

Medisinsk kunnskap før og nå: Fra teori til systematiske oversikter

Atle Fretheim^{1,2}

Seksjon for global helse, Nasjonalt kunnskapssenter for helsetjenesten

2) Institutt for helse og samfunn, Universitetet i Oslo

E-post: atle.fretheim@kunnskapssenteret.no Telefon: 91 64 98 28

SAMMENDRAG

Teorien om balansen mellom de fire kroppsvæskene – humoralpatologien – dominerte sykdomsforståelse og dermed også medisinsk behandling i 2000 år. Brekkmidler, årelating og avføringsmidler var hyppig brukte virkemidler for å rette opp i ”ubalanse” mellom kroppsvæskene, som var regnet som årsak til sykdom.

Det var ikke før på 1700-tallet at faktisk empirisk kunnskap ble etterspurt som begrunnelse for valg av behandlingsmetoder. Det å telle opp hvordan det gikk med pasienter som fikk en behandling, og sammenlikne med pasienter som ikke fikk behandlingen, var en ny måte å presentere behandlingseffekter på. Fra 1800-tallet av er det mange eksempler på at behandlingsmetoder ble prøvet ut ved å la annenhver pasient motta behandlingen.

Streptomycinstudien til UK Medical Research Council (1946) regnes for å være det første randomiserte kontrollerte forsøket innen medisinsk forskning, og for å ha satt standarden for hvordan kliniske utprøvinger bør utføres.

I boka ”Effectiveness and Efficiency. Random Reflections on Health Services” (1972) kritiserte Archie Cochrane mangelen på dokumentasjon for mange av tiltakene som utføres i helsetjenesten. Han ønsket seg langt flere randomiserte studier. Dessuten tok han til orde for at alle randomiserte studier burde samles, og at resultatene burde oppsummeres jevnlig. Iain Chalmers tok utfordringen og påbegynte et omfattende arbeid med å utarbeide en database over randomiserte kontrollerte forsøk. Med utgangspunkt i denne databasen ble det så utarbeidet systematiske oversikter over gjeldende kunnskap om effekt av tiltak i helsetjenesten. Dette ble starten på Cochrane-samarbeidet.

Fretheim A. **Medical knowledge then and now: From theory to systematic reviews.**

Nor J Epidemiol 2013; 23 (2): 113-118.

ENGLISH SUMMARY

The humoral medicine-theory dominated the understanding of diseases and thus also medical treatments over 2,000 years. Purgatives, bloodletting and laxatives were frequently used to correct ”imbalances” between body fluids, which were seen as the cause of disease.

It was not until the 18th century that actual empirical knowledge was seen as important as a basis for making decisions about choice of treatment. At this time, the idea of counting the outcomes among patients receiving a treatment and compare with patients who did not, was a new approach for presenting treatment effects. As of the 19th century there are many examples of treatments being tested by the use of alternation, i.e. allocating every other patient to the treatment under evaluation.

The UK Medical Research Council’s streptomycin-trial (1946) is regarded as the first randomised controlled trial in medicine, and for having set the standard for clinical trials.

In his book, ”Effectiveness and Efficiency. Random Reflections on Health Services” (1972), Archie Cochrane criticised the lack of evidence for many of the interventions being performed in the health services. He called for far more randomised trials. He also proposed that findings from randomised trials should be collated, and that the results should be summarised on a regular basis. Iain Chalmers took the challenge and engaged in the huge task of developing a database of randomised controlled trials. With the database as a starting point, several systematic reviews of current evidence on the effects of interventions in the health services were produced. This was the start of the Cochrane Collaboration.

This is an open access article distributed under the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gode beslutninger forutsetter god kunnskap. ”Kunnskapsbasert medisin” dreier seg nettopp om bruk av best mulig kunnskap når beslutninger tas. Ofte er det snakk om hva en skal velge å gjøre for å håndtere et problem: Er det tiltak A eller tiltak B som vil føre til det beste resultatet?

I mange tilfeller er det sprikende oppfatninger om hvilken løsning som er å foretrekke. Dette kan for

eksempel komme til uttrykk ved at én sykdom blir behandlet på forskjellig vis fra ett sykehus til et annet. Dersom to sykehus gjennomgående behandler pasienter med samme sykdomsbilde på hver sin måte, må nødvendigvis det ene sykehuset gi et bedre behandlingstilbud enn på det andre. Unntaket må være dersom de to behandlingsmåtene fører til like gode resultater, i gjennomsnitt. Dette er nettopp ett av de evige spørs-

målene i klinisk medisin, og i mange andre sammenhenger i helsetjenesten: Blant flere konkurrerende tiltak, hvilket er det beste?

FRA BREKKMIDLER TIL ELEKTRISK APPARATUR

Historisk sett har medisinsk behandling stort sett vært begrunnet med teoretiske argumenter. Ved å ta utgangspunkt i de rådende teorier har logiske argumenter blitt brukt som støtte for de behandlingstiltakene som ble foreslått. Tidens mest innflytelsesrike teori om sykdomsforståelse er utvilsomt humoralpatologien – læren om de fire kroppsvæskene og at sykdom skyldes ubalanse mellom disse (1). Brekkmidler, årelating og avføringsmidler var hyppig brukte virkemidler for å opprette ”balanse” mellom kroppsvæskene i de 2000 årene som humoralpatologien dominerte sykdomsforståelsen og dermed også medisinsk behandling.

