfirmeras och förnekas på varierande sätt. I EBM ställer man sig som helhet motvilligt till att dra kausala slutsatser ur data: de olika interventionernas effekter presenteras istället som den ökade nyttan eller risken i en population. Men i klinisk medicin ligger huvudintresset på den enskilda patienten, och kausala hypoteser är oundvikliga. Denna

slutsats leder oss till det femte avsnittet, där vi diskuterar metodologiska bekymmer genom att fokusera i synnerhet på randomiserade kontrollerade studier – en gyllene standard bland undersökningsmetoder i EBM. Det första huvudsakliga problemet gäller skillnaden mellan inre och yttre validitet. Det andra gäller frågor där andra undersökningsmetoder än randomiserade kontrollerade studier av en eller annan orsak är nödvändiga, och exemplifieras i synnerhet av skador som orsakas av medicinska interventioner. I det sjätte avsnittet diskuterar vi problematiken med att generalisera resultaten ur kliniska undersökningar till enskilda patienter. EBM:s ställningstaganden tycks förutsätta att samma intervention har liknande effekt på olika människor, men det är inte klart att detta antagande på bästa möjliga sätt representerar den kausala verkligheten i klinisk medicin. Vi avslutar genom att summera våra huvudsakliga slutsatser och understryka de frågor som är av intresse i framtiden.

EBM

EBM har sina rötter i den *kliniska epidemiologi* som utvecklades i Kanada och England under 1970- och 80-talet. Går man ännu längre bak i tiden kan man hitta tidiga ideologiska förfäder redan i början av 1800-talet, då Thomas Beddoes i England föreslog en systematisk klassificering av den vetenskapliga kunskapen och Pierre Louis i Frankrike ledde en rörelse vid namn *Médecine d'Observation* (Goodman 2003, 3–5). Tanken i *Médecine d'Observation* var att enskilda patientfall och "teoretisk spekulering" om sjukdomars mekanismer och natur var otillräckliga för att fatta beslut vid vården av patienter. Istället skulle läkare samla stora mängder observationell data och sammanfatta informationen numeriskt, vilket skulle avslöja vårdens sanna effekt (Vandenbroucke 1996, 1335).

Att läkare inte bör förlita sig på teoretisk spekulering kan anses utgöra en grundsten i EBM. I den artikel där EBM ursprungligen presenterades år 1992 heter det att EBM "minskar betydelse-

sen av intuition, osystematisk klinisk erfarenhet och patofysiologisk slutledning vid det kliniska beslutsfattandet, och understryker vikten av evidens ur kliniska undersökningar” (Evidence-based Medicine Working Group 1992, 2420). Med begreppet ”kliniska undersökningar” avses i detta sammanhang *komparativa kliniska studier*. Termen inbegriper såväl *randomiserade kontrollerade studier* (hädanefter RCT; eng. *randomised controlled trial*) som *observationella studier*. Eftersom dessa undersökningars särdrag är relevanta i sammanhanget kommer de kort att presenteras.

I en RCT, en gyllene standard i EBM, sorteras personer ur undersökningspopulationen slumpmässigt i en av två eller flera grupper. De som sorteras i *försöksgruppen* mottar den intervention vars effekt är av intresse. De som sorteras i *kontrollgruppen* mottar antingen placebo eller en redan vedertagen behandling för sjukdomstillståndet ifråga. Därefter jämförs de skillnader i mätta slutresultat som uppstått. En viktig komponent i en randomiserad studie är *blindning*: såväl de undersökta som undersökarna bör i mån av möjlighet vara omedvetna om vilken grupp den enskilda patienten tillhör för att denna kunskap inte skall snedvrída slutresultaten.

Observationella studier skiljer sig från RCT såtillvida som ingen särskild intervention läggs in. Observationella studier ingriper alltså inte direkt i vården, utan jämför skillnader som uppkommer eller redan har uppkommit mellan olika grupper som vårdas i icke-experimentell miljö. I en observationell studie är grupperna därmed inte slumpmässigt fördelade vad interventionen ankommer, det vill säga *randomiserade*, vilket anses utgöra en brist. Exempelvis existerar det en korrelation mellan kaffekonsumtion och hjärt- och kärlsjukdomar, men personer som dricker mera kaffe röker även i medeltal mer än personer som dricker mindre kaffe. Eftersom rökning är en riskfaktor för hjärt- och kärlsjukdomar fungerar den i detta fall som en *blandande faktor* med avseende på sambandet mellan kaffekonsumtion och hjärt- och kärlsjukdomar (Hammer et al 2009). Som blandande faktor fungerar vilken som helst variabel utöver den undersökta interventionen som inverkar på slutresultatet och

är ojämnt fördelad mellan undersökningsgrupperna. En blandande faktor kan relatera till såväl patienternas egenskaper (exempelvis ålder, kön, genetiska faktorer), sjukdomens egenskaper (etiologi, svårighetsgrad, duration) och undersökningens egenskaper (dosering, komplians, vårdpersonalens intresse, blindningens framgång). Genom att randomisera – fördela patienterna i antingen försöksgrupp eller kontrollgrupp slumpmässigt – tänker man sig att de blandande faktorerna fördelas jämnt, och att genuint kausala samband kan utvinnas.

Komparativa kliniska studiers särdrag kan anses vara att de förmår mäta effekten av en intervention direkt, utan att en förståelse om hur interventionen fungerar skulle vara nödvändig. Det är genom sin betoning på denna form av studier som EBM skiljer sig från det gamla "paradigmet" i medicin (Evidence-based Medicine Working Group 1992). Fortfarande under början av 1900-talet ansågs patofysiologisk härledning (evidens för hur olika sjukdomsförlopp utvecklas och reagerar på interventioner) utgöra pålitligare evidens än statistiskt härledd evidens. Genom en förståelse av fysiologi, anatomi och andra grundvetenskaper tänkte man sig kunna resonera sig till den bästa vården för individuella patienter. Flera av de medicinska framstegen som gjordes under 1800- och 1900-talen kan påstås ha varit avhängiga av just denna förståelse om *varför* något fungerar: exempelvis krävde utvecklingen av rabiesvaccinet en förståelse om virologi, medan utförandet av njurtransplantationer byggde på en kunskap om immunsystemets funktion (Le Fanu 1999). Tillvägagångssättet kan på ett större plan härledas åtminstone ända till läkare i hippokratisk tradition, som framställde sjukdomar som orsakade av enskilda verkande faktorer, nämligen obalans i de fyra kroppsvätskorna, och botemedlen följaktligen som korrigeringar av denna obalans (Bluhm & Borgerson 2011, 204).

EBM uttrycker explicit ståndpunkten att komparativa kliniska studier ger bättre evidens än andra metoder. Detta representeras i *evidenshierarkierna*, där evidensen rangordnas oftast enligt den undersökningsmetod som producerat evidensen ifråga. Existensen av en evidenshierarki kan anses utgöra den

första och äldsta EBM-principen (Murad et al 2016, 125). Idén om att rangordna evidens står i kärnan av EBM, och sättet den rangordnas på har genom tiden inte förändrats avsevärt (Howick 2011, 23). Grundidén kan sammanfattas på följande sätt (ibid, 4):

- (1) RCT eller systematiska granskningar av många RCT är generellt sett bättre än observationella studier
- (2) Komparativa kliniska studier är generellt sett bättre än mekanistisk härledning
- (3) Komparativa kliniska studier är generellt sett bättre än klinisk expertis

Vid skrivande stund har inte mindre än 195 olika evidenshierarkier identifierats (Blunt 2021). Vissa hierarkier rangordnar endast olika varianter av komparativa kliniska studier, medan andra inkorporerar exempelvis mekanistisk härledning och expertis, alltid på botten av hierarkin. Precis hur en enskild hierarki ska tolkas – vilken specifik information som kan dras om evidensen utgående från hierarkin ifråga – är oftast upp till granskaren att avgöra (Blunt 2015, 40). I tidigare EBM var tolkningen ofta absolut och stark. Exempelvis föreslår Straus et al (2005, 118) att du, ifall studien inte var randomiserad, ”slutar läsa och går till nästa artikel i din sökning. [...] Återvänd endast ifall du inte kan hitta randomiserade undersökningar om ämnet ifråga.” Moderna hierarkier som exempelvis GRADE (Guyatt et al 2008a) innehåller dock upp- och nedgraderingskriterier som möjliggör en viss rörelse mellan de olika nivåerna i hierarkin.

En modern utveckling i EBM är den ökade betoningen på förhandsbedömd evidens, som inte minst på grund av den stora massan komplex statistisk evidens visat sig vara oundviklig (Greenhalgh et al 2014). I systematiska granskningar summerar sakkunniga resultaten ur en eller flera primärstudier genom att använda sig av explicita kriterier vid insamlingen och bedömningen av resultat. I samband med systematiska granskningar görs ofta en *meta-analys*, där den relativa betydelsen av olika primärstudiers resultat statistiskt vägs i förhållande till

varandra och summeras till ett genomsnittresultat. Utgående från resultaten av en systematisk granskning kan en rekommendation eller riktlinje för praktiken utfärdas. Även om ett av de ursprungliga målen med EBM var att förbättra den enskilda läkarens förmåga att kritiskt bedöma evidensen ur olika kliniska undersökningar, utgör hon i nuvarande EBM oftast en användare av förhandsbedömd evidens, inte en evidensbedömare (se exempelvis Daly 2018, 91).

Optimalt kliniskt beslutsfattande kräver enligt EBM en medvetenhet om den bästa möjliga evidensen så som den definierats i evidenshierarkierna (Guyatt et al 2008b, 10). Sålunda gör EBM normativa påståenden om både medicinsk teori och praktik: den bästa evidensen fås ur komparativa kliniska undersökningar, i synnerhet RCT, och detta är den evidens de kliniska besluten skall "basera" sig på. Användningen och betydelsen av evidenshierarkierna begränsar sig dock inte enbart till medicinsk praktik. Principerna bakom evidenshierarkierna används numera på sätt eller annat vid utfärdandet nationella vårdrekommendationer – exempelvis God Medicinsk Praxis (Käypä Hoito) i Finland – och är således av betydelse vid frågor som berör exempelvis försäkringsersättningar (Bolt 2015, 20). Att termer som "evidensbaserad policy" och "evidensbaserad hälsovård" fått genomslag tyder på att EBM antagit en "makrodimension" med avsevärd samhällsrelaterad betydelse (ibid, 17).

Evidensbegreppet i EBM

Tyngdpunkten i diskussionerna om evidens i EBM har för det mesta legat på metodologi – frågor om olika metoders för- och nackdelar, källor till snedvridning och resultatens giltighet utanför studiepopulationen, till exempel. Ett annat vanligt fokus har varit *vems* evidens det egentligen är som används i EBM, och för vilka ändamål (se exempelvis Witkin & Harrison 2001; Goldenberg 2006; De Vries et al 2008). En fråga som däremot tenderat bli förbisedd är precis vad evidens egentligen är. Det är inte oväsentligt i sammanhanget. Som vi kommer att se kan vissa av kontroverserna kring EBM påstås bota i oklarheter

kring evidensens natur och dess roll i klinisk praktik. I detta avsnitt diskuterar vi EBM:s syn på evidensens natur och dess förhållande till hypoteser, för att i nästa avsnitt diskutera hur evidensen förhåller sig till en praktik.

Ordet evidens intar en central roll i såväl epistemologi som vetenskapsfilosofi. Den förmodligen vanligaste och minst kontroversiella definitionen på evidens är som "det som berättigar en tro" (Kelly 2016). Att säga att E är evidens för H betyder att E stöder en tro på att H är sann. Detta är hur ordet vanligtvis används i vetenskap och i EBM (Djulbegovic et al 2009, 160). En dylik definition berättar inget om vilka saker kan utgöra evidens eller till vilken grad en hypotes trovärdighet bör påverkas för att vi skall kunna tala om konfirmerande evidens. Inte heller finns det något i konceptet som automatiskt placerar till exempel RCT ovanom all annan form av data. Stödfunktionen som evidensen utför kan i princip grunda sig på empiriska observationer, sensoriska upplevelser eller annan form av information. Detta är också en av EBM:s explicita teser (Djulbegovic et al 2015).

Enligt vissa probabilistiska konceptioner på evidens är E evidens för H om E höjer sannolikheten för att H är sann. Men det finns naturligtvis situationer där E höjer sannolikheten för att H är sann så litet att det inte orsakar en förändring i vår uppfattning om evidensens styrka eller graden av konfirmation. Inhandlandet av en Eurojackpotkupong ökar visserligen sannolikheten för att vinna huvudvinsten i Eurojackpot men att påstå att kupongen är "evidens för" en huvudvinst ter sig, åtminstone i vetenskapliga sammanhang, alltför inklusivt. EBM har således till synes accepterat ett tilläggskrav som Worrall (2010a, 358) presenterat: för att kunna räknas vara i stark favör för en teori bör evidensen "inte endast förklaras av teorin, utan också i annat fall 'vara osannolik'". Worrall (2007a, 1015) kallar detta "vetenskapligt sunt förnuft". God evidens talar samtidigt emot andra hypoteser och gör dem mindre trovärdiga. Det är denna princip som ligger bakom idén för att kontrollera en studie med en grupp som inte mottar den undersökta interventionen. Vår

evidens är inte stark om en teori eller tidigare kunskap innefattar en möjlig blandande faktor som förklarar det resultat vi fått.

Det viktiga med evidensen är alltså dess funktion, sättet på vilket den stöder konfirmerandet eller förnekandet av en hypotes. I denna mening är evidens inte ett objekt eller fenomen i avsaknad av ett förhållande till en hypotes. För att använda ett slitet exempel utgör en blodig kniv som sådan inte "evidens" i sig själv. Däremot gör den det i förhållande till hypotesen att butlern verkligen är förövaren. Denna skillnad mellan så kallad *evidens simpliciter* och evidens som står i förhållande till en hypotes har ställvis förgrumlats i EBM. I EBM jämföras evidens ibland med undersökningsresultat, det vill säga individuell data som aggregerats och utsatts för ett eller fler statistiskt test (Clarke 2014). Exempelvis definierar Sackett et al (2000, 1) den bästa evidensen som:

...kliniskt relevant undersökning, ofta ur medicinsk grundvetenskap, men i synnerhet ur patientcentrerad klinisk undersökning beträffande träffsäkerheten och precisionen av diagnostiska test (den kliniska undersökningen inbegripet), kraften av prognostiska markörer och effekten samt säkerheten av terapeutiska, rehabiliterande och preventiva procedurer.

I synnerhet i en klinisk kontext är ett likställande av undersökningsresultat med evidens inte att föredra. I EBM används evidens som ett "pålitligt tecken på, symptom av eller märke för det evidensen är evidensen för", och pekar således "bortom sig själv" genom att öka trovärdigheten för att ett visst påstående är sant (Djulgovic et al 2009, 160). I evidenshierarkierna ges evidensen egenskaper som styrka och pålitlighet, och binds därigenom till en viss hypotes. Evidenshierarkierna är verktyg som används för att påverka medicinsk praktik, och som praktik kräver medicin att en viss teori tillämpas och att andra teorier förkastas, att potentiellt motstridig evidens vägs och bedöms och att värderingar bundna till det specifika problem som föreligger beaktas. För att ett undersökningsresultat eller ett

faktum, att patienten har ont i höften till exempel, skall bli evidens bör det alltså stå i förhållande till en hypotes – att patienten har artros i höften. I denna mening är evidens inom EBM alltid evidens "för någon, om något" (Martini 2020, 302).

Precis hur evidensen inverkar på olika hypotesers sanningshalt är en komplicerad fråga. Djulbegovic et al (2009, 158) skriver att EBM:s teoretiska grund "föreslår ett specifikt förhållande mellan teori, evidens och kunskap" och "kan förstås som ett epistemologiskt system". EBM intar dock inte enligt författarna "en rigorös epistemologisk ståndpunkt", utan

...använder sig entusiastiskt av alla betydande filosofiska teorier berörande vetenskaplig evidens [...] EBM borde inte konstrueras som en ny vetenskaplig eller filosofisk teori som förändrar medicinens natur eller vår förståelse av den. I stället borde EBM förstås som en ständigt utvecklande heuristisk struktur för optimerandet av klinisk verksamhet.

Oaktat skillnaden mellan ett system som är teoretiskt adekvat och heuristiskt adekvat är det värt att notera att EBM, trots denna uttryckta neutralitet vad filosofiska evidenskonceptioner ankommer, i praktiken bäst verkar lämpa sig till en evidentialistisk ståndpunkt (Blunt 2015, 30). Ur en evidentialistisk position är det uteslutande evidensen som avgör vad en rationell aktör är berättigad att tro, och enligt en stark evidentialistisk position kommer två rationella aktörer som begrundar samma uppsättning hypoteser och har tillgång till samma evidens att uppnå exakt samma uppfattningar (Kelly 2016). Detta är även vad evidenshierarkierna tycks deklarerar: i dem ges ofta absoluta mått på den tro en hypotes åtnjuter enbart utgående från den evidens granskaren har tillgång till. Ingen EBM-beskrivning tycks ta på allvar den bayesianska idén att samma evidens kan leda till accepterandet av vitt åtskilda hypoteser beroende på de förhandsuppfattningar granskarna har. EBM ger dessutom randomisering en särskild epistemisk roll, vilket bayesia-

ner som utgångspunkt inte gör (även om stora skillnader mellan dem som betraktar sig som bayesianer existerar).² I EBM är det fortfarande kutym att förneka eller konfirmera hypoteser genom så kallad *klassisk hypotesprövning*, där konfirmationen eller förnekandet av en hypotes sker vid ett på förhand bestämt tröskelvärde eller konfidensintervall, utan att sannolikheten för att hypotesen är sann beaktas. Ur ett evidenshierarkiperspektiv tycks även en falsifikationistisk modell passa dåligt in i ett EBM-konceptualiserat system såtillvida som frågeställningarna i hierarkier och riktlinjer allt som oftast handlar om evidens för olika hypoteser, istället för evidens mot dem (Blunt 2015, 30).

För att sammanfatta är evidens i EBM något som berättigar en tro. Evidensrelationen handlar om stöd: evidensen konfirmerar eller förnekar en hypotes. EBM uttrycker inte explicit en rigorös epistemologisk ståndpunkt gällande evidensens natur, men verkar i praktiken och speciellt ur ett evidenshierarkiperspektiv inta en evidentialistisk position. Principiellt kan vilka som helst empiriska observationer eller sensoriska upplevelser utgöra evidens i EBM, men god evidens talar samtidigt emot tävlande hypoteser, vilket är "vetenskapligt sunt förnuft".

Om att "basera" sig på evidens

Diskussionen hittills har rört förhållandet mellan evidens och isolerade hypoteser. Innan vi lägger fokus på vad det egentligen är vi vill ha evidens för i EBM är det lönt att diskutera hur evidensen förhåller sig till en praktik, inte minst eftersom det varit en stöttesten i diskussioner om EBM. EBM har exempelvis ända sedan sin begynnelse kritiserats för att förfäktat "kokboksmedicin", det vill säga slaviskt följande av evidensbaserade riktlinjer utan hänsyn till läkarens kliniska erfarenhet och den specifika kliniska situationen (Knaapen 2014). I synnerhet patienternas värderingar har ansetts blivit förbisedda. På ömse sidor av diskussionen har överdrifter gjorts. EBM-anhängare har

² För frågan om huruvida bayesianer ger randomisering en särskild epistemisk roll, se exempelvis Worrall 2010a och Worrall 2007b, s. 452-454.

hävdat att EBM på ett "dramatiskt sätt förändrat hur vi tänker, förklarar och utövar medicin" (Bagshaw & Bellomo 2008). Kritiker i sin tur har hävdat att EBM "är ett kraftfullt verktyg som objektifierar världen, tillintetgör subjektet och undergräver den kliniska friheten" (Parker 2002, 273). Som vi kommer att se gör dylika representationer alltför grova antaganden om den magnitud till vilken EBM förändrat medicinsk praktik.

I synnerhet i tidigare EBM gavs evidensen onekligen en väldigt framstående roll. Evidensen, hette det, "talar sitt klara språk", och liknades till sin natur vid en "neutral, objektiv domare" som hade förmågan att "ersätta tidigare teori" (Djulbegovic et al 2009, 165). Den kunde alltså insamlas, analyseras och tolkas utan att avsevärt påverkas av betraktarens förhandsuppfattningar. I de första versionerna av EBM hänvisades det överlag inte till patienternas värderingar, och även i senare EBM har en mer ingående granskning av värderingarnas betydelse för praktiken för det mesta uteblivit. Praktiken har ställvis rentav framställts som enkel derivation, där ordet "basera" ger evidensen en roll som ett fundament varur handlingsregler enkelt kan utvinnas och tillämpas. Kvernbekk (2018, 138) citerar David Hargreaves, en tidig anhängare av evidensbaserad praktik, när han säger att pedagogiken behöver forskning som:

[...] (i) slutgiltigt demonstrerar att en förändring i praktiken från x till y leder till en signifikant och varaktig förbättring i såväl pedagogik som inläring och (ii) har utvecklat en metod som övertygar lärarna om fördelarna av, och medlen för, att byta från x till y.

Som Kvernbekk påpekar är det för det första tvivelaktigt om någon evidens över huvud taget är "slutgiltig" om man med ordet avser icke-fallibel. Fallibilism, det vill säga tesen om att den vetenskapliga kunskapen alltid är ofullständig och all evidens kan visa sig vara felaktig, har även varit en utgångspunkt i EBM (Djulbegovic et al 2009, 160). Dessutom är det inte så mycket evidensen som hypoteserna evidensen står i stöd för som utnyttjas i EBM, och som önskas ha effekt. Detta betyder

inte att en praktik inte skulle behöva evidens: evidensens funktion i en praktik är precis densamma som för hypoteser, att ge stöd. De berättiganden som används i en praktik kommer dock att inbegripa annat än enbart evidens ur kliniska studier, eftersom en praktik även bör vara berättigad ur ett moraliskt, socialt och jämställdhetsperspektiv. Det är inte självklart att den mest effektiva vårdmetoden alltid är den bästa (Louhiala & Hemilä 2005). Kliniska beslut omges av en mängd olika faktorer utöver evidensen, och något så vardagligt som diagnostiseringen och vården av mellanöroninflammation innefattar en mängd olika värdebeslut (ibid, 1323–1324). Visserligen är värderingarna och målen mellan patient och läkare allt som oftast tillräckligt lika för att inga stora värdeskillnader ska uppstå, men värderingar som inte formuleras explicit kan bli ett problem. I synnerhet offentliga åtgärders underliggande natur som politiska och värdeladdade tenderar grumlas om man kräver att alla beslut som fattas skall "basera" sig på värdefri, vetenskaplig evidens (Canali & Jukola 2020).

Ett annat sätt att granska den inneboende värdeladdningen i en praktik är genom argumentet om *induktiv risk* (Douglas 2000). Eftersom all vetenskaplig evidens löper risken av att vara felaktig kommer ställandet av kriterium för vad som är tillräcklig evidens för att (låta bli att) agera att kräva en mängd icke-epistemiska bedömningar, till exempel etiska, sociala och politiska omdömen. I EBM avses med begreppet "evidensbaserat" allt som oftast den standard där praktiken kan anses vara berättigad och representeras i evidenshierarkierna (Goldenberg 2006, 2623). Men denna standard där vissa vetenskapliga hypoteser accepteras och andra förkastas är avhängig vilken mängd osäkerhet som tolereras, och vilka följderna är ifall bedömningarna som gjorts exempelvis visar sig vara felaktiga. I medicin utgör mammografiscreeningar ett exempel. I mammografiscreeningar färgas allt från bedömningen av epidemiologisk data berörande prevalens och mortalitet, utformningen av undersökningar, tolkningen av resultat samt kommunikationen av screeningarnas för- och nackdelar av vår tolerans för felbedöm-

ningar (Plutynski 2017). Det finns inget som tyder på att risktoleransen vid olika beslut skulle vara statisk och kontextobunden, även om evidenshierarkierna är det.

I senare EBM har antalet framställningar där evidensen ses som självtillräcklig minskat. Där kliniskt omdöme i tidigare EBM representerades som sämre än – och ofta som stående i motsats till – evidens i sig, heter det i senare EBM att den bästa möjliga praktiken kräver att evidens, kliniskt omdöme och patientpreferenser integreras. Precis hur detta skall göras är dock fortfarande rätt så oklart – analysen stannar allt som oftast där. Men även om ett gemensamt, väldefinierat mål kunde identifieras kan evidensen enligt EBM aldrig på egen hand bestämma det optimala tillvägagångssättet, eftersom de flesta fall kräver en analys av de förväntade nyttorna och skadorna. Vid valet av tillvägagångssätt bör dessutom patientens preferenser och värderingar beaktas (Djulbegovic & Guyatt 2019).

Dessutom anses evidensen i EBM inte längre till sin natur vara oberoende av våra teorier och som sådan inte heller absolut värdefri. Slutsatserna som dras ur en evidens, heter det, kan bara göras i en kontextuell miljö där såväl den teoretiska förståelsen och värdegrunden är delad (ibid). En orsak till detta är att inte ens en meta-analys – som ofta anses vara en av de mest tillförlitliga metoderna i medicinsk forskning – är en statisk och mekanisk teknik, utan inbegriper värdebedömningar i alla väsentliga delar av processen (Stegenga 2011; Louhiala & Hemilä 2005, 1319–1321). Meta-analyser som utgår från samma primärundersökningar kan uppnå olika resultat, och i synnerhet undersökarens kommersiella intressen tycks skeva resultaten (Stegenga 2011, 499). Och även om själva undersökningsförloppet kunde tänkas vara utom räckhåll för forskares individuella bedömningar är det inte klart att processen kring undersökningarna kan vara det. För att ett svar skall kunna produceras och tolkas med så få värdebedömningar som möjligt, bör frågan vara specifik och sätten att besvara den på begränsade. Detta kräver dock att värderingar inkorporeras i själva frågan. Om intervention X i en undersökning visas vara effektiv för depression, är effektiviteten alltså inte enbart ett empiriskt förhållande,

utan en variabel som är avhängig den definition på depression (och följaktligen även på "normal" sorg) som implementerats i undersökningsdesignen (Stegenga 2015, 63–65). Likaså kräver en undersökning av en intervention för avvikande sexuell lust en bild av vad normal sexuell funktion utgörs av (Bueter & Jukola 2020), och ersättningsvård vid missbruk en specificering av vilken sorts tillstånd eller störning missbruk egentligen är (Uusitalo 2021).

Man kunde alltså påstå att EBM gått varvet runt. Såväl tolkningen av evidensens betydelse som dess praktiska tillämpning kräver teoretisk förståelse och delad värdegrund, eftersom evidensen "aldrig på egen hand talar sitt klara språk eller förmedlar sanningen" (Djulgovic & Guyatt 2019, 725). Den kliniska praktiken består av mer än endast tillämpning av god evidens. EBM som sådan förfäktar inte "kokboksmedicin", även om de tidigare versionerna av EBM ofta förenklade såväl hoppet från evidens till riktlinjer som hoppet från riktlinjer till praktik.

Evidens för vad?

I en praktik är det inte evidensen som sådan vi är intresserade av, utan hypoteserna evidensen fungerar som evidens för. Det är hypoteserna vi tillämpar i praktiken, och som vi önskar ha en effekt. Har denna distinktion en betydelse? Åtminstone så tillvida att en fråga som ofta blir förbisedd i diskussioner om evidens i EBM är precis vilken sorts hypoteser detta kunde vara. Hur pålitligt ett stycke evidens är kommer rimligtvis att vara avhängigt av vilken hypotes det är vi vill ha evidens för. Evidensens "kvalitet" är så att säga inte en inneboende egenskap, utan ett förhållande mellan evidensen och den hypotes den stöder. I samma anda är även probabilistiska definitioner på evidens – till exempel att god evidens höjer sannolikheten för att en viss hypotes är sann över ett visst tröskelvärde – i grund och botten inte särskilt informativa. I viss mån tycks de "sätta kärran före hästen": användningen av dem kräver att vi redan vet vad som är relevant vid bedömningen av sannolikhetsförändringen.

Vad är det då vi vill ha evidens för i EBM? Som i "medicin" överlag kan målet i EBM anses vara att som helhet främja patienters hälsa och välmående genom att instrumentellt använda sig av vetenskaplig kunskap. Således ligger det främsta intresset i EBM på evidens som stöder en kausal tolkning mellan en intervention eller ett tillvägagångssätt och en patients ökade välmående. I sin mest generella form kan en dylik kausal hypotes uttryckas i formen "om vi gör X, får vi resultat Y". Det vi vill ha evidens för är i sin mest grundläggande form evidens som stöder att X verkligen leder till Y. Eller, som det ibland uttrycks, evidens för att X "fungerar".

Låt oss för tillfället nöja oss med denna breda term, "fungera". Vad som egentligen "fungerar" är nämligen en mer komplicerad fråga än det möjligtvis ytligt sett ter sig. Om vi till exempel påstår att en intervention "fungerar" eftersom en RCT har bevisat det – vilket när allt kommer omkring är vad evidenshierarkierna tycks antyda – är det inte bara ett påstående om hur kunskap genereras. Det är också ett påstående om kausalitetens natur, om vad det betyder att något "fungerar". Vare sig vi tänker oss att RCT avslöjar kausala förhållanden, eller att de berättigar vår tolkning om kausala förhållandens existens, tycks kausaliteten till sin natur vara avhängig metodologin i RCT. Vårt svar på frågan om vad som "fungerar" är alltså inte enbart epistemologiskt, utan även ontologiskt, laddat (Kerry 2018, 83).

Låt oss granska detta närmare. Ett argument för upphöjandet av RCT är att de till skillnad från observationella studier kan skilja kausalitet från korrelation (Worrall 2007a). Om alla blandande faktorer – såväl de vi känner till som de vi inte känner till – är fördelade lika mellan undersökningsgrupperna kan vi deduktivt dra slutsatsen att skillnaden som uppstått mellan grupperna beror på interventionen, det vill säga att hypotesen att interventionen har en effekt är sann. Vad berättar en dylik argumentation om vår bild av kausalitetens natur? Här tycks kausalitet åtminstone vara något annat än korrelation. Skillnaden mellan observationella studier och RCT ligger i att RCT kan kontrollera för alla blandande faktorer, såväl kända som

okända. Den enda väsentliga skillnaden som existerar mellan grupperna i en RCT är enligt detta resonemang den kausala faktor som undersökts, nämligen interventionen ifråga. Kausaliteten fastställs i mätningen mot en annan grupp. Med andra ord är den *kontrafaktisk*; sanningsmakaren är den grupp som fått placebo och vars slutresultat skiljer sig från interventionsgruppens.

Kausala påståenden konfirmeras och förnekas dock i EBM även på andra sätt än genom RCT. Det som kallas *effektivitetsparadoxen* går ut på att flera av de mest allmänt accepterade interventioner som används i medicin – aspirin för huvudvärk, penicillin för lunginflammation, vätska mot uttorkning och heimlichmanövern (den första hjälpen-manöver genom vilken man försöker få till exempel en matbit som fastnat i luftstrupen att lossna) – aldrig har bevisats vara effektiva i RCT. Detta beror åtminstone delvis på att dessa interventioners effektivitet anses vara omöjlig att betvivla. I dylika fall kan utförandet av RCT framför allt anses vara oetiskt, eftersom det utsätter patienterna i kontrollgruppen för omotiverad risk. Detta var fallet exempelvis berörande barn med persisterande pulmonell hypertension (PPHS) och ECMO-behandling (Worrall 2008). PPHS är en sjukdom som leder till att blodet hos spädbarn syresätts dåligt, och med hjälp av en ECMO-maskin kan blodet syresättas utanför kroppen. Observationella studier visade tidigt på en väldigt klar nytta av ECMO-behandling vid PPHS, men det ansågs likväl nödvändigt att utföra RCT för att vidare utreda deras effekt. Minst fem påföljande RCT gjordes, vilket – eftersom ECMO-behandling är effektivt för PPHS – ledde till att patienterna i kontrollgrupperna inte fick bästa tillgängliga vård. Huruvida dessa påföljande RCT verkligen var nödvändiga har diskuterats. Mekanismen genom vilken ECMO-behandling verkar var känd, och flera observationella studier av hög kvalitet fanns att tillgå.

Ett annat exempel gäller homeopati. Vissa RCT och metaanalyser tycks ha "bevisat" effekten av homeopati (Vandenbroucke & de Craen 2001, 507). En framstående tes i homeopatiska behandlingar är att den gynnsamma verkan i ett boteme-

del blir större ju mer utspädd den verkande substansen i bote-medlet blir; de homeopatiska substanserna är ofta utspädda till en punkt där inte en enda aktiv molekyl nödvändigtvis finns kvar av det som tänks verka (ibid). Ur en naturvetenskaplig synvinkel finns det ingen möjlig mekanism som förklarar varför homeopati skulle vara effektivt, vilket även är orsaken till att homeopati aldrig godkänts som gängse behandling i EBM och att förnuftet i att testa dess effekt i meta-analyser från första början ifrågasatts (Vandenbroucke 1997). Avsaknaden av en trovärdig förklaring för den mekanism genom vilken den kausala faktorn verkar tycks i detta fall väga mer än den rena associationen. Att så är fallet är i linje med de nio inflytelserika "kriterier" för att bedöma existensen av en kausal verkare som Sir Austin Bradford Hill, förfadern till de moderna kliniska undersökningarna, presenterat. Av dessa relaterar fyra (temporalitet, teoretisk trovärdighet, koherens, och experimentell evidens) i synnerhet till evidens för den mekanism genom vilken den kausala faktorn verkar, och är jämställda de kriterier som relaterar till evidens för det direkta probabilistiska sambandet mellan intervention och slutfall (i synnerhet associationens styrka, följdriktighet, dos-responsförhållande) (Russo & Williamson 2007). Däremot passar det sämre in i en EBM-motiverad epistemologi, där evidens för det probabilistiska sambandet mellan intervention och slutfall enligt evidenshierarkierna anses utgöra mer trovärdig evidens än evidens för mekanismer.

Ur denna synvinkel tycks EBM i grunden arbeta med flera olika uppfattningar av vad kausalitet egentligen är och sakna en enhetlig bild av hur kausala påståenden konfirmeras och förnekas. Sådillvida är det inte heller en överraskning att man i EBM - i likhet med epidemiologisk forskning överlag - är försiktig med att dra kausala slutsatser ur data. Mekanismen genom vilken den kausala faktorn bidrar med en effekt är generellt sett underordnad kunskapen att den har en effekt. De olika verkarnas effekter presenteras enbart som en sorts ökad risk: vi kan säga att rökning ökar risken för lungcancer eller att kombinationspiller ökar risken för djupvenstromboser, till exempel. Resultaten framställs som den ökade incidensen av ett slutfall i

en grupp i jämförelse med en annan; ju robustare och mer universell korrelation, desto mer evidens för existensen av en kausal faktor.

Men även om vi i EBM kan undgå att befatta oss med kausalitet är det inte klart att vi kan göra det i klinisk medicin. Det vi är intresserade av i klinisk medicin är inte data i sig, utan av att kunna generalisera resultaten, göra prognoser och ordinera interventioner i andra populationer eller individer än den undersökta, vilket förutsätter en kausal hypotes. Låt oss granska detta med några exempel ur den kliniska praktiken. Eftersom de flesta orsakssamband i medicin är svaga på populationsnivå använder man sig i EBM av ett så kallat *number needed to treat*-tal (NNT) för att beskriva en interventions nytta. NNT beskriver enkelt hur många patienter bör vårdas med en viss intervention under en viss tid för att en person skall kunna förväntas få den undersökta nyttan. Omvänt används ett NNH-tal (*number needed to harm*) för att beskriva hur många som bör vårdas för att en person skall få ett visst negativt utfall. För selektiva serotonininhibitorer (SSRI:s), den vanligaste klassen depressionsmediciner, har NNT uppskattats vara runt 6-7 (Isometsä 2020). För statiner (kolesterolsänkande mediciner) som förhindrare av hjärtattack hos personer utan känd hjärtsjukdom har NNT uppskattats vara 217 (Abramson 2017), och för kombinationspreventivpiller som orsakare av ventrombos ligger NNH runt 1000 (Stegeman et al 2013). Så när det i rubriken till den *God Medicinsk Praxis* -artikel varur ovannämnda NNT för SSRI:s hämtats (Isometsä 2020) heter att "Effektiviteten av antidepressiva mediciner har påvisats även för patienter med depression i primärhälsovården" betyder det alltså att sex till sju patienter bör vårdas för att i medeltal en patient skall få den sökta effekten. Vad berättar detta påstående om de enskilda patienternas benägenhet? Att det är ett faktum att i medeltal en av sju personer i undersökningsgruppen fick en viss effekt betyder inte att varje enskild patients benägenhet i gruppen är en på sju - att anta detta är inte minst en form av ekologiskt felslut (Hickey et al 2013, 70). Likaså betyder inte faktumet att frekvensen för djupvenstromboser orsakade av kombinationspiller är en på tusen

användare att varje enskild användares benägenhet är en på tusen. Snarare är den i kausal mening antingen 100% eller 0%, bara vi klassificerar patienten i korrekt undergrupp (Kerry et al 2012).

Poängen är inte att de flesta läkare antar att en patients individuella benägenhet att få djupvenstromboser av kombinationspiller är en på tusen, utan att de i de flesta fall troligtvis *inte* gör det, av goda skäl. Om en intervention fungerar i en grupp eller patient men inte i en annan antar vi att det existerar ett kausalt samband som förklarar skillnaden. För att ta ett extremt exempel försöker vi inte, i det fall en patient tidigare använt sig av en sorts antidepressiv medicin utan effekt, ge henne samma medicin sju gånger om, i hopp om att den möjligtvis denna gång fungerar. Snarare antar vi att något i den kausala kedja genom vilken medicinen verkar i patientens kropp (om hon ätit medicinen) inte är densamma som hos de patienter på vilken medicinen verkat, och ordinerar en annan medicin. Det är även orsaken till att patienter med hög benägenhet för ventromboser överlag råds att undvika kombinationspiller, och att man försöker kanalisera användningen av statiner främst till patienter med hög risk för komplikationer av hjärt- och kärlsjukdomar. Kort sagt försöker vi definiera rätt undergrupp för varje enskild patient. I NNT-termer ser vi effekten av detta då gruppen som mottar statiner för att förhindra hjärtattacker omdefinieras som personer med redan känd hjärtsjukdom: då är NNT inte längre 217, utan 39 (The NNT 2013).

Denna form av härledning leder oundvikligen till att den mest relevanta undergruppen i klinisk medicin är den enskilda patienten, och att den mest trovärdiga metodologin är en *N-av-1 undersökning* (Anjum et al 2015, E11). En N-av-1 undersökning är i princip en RCT utförd på en enda patient. Genom att slumpmässigt fördela den experimentella interventionen och kontrollinterventionen på patienten av intresse kan vi komma fram till den individuella effekten hos den person vi vill vårda. Och det är denna individuella effekt vi är intresserade av i klinisk medicin, inte medelvärdeseffekten i en grupp. Att antidepressiva mediciner i medeltal "fungerar" på en patient av sju i en

viss grupp är naturligtvis inte värdelös information, men vad vi egentligen vill veta i klinisk medicin – vad vi vill ha evidens för, så att säga – är att antidepressiva mediciner ”fungerar” för just vår patient. Och denna skillnad medför, som vi kommer att se, vissa metodologiska bekymmer.

Vilka hypoteser är det vi får evidens för i kliniska undersökningar?

I det förra avsnittet slog vi fast att vi i klinisk medicin är intresserade av vad som ”fungerar” för den mest relevanta undergruppen, nämligen den enskilda patienten, och att den mest relevanta och tillförlitliga undersökningen är en N-av-1 undersökning på patienten av intresse. Detta skapar en ontologisk spänning mellan klinisk medicin och EBM, eftersom EBM anser att den mest tillförlitliga undersökningen är en RCT eller sammanslagningar av flera RCT, där resultatet är medelvärde ur en stor mängd patienter. Denna spänning uttrycker sig även som metodologiska bekymmer, eftersom ontologiska ställningstaganden inte enkelt kan skiljas från epistemologiska. I detta avsnitt granskar vi några av dessa problem genom att fokusera på RCT. För det första frågar vi oss om de resultat vi får ur RCT som sådana är överförbara på de enskilda patienter vi vill vårda. Därefter frågar vi oss på ett bredare plan vilka hypoteser det är vi egentligen besvarar – och möjligtvis underlåter att besvara – i RCT genom att diskutera skador som orsakas av medicinska interventioner.

RCT ger verkligen stark data. Cartwright har genom en probabilistisk definition på kausalitet och särskilda premisser presenterat vad hon kallar en ideal RCT, som kan isolera kausalitet (Cartwright 2007). I en ideal RCT är de två grupperna som undersöks identiska för alla faktorer förutom interventionen, det vill säga alla blandande faktorer, såväl kända som okända, är fördelade lika mellan grupperna. Vi kan deduktivt dra slutsatsen att hypotesen att interventionen har en effekt är sann.

Huruvida en dylik RCT existerar i verkligheten har debatterats (se exempelvis Worrall 2007b, Cartwright 2007, Senn 2013).

Som Worrall påpekat existerar det i verkligheten en oändlig mängd blandande faktorer, vissa av dem kända, andra okända. Randomiseringen som process är just det, slumpmässig, och inget i sig betryggat att alla väsentliga faktorer fördelas jämnt, i synnerhet som randomiseringen oftast utförs endast en gång. Egentligen vet vi inte ens när vi närmat oss en perfekt randomisering. När vi i praktiken utför en RCT är det alltså inte endast hypotesen att interventionen har en effekt som testas. Vi frågar oss även om undersökningen använde sig av rätt dos, om patienterna i undersökningen strikt följde det ordinerade programmet, om de kända blandande faktorerna var fördelade jämnt, och så vidare. Dessa stödhypoteser kan, exempelvis i det fall vi märker att den ena gruppen i grunden var sjukare än den andra, modifiera vår uppfattning om den primärt testade hypotesens sanningshalt, och vice versa. Detta kan betraktas som en variant på det som i vetenskapsfilosofin kallas *Duhem-Quine problemet*: nödvändighet att inbegripa stödhypoteser vid testandet av enskilda vetenskapliga hypoteser (Chin-Yee 2014).

Hur som helst är det värt att notera att räckvidden i en ideal RCT, även om en dylik tänks existera, är väldigt begränsad. De slutsatser vi drar kan inte vara starkare än de strikta premisserna tillåter. Strikt taget gäller det kausala påståendet endast för de kausalt homogena grupper som undersökts – inte för de enskilda personerna i dessa grupper, utan för grupperna som helhet (Cartwright 2007). Detta behöver naturligtvis inte vara ett stort problem så länge som de enskilda patienterna i grupperna är tillräckligt lika. Men det vi åtminstone måste fråga oss är huruvida grupperna är tillräckligt lika de patienter vi vill vårda, det vill säga de enskilda patienterna utanför undersökningspopulationen.

För att granska om detta är fallet i praktiken lönar det sig att återgå till orsaken till att vi kontrollerar en undersökning. Vi skrev tidigare att logiken bakom att kontrollera en undersökning är att minska på andra hypotesers sannolikhet, till exempel hypotesen att den ena gruppen var yngre och friskare än den andra, vilket gav slutresultatet. Om de två grupperna som undersöks är olika med tanke på blandande faktorer kan vi inte

fastställa att slutresultatet kausalt beror på interventionen. Alltså räcker inte heller vilken som helst undersökningspopulation till: den måste vara tillräckligt lika med tanke på faktorer vi föreställer oss att kan inverka på slutresultatet. Detta för att den *inre validiteten*, det vill säga den grad till vilken vi kan vara säkra på att den mätta effekten eller associationen i en undersökning är korrekt, skall vara hög. Så kan den inre validiteten i en RCT förbättras till exempel genom att exkludera en stor mängd patienter (detta kan också göras helt enkelt för att maximera sannolikheten av att effekten av intervention bevisas) (Stegenga 2018, 141). En betydande del potentiella deltagare kan hamna utanför en undersökning utifrån dåligt rapporterade och slumpmässigt valda exklusionskriterium (Howick et al 2013, 277). I synnerhet kvinnor, äldre personer, etniska minoriteter, mångsjuka patienter och patienter med ovanligt uttryck av en sjukdom tenderar vara underrepresenterade (Rawlins 2008, 2155). Så var fallet exempelvis i RECORD-undersökningen som gjordes för att utreda huruvida rosiglitazon, en diabetesmedicin, ökar risken för hjärtattacker (Stegenga 2016, 492). Undersökningen hade sju inklusionskriterium och sexton exklusionskriterium. Delvis som följd av detta var både kontrollgrupp och interventionsgrupp i undersökningen avsevärt friskare än motsvarande målpopulation i klinisk verksamhet: undersökningspopulationens hjärtattacksfrekvens uppskattas ha varit ungefär 40 % lägre än för motsvarande demografiska grupp i den bredare populationen.

Inte heller kontexten är oviktig: populationen i en undersökning är inte slumpmässigt vald såtillvida som medverkan i en undersökning korrelerar med psykologiska, fysiologiska och ekonomiska karakteristika. Dessutom är undersökningsmiljöer i allmänhet mer reglerade än vårdmiljön utanför undersökningar är: läkemedelsundersökningar utförs ofta på sjukhus, i specialiserade miljöer där kompliansen uppföljs och doseringen är klart och specifikt skriven (Blunt 2015, 147–148). Detta för att stödhypoteserna, exempelvis att patienterna till olika grad följer det ordinerade programmet, inte skall underminera den primärt undersökta hypotesens trovärdighet.

Alla av dessa faktorer inverkar på den *yttre validiteten*, det vill säga trovärdigheten att resultaten kan generaliseras att gälla andra populationer än den undersökta. Vi vill naturligtvis att den inre validiteten är god – i annat fall skulle vi knappast ha skäl att generalisera över huvud taget. Men beträffande RCT sker upphöjandet av den inre validiteten allt som oftast till pris av den yttre. Ju närmare vi kommer en ideal RCT, desto mindre allmängiltigt blir generellt sett det resultat vi fått. Cartwright (2007, 18) kallar detta ”utbytet” mellan inre och yttre validitet. Teorin som prövats i en RCT är alltså inte den vi är intresserade av i klinisk medicin – att interventionen fungerar för den patient vi vill administrera den åt – utan att den är effektiv i jämförelse med en kontrollintervention, oftast placebo, när den administreras på ett visst sätt för en viss sorts patienter under en viss begränsad tid. Denna skillnad behöver som utgångspunkt inte vara förödande, men den är åtminstone inte betydelslös. Ett exempel är fallet Benoxaprofen (Opren), en icke-steroidanti-inflammatorisk medicinering (NSAID) vars effekt vid ledinflammation och muskelsmärta bekräftades i RCT. Eftersom undersökningarna hade ålderskriterium på 18-65 år och Opren i klinisk praxis gavs åt en i medeltal äldre population, orsakade den i klinisk praxis lever- och njursvikt, och måste tas ur bruk (Worrall 2010a, 361).

Problemet förstärks av att frågor berörande yttre validitet historiskt sett förbisetts i EBM (Rothwell 2005, 82). De flesta evidenshierarkier tycks i grunden rangordna evidensen utifrån de olika metodernas inre validitet i förhållande till en hypotes som gäller effektiviteten av en medicinsk intervention, men detta sägs sällan explicit. Sådillvida är det nära till hands att anta att det som rangordnas i evidenshierarkier helt enkelt är den kliniska effekten, och att ingen skillnad mellan inre och yttre validitet existerar, i synnerhet som inga hierarkier med yttre validitet som explicit tolkningsegenskap existerar (Blunt 2015, 56). Ingen evidenshierarki skiljer heller på så kallade pragmatiska RCT, som (i utbyte mot sämre inre validitet) utförs på patienter som till högre grad motsvarar målgruppen i klinisk miljö, och RCT gjorda i experimentell miljö (ibid, 150). I evidenshierarkier

där den rangordnade egenskapen är inre validitet borde pragmatiska RCT som utgångspunkt placera sig lägre än RCT utförda i experimentell miljö.

Idealet för hög inre validitet medför även begränsningar på vilka frågor som kan besvaras, och på vilket sätt. Ett flervalssformulär, till exempel, har visserligen låg risk för snedvridning så tillvida som den allt som oftast ger samma resultat oavsett undersökarens värderingar och egenheter, men detta har sitt pris: en viss mängd väsentlig information kan aldrig formuleras som något så strikt som flervalssfrågor. På samma sätt kan inte RCT, precis på grund av deras strikta metodologi, svara på alla relevanta frågor vi vill ha svar på i klinisk medicin. Ur denna synvinkel finns det en reell risk att idealet för RCT med hög inre validitet i praktiken inte enbart leder till mindre allmängiltig evidens, utan att det också begränsar vilka frågor som ställs och anses vara viktiga. Vi kan, så att säga, inte bedöma vad det är vi vill veta på basis av den information vi är bra på att producera.

Det mest uppenbara fallet där frågan som besvaras i en RCT är formulerad på ett för den enskilda patienten otillfredsställande sätt är det som kallas *surrogatresultat*. De flesta sjukdomar utvecklas långsamt, och att mäta skillnader i exempelvis dödlighet eller hjärtattacker i RCT är tidskrävande och dyrt. De flesta RCT håller på i sex till tjugofyra månader, även om patienterna i verkligheten använder medicinen längre, möjligtvis livet ut (Rawlins 2008, 583). Däremot är det lättare att mäta skillnader i exempelvis LDL (*Low-Density Lipoprotein*) kolesterol, som anses utgöra en pålitlig markör för risken för att få hjärtattack. Ju lägre LDL, desto mindre risk för hjärtattack, är tanken. Men det existerar naturligtvis fortfarande en stor skillnad i att en medicin effektivt sänker kolesterolvärden och att den effektivt förhindrar hjärtattacker eller sänker dödligheten. Ett exempel kommer ur ILLUMINATE-undersökningen berörande hjärt- och kärlsjukdomar, som var tvungen att stoppas som följd av större dödlighet i interventionsgruppen, även om interventionen effektivt sänkte LDL-kolesterol (Krumholz & Lee 2008). Andra surrogatmarkörer i medicin är till exempel PSA-värden för prostatacancer, systoliskt blodtryck för kardiovaskulära

komplikationer, bentäthet för benbrott och laboratorievärdet HbA1c för diabetiska komplikationer (Twaddell 2009).

Ett annat, mer subtill exempel, där för patienten relevanta frågor riskerar förbli obesvarade i RCT gäller skador medicinska interventioner orsakar. För det första är skador till sin natur ofta oväntade, oförutsägbara och oavsiktliga, vilket gör dem svåra att mäta i RCT – RCT kan som utgångspunkt endast mäta sådana kausala förhållanden vi redan förväntar oss att existerar. Dessutom är RCT oftast designade för att ha tillräcklig statistisk kraft för att upptäcka nyttan i en intervention, ofta just till priset av förmågan att upptäcka skadorna av en intervention (Stegenga 2016, 489–494). Eftersom skador utvecklas under en lång tid och dessutom är ovanliga är RCT generellt sett för korta för att systematiskt kunna upptäcka skador. Många skador upptäcks först när interventionen är på marknaden, de flesta genom icke-kontrollerade studier (ibid, 494–495).

Såttillvida som skador är oväntade, oförutsägbara och inte länkade till indikationen för vård förlorar blindningen och randomiseringen en stor del av sina värdefulla funktioner – så att säga är det knappast nödvändigt att blinda något som redan är blint, eller randomisera något som redan är slumpmässigt (Vandenbroucke 2006). Dessutom gör de strikta kriterierna för medverkan som ofta inkorporeras i RCT att precis de personer som troligtvis är mest utsatta för skadliga effekter av interventioner inte är med i själva undersökningarna. Worrall (2010b) har noterat hur ett exklusionskriterium i ASSENT-2-undersökningen var ”vilken som helst annan sjukdom som skulle placera en patient i högre risk”. Det finns naturligtvis etiska begränsningar för att undersöka något potentiellt farligt i en RCT, men att något inte kan undersökas betyder naturligtvis inte att problemet försvinner då interventionen når marknaden.

Flera skribenter har krävt att evidens av ”lägre grad” ska tas allvarligt i beaktande vid bedömningen av riskerna för en intervention, eller att evidenshierarkin rentav bör inverteras vad skador ankommer (Osimani 2013; Vandenbroucke 2008; Stegenga 2016). Även om EBM accepterar att studiens design bör

motsvara frågan som ställts, är det som Greenhalgh (2020) noterar inte i praktiken klart vilka dessa "andra" frågor är. Exempelvis kritiserades en undersökning om ansiktsmaskers effektivitet, gjord av Royal Society i Storbritannien, som använde sig av laboratorievetenskap, matematiska modeller och policystudier av epidemiologer för att vara icke-systematisk, och för att rekommendera policybeslut i avsaknaden av en kvantitativ uppskattning om effektstorleken ur välgjorda RCT. Greenhalgh noterar att det inte är klart att en dylik fråga vare sig kan eller bör besvaras med hjälp av precis kvantifikation, eller att avsaknaden av en dylik förhindrar aktion (argumentet om induktiv risk).

Dessutom kvarstår det faktum att det existerar få, om några alls, evidenshierarkier för andra frågor än just frågor om effektivitet. Om evidenshierarkierna verkligen endast är hierarkier för frågor som gäller effektivitet ("frågor som berör terapi") och annan evidens spelar en väsentlig roll bör det åtminstone finnas riktlinjer för hur denna helhet skall sammanbindas och användas i praktiken, exempelvis i de fall då RCT inte av metodologiska orsaker ger tillfredsställande svar eller av andra skäl kan användas. Är rangordningen av evidens densamma vid vården av en enskild patient som vid policybeslut? Inkorporeras påståenden om medicinens skadeverkningar i "frågor som berör intervention"? Hur förhåller sig effektivitetshypoteser över huvud taget till andra hypoteser vid kliniskt beslutsfattande? Klinisk expertis tycks spela en väsentlig roll i detta sammanhang. Men om klinisk expertis är ett strikt följande av evidenshierarkierna är den knappast värdefull, trots att EBM explicit kräver att den integreras vid beslutsfattandet. Och om den är något mer, är det oklart vad som händer exempelvis när klinisk expertis och evidenshierarkierna kolliderar. Som verktyg är evidenshierarkierna inte gjorda för att kombinera evidens, utan pekar i riktning mot toppen. I den kontextuella miljö evidenshierarkierna befinner sig i tycks antagandet vara att samma metoder kan besvara de väsentliga frågorna berörande inte bara en interventions effektivitet, utan medicinsk praxis överlag.

Från undersökning till praktik

EBM:s epistemologiska och ontologiska ställningstaganden tycks förutsätta att kausal kunskap kan flöda från det generella till det partikulära. Mer specifikt måste information kunna flöda från en hypotetisk medelvärdespatient i en undersökningspopulation till en enskild individ utanför undersökningspopulationen. Detta gör *homogenitet* – tanken om att samma intervention kommer att orsaka liknande effekter hos olika människor – till en central utgångspunkt för EBM (Anjum 2016). Antagandet om homogenitet ligger bakom *simpel extrapolation* (simpel induktion), den metod genom vilken EBM föreslår att resultaten ur undersökningar skall generaliseras till målpopulationen:

... fråga om det finns en övertygande orsak varför studiens resultat inte borde tillämpas på patienten. Ett övertygande resultat kommer inte vanligtvis att hittas, och allt som oftast kan du generalisera till vården av din patient med tillit (Dans et al 1998, 545).

Simpel extrapolation är en kontextobunden metod, även om beslutet att extrapolera eller inte knappast är oavhängigt de följder vi förväntar oss om vi har rätt eller fel i våra slutsatser. Att notera är även åt vilket håll frågan är ställd. Det som frågas är inte om det finns trovärdiga skäl som berättigar en generalisering, utan om det finns trovärdiga skäl som tyder på att vi inte kan göra det. Men bara för att vi inte känner till en orsak till varför effektstorleken skulle vara en annan i målpopulationen kan vi inte härleda att en dylik inte existerar, speciellt som faktumet att vår kunskapsbas är begränsad var det som drev oss till komparativa kliniska studier från första början. Som Fuller (2019, 2932–2934) påpekar är det både oförnuftigt och inkonsekvent att oroa sig för okända blandande faktorer då det handlar om den inre validiteten, men inte då det handlar om den yttre validiteten.

Simpel extrapolation baserar sig i grunden på den obevisade tesen att skillnader mellan undersökningspopulation och målpopulation är sällsynta. Men representerar detta verkligheten i

klinisk medicin? I synnerhet med tanke på de strikta kriterier för medverkan som allt som oftast inkorporeras i en RCT kan detta antagande anses vara tvivelaktigt. Dessutom kvarstår faktum att få interventioner – om några alls – fungerar för alla, vilket man även inom EBM från början varit medveten om och explicit förmedlat. I denna mening uppstår problemen med generaliseringar inte endast mellan studiepopulation och målpopulation, utan även inuti en enskild studie. Ett exempel berör endarterektomi (en procedur där ett plack på blodkärlsväggen mekaniskt borttas) vid stenos av karotidartärerna (halsartärerna). Vid stenos bildas ett plack i artärväggen som kan gå sönder eller lossna och därefter orsaka störningar i hjärnans blodförsörjning – så kallade TIA:s (transitorisk ischemisk attack, en övergående störning i hjärnans blodcirkulation) eller strokes. Endarterektomi som intervention kan orsaka dödsfall, eftersom placket under proceduren riskerar lösgöras och orsaka stroke. I inledande undersökningar verkade endarterektomi som helhet öka mortaliteten med 0,5% (Howick et al 2013, 277). Däremot visade sig endarterektomi vara fördelaktigt i en undergrupp bestående av patienter med allvarlig stenos. Detta eftersom risken för stroke vid stenos av karotidartärerna stiger i förhållande med stenosens grad. Vid allvarlig stenos är risken för stroke som följd av sjukdomen större än den risk endarterektomin som procedur orsakar, medan det omvända tycks gälla för patienter med moderat eller lindrig stenos. Huruvida endarterektomi är en effektiv intervention beror alltså på stenosens grad. Att utföra endarterektomi på patienter för vilka strokerisken i samband med interventionen överstiger den som stenosen orsakar är inte önskvärt. Samtidigt representerar inte talet 0,5% alla patienter: patienter med allvarlig stenos går miste om en potentiellt effektiv intervention om endarterektomi antas vara en ineffektiv procedur som helhet.

Endarterektomifallet är ett exempel på en så kallad *paradoxal effekt*: proceduren höjer kortvarigt risken för akut stroke, men en lyckad endarterektomi minskar den långvariga risken för stroke. Andra exempel på paradoxala effekter inbegriper exempelvis beta-2-agonister, som används mot astma och kronisk

obstruktiv lungsjukdom (KOL) men som kan öka på astma- och KOL-relaterade dödsfall (Smith et al 2012, 184), och antidepressiva SSRI:s, som tycks kunna öka självmordsbenägenheten i synnerhet i början av medicineringen hos äldre personer (Juurink 2006). Ur denna synvinkel ter sig antagandet om en enda sann effekt för en intervention tvivelaktigt: vissa interventioner har flera, lika sanna effekter. Problemet kan återigen anses hänföra sig till referensklasser: varje patient tillhör flera undergrupper, och det är svårt att veta precis hur en enskild patient bör klassificeras. Speciellt svårt blir valet av referensklass i fall där sjukdomar med olika kausala mekanismer uttrycker sig på liknande sätt. Exempelvis består sjukdomen bröstcancer av flera olika tumörer som kan skiljas åt utifrån de receptorer som uttrycks i tumören i fråga. Även själva tumören kan uppvisa heterogenitet, det vill säga olika uttryck på olika områden av tumören. Olika typer av tumörer svarar olika bra på olika interventioner, och undersökningen av de molekylära och cellulära mekanismerna för denna heterogenitet som är relevanta för diagnos, prognos och terapi är därmed av stort intresse (Turashvili & Brogi 2017).

Det finns även faktorer utanför ren biologisk variation som måste tas i beaktande vid generaliseringar. Ett exempel som Cartwright (2012, 982–984) lyft fram berör en intervention stödd av Världsbanken med målet att förbättra barns närings-tillförsel i Tamil Nadu i Indien. Genom att informera mammor och ge dem antingen mat eller pengar för att handla matvaror tänktes barnens näringsstillstånd förbättras. I Tamil Nadu var interventionen en succé, men när samma metod tillämpades i Bangladesh uteblev dock resultaten. Varför? Åtminstone delvis för att mammorna i Indien och mammor i Bangladesh inte har samma uppgifter i hemmet, och för att generaliseringen baserade sig på simpel extrapolation. I Bangladesh är det inte mammorna som sköter inhandlandet av matvaror, utan papporna (åtminstone i lika hög grad som i Tamil Nadu). Dessutom är strukturen sådan att mammor som bor i gemensamma hushåll – vilket en avsevärd del gör – är underordnade sina svärmödrar

i samma hushåll. Samma sociala mekanism och kausala struktur verkar inte i Bangladesh som i Tamil Nadu. Denna form av "behaviorella effektmodifikatorer" (Victora et al 2004) existerar även i klinisk medicin. Ett exempel är kompliansproblem, som leder till över- eller underdosering. Blunt (2015, 224–225) lyfter fram så kallad megavitaminterapi, där föräldrar till barn med kronisk vitaminöverdosering tolkar de därpå följande symtomen – trötthet, illamående, anorexi – som brist på vitaminer och därmed fortsättningsvis ökar doseringen.

Generaliseringarna utgör naturligtvis inte ett problem endast för RCT. Alla undersökningar som inte utförs direkt på målpopulationen kan anses lida av samma problem. Poängen är snarare att det inte är ett nollsummespel: i verkligheten har olika former av evidens sina för- och nackdelar, och generaliseringen från undersökningspopulation till målpopulation kräver allt som oftast att dessa kombineras. Att undersöka vitaminfallen kunde tänkas kräva en pragmatisk studie eller andra kvalitativa metoder, medan den information som saknades i Bangladesh till sin natur inte kan utvinnas ur RCT, utan genom lokal information om samhället ifråga – en motsvarighet till medicinens patofysiologi eller mekanistisk härledning. På ett större plan är det inte klart att vi kan extrapolera något alls utan att förstå den kausala strukturen som orsakar slutresultatet åtminstone till viss mån.

La Caze (2008) har framställt en evidenshierarki där den explicita tolkningsegenskapen är inre validitet, och där tolkningen inte är kategorisk. Denna form av hierarki förlorar möjligheten att tala om evidensens kvalitet: för att härleda evidensens kvalitet är man tvungen att dra slutledningar som baserar sig på teori och mekanistisk härledning, som ur EBM:s perspektiv placerar sig längst ner i hierarkin. Den bästa evidensen ur EBM:s perspektiv tycks alltså, för att kunna tillämpas i verkligheten, kräva den sortens "origorösa" härledningar som drev oss till densamma från första början.

Avslutning

EBM har sina rötter i klinisk epidemiologi, där evidens utvinns genom stora populationsundersökningar och statistiska medelvärden. Således föredrar EBM empirisk evidens framom teori, och korrelation framom kausalitet. För att göra hoppet mellan medicinsk forskning och klinisk medicin, där huvudintresset ligger på enskilda patienter, antar EBM homogenitet, det vill säga att samma intervention har liknande effekt på olika människor.

I praktiken existerar flera fall där de undersökningsmetoder EBM anser vara de mest tillförlitliga inte på egen hand producerar tillfredsställande svar på de frågor som i klinisk medicin är av intresse. Problemet förstärks av att det är oklart precis hur evidenshierarkierna skall tolkas, och att ingen skillnad mellan yttre validitet och inre validitet görs. Dessutom är det oklart var gränsen mellan frågor som berör effektivitet och andra frågor går: i praktiken existerar inga andra evidenshierarkier än de som gäller frågor som berör effektiviteten av en intervention. I synnerhet de skadeverkningar interventioner orsakar utgör ett bekymmer, då evidens för skador ofta härstammar ur studier som i evidenshierarkierna anses vara opålitliga.

Flera skribenter har på senare tid yrkat på en "renässans" i EBM (Greenhalgh et al 2014), inte minst för att göra EBM mer lämpat för individuella patienter (Anjum et al 2020). Huruvida detta lyckas återstår att se. Åtminstone tycks en genuin förändring i fokus från stora populationsundersökningar till den enskilda patienten kräva mer än ytliga tillägg. I denna mening går också de problem i nuvarande EBM som i denna artikel lyfts upp möjligtvis djupare än enbart evidenshierarkier och metodologi. Den medicinska kunskapen är mångbottnad och olika metoder ger olik kausal kunskap. Utan en teori om hur och varför kausala påståenden konfirmeras och förnekas i medicin riskerar enkla representationer som evidenshierarkierna förbli

förhärskande. I en värld med allt större volym högklassig evidens ter sig detta alternativ inte som det förnuftigaste sättet att utöva och tänka medicin.

Helsingfors universitet

Tammerfors universitet

Litteratur

- Abramson, John (2017). "Statins in Persons at Low Risk of Cardiovascular Disease" *Am Fam Physician* 96(9). <https://www.aafp.org/afp/2017/1101/od1.html>
- Anjum, Rani, Kerry, Roger och Mumford, Stephen (2015). "Evidence based on what?", *J Eval Clin Pract* 2, E11-E12.
- Anjum, Rani (2016). "Evidence-based or person-centered? An ontological debate", *European Journal for Person Centered Healthcare* 4(2), 421-429.
- Anjum, Rani, Copeland, Samantha och Rocca, Elena (2020). "Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence'" *BMJ Evidence-Based Medicine* 25, 6-8.
- Bagshaw, Sean och Bellomo, Rinaldo (2008). "The need to Reform Our Assessment of Evidence from Clinical Trials: A Commentary", *Philosophy, Ethics, and Humanities in Medicine* 3, 23.
- Bluhm, Robyn och Borgerson, Kirstin (2011). "Evidence-Based Medicine", i Gifford, Fred (red.), *Handbook of the Philosophy of Science: Volume 16: Philosophy of Medicine*, 203-238. Amsterdam: Elsevier.
- Blunt, Christopher (2015). Hierarchies of evidence in evidence-based medicine. Fil. dr avhandling, The London School of Economics and Political Science (LSE).
- Blunt, Christopher (2021). "Hierarchies of evidence" (version från 10.05.2021), <http://cjblunt.com/hierarchies-evidence/>
- Bolt, Timo (2015). *A Doctor's Order: the Dutch Case of Evidence-Based Medicine*. Antwerpen/Apeldoorn: Garant.
- Bueter, Anke och Jukola, Saana (2020). "Sex, drugs, and how to deal with criticism: The case of flibanserin", i LaCaze, Adam och Osmani, Barbara (red.), *Uncertainty in Pharmacology*, 451-470. Cham: Springer.
- Canali, Stefano och Jukola, Saana (2020). "Näyttöön perustuvan pandemiapolitiikan haaste", *Tiede & Edistys* 45(3), 212-227.

- Cartwright, Nancy (2007). "Are RCTs the gold standard?", *BioSocieties* 2(1), 11–20.
- Cartwright, Nancy (2012). "Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps", *Philosophy of Science* 79, 973–989.
- Chin-Yee, Benjamin (2014). "Underdetermination in EBM", *J Eval Clin Pract* 20, 921–927.
- Clarke, Brendan (2014). "What's the difference between data and evidence", <https://blogs.kent.ac.uk/ebmplus/2014/09/19/whats-the-difference-between-data-and-evidence/> (20.8.2021).
- Daly, Jeanne (2005). *Evidence-based medicine and the search for a science of clinical care*, Berkeley/London/Los Angeles: University of California Press.
- Dans, Antonio, Dans, Leonila, Guyatt, Gordon och Richardson, Scott (1998). "Users' guides to the medical literature. XIV. How to decide on the applicability of clinical trial results to your patient", *Journal of the American Medical Association* 279(7), 545–549.
- De Vries, Raymond, Lemmens, Trudo och Bosk, Charles (2008). "The Subjectivity of Objectivity: The Social, Cultural and Political Shaping of Evidence-Based Medicine", i Bennet, Belinda, Carney, Terry och Karpin, Isabel (red.), *The Brave New World of Health*, Annandale: Federation Press.
- Djulbegovic, Benjamin, Guyatt, Gordon och Ashcroft, Richard (2009). "Epistemologic inquiries in evidence-based medicine", *Cancer Control: Journal of the Moffitt Cancer Center* 16(2), 158–168.
- Djulbegovic, Benjamin, Guyatt, Gordon och Ashcroft, Richard (2015). "Evidence-Based Medicine and the Theory of Knowledge", i Guyatt, Gordon, Rennie, Drummond, Meade, Maureen O. och Cook, Deborah J. (red.), *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*, 3rd ed. New York: McGraw Hill Medical.
- Djulbegovic, Benjamin och Guyatt, Gordon (2019). "Evidence vs Consensus in Clinical Practice Guidelines", *JAMA* 322(8), 725–726.
- Douglas, Heather (2000). "Inductive risk and values in science", *Philosophy of science* 67(4), 559–579.
- Evidence-based Medicine Working Group (1992). "Evidence-based Medicine: a New Approach to Teaching the Practice of Medicine", *Journal of the American Medical Association* 268, 2420–2425.
- Fuller, Jonathan (2019). "The myth and fallacy of simple extrapolation in medicine", *Synthese* 198, 2919–2939.

- Goldenberg, Maya (2006). "On Evidence and Evidence-Based Medicine: Lessons from the philosophy of science", *Social & Science Medicine* 62 (11), 2621–2632.
- Goodman, Kenneth (2003) *Ethics and Evidence-Based Medicine: Fallibility and Responsibility in Clinical Science*. Cambridge: Cambridge University Press.
- Greenhalgh, Trisha, Howick, Jeremy och Maskrey, Neal (2014). "Evidence based medicine: a movement in crisis?", *BMJ* 348, g3725.
- Greenhalgh, Trisha (2020). "Will COVID-19 be evidence-based medicine's nemesis?", *PLoS Med* 17(6), e1003266. <https://doi.org/10.1371/journal.pmed.1003266>
- Guyatt, Gordon, Oxman, Andrew, Vist, Gunn, Kunz, Regina, Falck-Ytter, Yngve, Alonso-Coello, Pablo och Schünemann, Holger (2008a). "GRADE: an emerging consensus on rating quality of evidence and strength of recommendations", *BMJ* 336, 924–926.
- Guyatt, Gordon, Haynes, Brian, Jaeschke, Roman, Meade, Maureen O., Wilson, Mark, Montori, Victor och Richardson, Scott (2008b). "The Philosophy of Evidence-Based Medicine", i Guyatt, Gordon, Rennie, Drummond, Meade, Maureen O. och Cook, Deborah J. (red.), *Users' Guides to the Medical Literature*, 9–16. New York: McGraw Hill Medical.
- Hammer, Gaël, du Prel, Jean-Baptist och Blettner Maria (2009). "Avoiding bias in observational studies: part 8 in a series of articles on evaluation of scientific publications", *Deutsches Ärzteblatt International* 106(41), 664–668.
- Hickey, Stephen, Hickey, Andrew och Noriega, Leonardo (2013). "The failure of evidence-based medicine?", *European Journal for Person Centered Healthcare* 1(1), 69–79.
- Howick, Jeremy, Glasziou, Paul och Aronson, Jeffrey (2009). "The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute?", *Journal of the Royal Society of Medicine* 102(5), 186–194.
- Howick, Jeremy (2011). *The philosophy of evidence-based medicine*. Chichester, West Sussex: Wiley.
- Howick, Jeremy, Glasziou, Paul och Aronson, Jeffrey (2013). "Problems with using mechanisms to solve the problem of extrapolation" *Theoretical medicine and bioethics* 34, 275–291.
- Isometsä, Erkki (2020). "Masennuslääkkeet perusterveydenhuollossa". *God medicinsk Praxis, Duodecim. Näytönastekatsaus; nak04329*. <https://www.kaypahoito.fi/nak04329> (26.08.2021).

- Juurink, David, Mamdani, Muhammad, Kopp, Alexander och Redelmeier, Donald (2006). "The risk of suicide with selective serotonin reuptake inhibitors in the elderly", *Am J Psychiatry* 163(5), 813–821.
- Kelly, Thomas (2016). "Evidence", *The Stanford Encyclopedia of Philosophy* (version från vintern 2016), red. E. N. Zalta. <https://plato.stanford.edu/archives/win2016/entries/evidence>.
- Kerry, Roger, Eriksen, Thor Eirik, Lie, Svein Anders Noer, Mumford, Stephen och Anjum, Rani (2012). "Causation and evidence-based practice: An ontological review", *J Eval Clin Pract* 18(5), 1006–1012.
- Kerry, Roger (2018). "Reconceptualising causation in evidence-based physiotherapy", i Gibson, Barbara, Nicholls, David, Setchell, Jenny och Synne Groven, Karen (red.), *Manipulating practices: A critical physiotherapy reader*, 77–100. Oslo: Cappelen Damm Akademisk.
- Knaapen, Loes (2014). "Evidence-Based Medicine or Cookbook Medicine? Addressing Concerns over the Standardization of Care", *Sociology Compass* 8, 823–836.
- Krumholz, Harlan och Lee, Thomas (2008). "Redefining Quality - Implications of Recent Clinical Trials", *N Engl J Med* 358, 2537–2539.
- Kvernbekk, Tone (2018). "Evidensbasert pedagogisk praksis: Utvalgte kontroverser", *Nordisk tidsskrift for pedagogikk og kritikk* 4, 136–153.
- La Caze, Adam (2008). "Evidence based medicine can't be...", *Social Epistemology* 22(4), 353–370.
- Le Fanu, James (1999). *The rise and fall of modern medicine*. London: Little, Brown & Co.
- Louhiala, Pekka och Hemilä, Harri (2005) "Näyttöön perustuva lääketiede: hyvä renki mutta huono isäntä", *Duodecim* 121(12), 1317–1325.
- Martini, Carlo (2020). "What 'Evidence' in Evidence-Based Medicine?", *Topoi* 40, 299–305.
- Murad, M. Hassan, Asi, Noori och Alsawas, Mouaz och Alahbad, Fares (2016). "New evidence pyramid" *BMJ Evidence-Based Medicine* 21, 125–127.
- Osimani, Barbara (2013). "Until RCT proven? On the asymmetry of evidence requirements for risk assessment", *Journal of evaluation in clinical practice* 19, 454–462.
- Parker, Malcolm (2002). "Whither Our Art? Clinical Wisdom and Evidence-based Medicine." *Medicine, Health Care and Philosophy* 5, 273–280.

- Plutynski, Anya (2017). "Safe or Sorry? Cancer Screening and Inductive Risk", i Elliot, Kevin och Richards, Ted (red.), *Exploring Inductive Risk: Case Studies of Values in Science*, 149-169.
- Rawlins, Michael (2008). "De testimonio: on the evidence for decisions about the use of therapeutic interventions", *The Lancet* 372(9656), 2152-2161.
- Rothwell, Peter (2005). "External validity of randomized controlled trials: "To whom do the results of this trial apply"", *The Lancet*, 365(9453), 82-93.
- Russo, Federica och Williamson, Jon (2007). "Interpreting Causality in the Health Sciences", *International Studies in the Philosophy of Science* 21(2), 157-170.
- Sackett, David, Rosenberg, William, Gray, Muir, Haynes, Brian och Richardson, W. Scott (1996). "Evidence based medicine: What it is and what it isn't", *British Medical Journal* 312(7023), 71-72.
- Sackett, David, Straus, Sharon, Richardson, W. Scott, Rosenberg, William och Haynes, R. Brian (2000). *Evidence-based Medicine: How to Practice and Teach EBM*, 2nd edn. Edinburgh: Churchill Livingstone.
- Senn, Stephen (2013). "Seven myths of randomization in clinical trials", *Statist. Med.* 32, 1439-1450.
- Smith, Silas, Hauben, Manfred och Aronson, Jeffrey (2012). "Paradoxical and Bidirectional Drug Effects" *Drug Safety* 35, 173-189.
- Stegeman, Bernardine, de Bastos, Marcos, Rosendaal, Frits, van Hylckama Vlieg, Astrid, Helmerhorst, Frans, Stijnen, Theo och Dekkers, Olaf (2013). "Different combined oral contraceptives and the risk of venous thrombosis: systematic review and network meta-analysis", *BMJ* 347, f5298.
- Stegenga, Jacob (2011). "Is meta-analysis the platinum standard of evidence?", *Studies in history and philosophy of science part C: Studies in history and philosophy of biological and biomedical sciences* 42(4), 497-507.
- Stegenga, Jacob (2015). "Measuring effectiveness", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 54, 62-71.
- Stegenga, Jacob (2016). "Hollow Hunt for Harms" *Perspectives on Science* 24(5), 481-504.
- Stegenga, Jacob (2018). *Medical nihilism*. Oxford: Oxford University Press.

- Straus, Sharon, Richardson, W. Scott, Glasziou, Paul och Haynes, R. Brian (2005). *Evidence-based Medicine: How to Practice and Teach EBM*. Toronto: Elsevier.
- The NNT (2013). "Statins given for 5 years for heart disease prevention (with known heart disease)", < <https://www.thennt.com/nnt/statins-for-heart-disease-prevention-with-known-heart-disease/> > (29.08.2021).
- Turashvili, Gulisa och Brogi, Edi (2017). "Tumor Heterogeneity in Breast Cancer", *Front Med (Lausanne)* 4, 227.
- Twaddell, Scott (2009). "Surrogate outcome markers in research and clinical practice", *Australian Prescriber* 32, 47–50.
- Uusitalo, Susanne (2021). "Eettisiä näkökulmia päihderiippuvuudesta toipumisen tutkimukseen ja sen vaikutuksiin", *Niin & Näin* 2, 105–115.
- Vandenbroucke, Jan (1996). "Evidence-based Medicine and 'Médecine d'Observation'", *Journal of Clinical Epidemiology* 49, 1335–1338.
- Vandenbroucke, Jan (1997). "Homoeopathy trials: going nowhere", *Lancet* 350(9081), 824.
- Vandenbroucke, Jan och de Craen, Anton (2001). "Alternative medicine: a "mirror image" for scientific reasoning in conventional medicine" *Ann Intern Med* 135(7), 507–513.
- Vandenbroucke, Jan (2005). "Homeopathy and 'the growth of truth'", *Lancet* 366(9487), 691–692.
- Vandenbroucke, Jan (2006). "What is the best evidence for determining harms of medical treatment?", *CMAJ* 174(5), 645–646.
- Vandenbroucke, Jan (2008). "Observational research, randomised trials, and two views of medical science", *PLoS Med* 5(3), e67.
- Victoria, Cesar, Habicht, Jean-Pierre och Bryce, Jennifer (2004). "Evidence-Based Public Health: Moving Beyond Randomized Trials", *American Journal of Public Health* 94, 400–405.
- Witkin, Stanley och Harrison, W. David (2001). "Whose Evidence and for What Purpose?", *Social Work* 46(4), 293–296.
- Worrall, John (2007a). "Evidence in Medicine and Evidence-Based Medicine", *Philosophy Compass* 2(6), 981–1022.
- Worrall, John (2007b). "Why There's No Cause to Randomize", *The British Journal for the Philosophy of Science* 58(3), 451–488.
- Worrall, John (2008). Evidence and Ethics in Medicine. *Perspectives in Biology and Medicine* 51(3), 418–431.
- Worrall, John (2010a). "Evidence: philosophy of science meets medicine", *Journal of Evaluation in Clinical Practice* 16(2), 356–362.

Worrall, John (2010b). "Do we need some large, simple randomized trials in medicine?", i Suárez, Mauricio, Dorato, Mauro och Rédei, Miklós, (red.) *Epsa Philosophical Issues in the Sciences: Launch of the European Philosophy of Science Association*, 289-302. London: Springer.