Med opplysningstiden endret det medisinske kunnskapsbildet seg radikalt, og økt kunnskap innen anatomi, fysiologi, mikrobiologi osv. førte til stor begeistring. Det fant også sted en rivende utvikling innen kjemifaget – en rekke stoffer kunne nå fremstilles i laboratoriet – og etter hvert ble både elektrisitet og røntgenstråler tilgjengelig. De medisinske tidsskriftene gav betydelig spalteplass til entusiastiske beskrivelser av nye behandlingsmetoder. Den teoretiske begrunnelsen for metoden ble som regel tillagt stor vekt, gjerne supplert med beskrivelser av enkeltpasienter eller mindre grupper av pasienter som – etter sigende – hadde svært god nytte av den nye behandlingen. En god illustrasjon er de mange aparte elektriske instrumentene som ble benyttet mot slutten av 1800-tallet, for å tilføre kroppen – gjerne også sinnet – energi og vitalitet.

Historikerne har funnet få eksempler fra før opplysningstiden på at leger eller andre etterspurte faktisk empirisk kunnskap om virkningene av behandlingsmetoder. Men fra 1700-tallet av begynte ideen om bruk av tall i medisinsk forskning å bre om seg: Stadig flere argumenterte for at det å foreta en opptelling av hvordan det gikk med pasientene var nødvendig for å kunne si noe sikkert om virkningene av et tiltak.

Ett tidlig eksempel gjaldt behandling av febersykdom i tropiske strøk, som var et betydelig problem for britiske kolonialister. Årelating og brekningsmidler var standardbehandlingen, men bruk av kinabark (som inneholder malariamiddelet kinin) var et nytt behandlingsprinsipp som flere tok til orde for. Den skotske legen John Millar undersøkte dette nærmere ved å sette opp en tabell som viste overlevelsen blant pasienter som hadde fått standardbehandling sammenliknet med dem som hadde fått kinabark. Slike enkle regnestykker, hevdet han, er den mest overbevisende form for dokumentasjon om virkningene av en behandlingsmetode: “the stubborn evidence of arithmetical demonstration could not be shaken by argument” (2).

Et annet eksempel: Da den tidlige varianten av koppevaksinering (”variolisering”) kom til Europa i

1720-årene – etter å ha vært brukt i århundrer i Asia – var det stor uenighet om dette var verdt å satse på. Særlig ble det stilt spørsmålstegn ved om metoden var trygg, ettersom de aller fleste ble syke og noen til og med døde som følge av prosedyren. Noen foreslo da å telle opp andelen døde blant dem som ble vaksinert og sammenlikne med hvor mange som døde blant dem som fikk kopper på naturlig vis. Tallenes tale var meget overbevisende, og mange som var skeptiske til variolisering snudde. Dette er ett av de tidligste eksemplene på at en medisinsk uenighet langt på vei ble avklart gjennom bruk av statistikk (tall) (2).

I dag tas det som en selvfølge at påstander om effekt begrunnes kvantitativt, men den gang var dette en ny ide, og representerte en betydelig trussel mot det medisinske establishment. Veletablerte teoretiske forklaringer kunne nå bli utfordret, til og med forkastet, på empirisk grunnlag.

Selv om empiriske metoder og bruk av tall dominerer medisinsk vitenskap i dag, er det fremdeles spenninger knyttet til forholdet mellom teori og empiri. Det kan hevdes at teoriutvikling og en dypere forståelse av sykdom og behandlingsmetoder får for lite oppmerksomhet ettersom dagens medisinske forskning i hovedsak dreier seg om å bruke statistiske metoder for å påvise mer eller mindre viktige ”signifikante” forskjeller. Og selv om en behandlingsmetode gir en viss gevinst i gjennomsnitt, mangler ofte teoretisk innsikt om hvordan eller hvorfor behandlingen virker. I noen tilfeller kan mangel på teoretisk forståelse gjøre det vanskelig å applisere empirisk kunnskap på den enkelte pasient – gjennomsnittspasienten finnes tross alt ikke i virkeligheten.

Their Ages.	Persons inoculated.	Had the Small-Pox by Inoculation.	Had an Imperfect small pox.	Had no Effect.	Suspected to have died of Inoculation.
From 9 months to 2 years old.	06	06	00	00	00
2 to 5	14	14	00	00	00
5 to 10	16	16	00	00	00
10 to 15	29	29	00	00	00
15 to 20	51	51	00	00	01
20 to 30	62	60	00	02	01
30 to 40	44	42	00	02	01
40 to 50	08	07	00	01	00
50 to 60	07	06	00	01	02
60 to 67	07	07	00	00	01
Total Inoculated by Drs. Roby and Thompson in Roxbury and Cambridge.	244	238	00	06	06
Total	280	274	00	06	06

Bilde 1. Tabell som viser lav dødelighet blant dem som ble vaksinert mot kopper (Z. Boylston, 1726). Hentet fra James Lind Library.

PÅ VEI MOT MODERNE KLINISK FORSKNING

Bruk av kvantitative sammenlikninger bredte om seg, men det tok ikke lang tid før noen av utfordringene med denne type analyser ble erkjent. Motstanderne av variolisering viste for eksempel til at gravide og enkelte andre grupper ikke ble vaksinert, og dermed ble det misvisende å sammenlikne dødelighet blant dem som ble vaksinert med dem som ikke ble det. Dessuten var aldersfordelingen forskjellig: De som ble vaksinert var yngre enn de som fikk sykdommen på naturlig måte, ble det hevdet (2). For å ta høyde for slike skeivheter begynte en tidlig å gjøre separate sammenlikninger for forskjellige aldersgrupper.

Dette problemet er like aktuelt den dag i dag: En av de største utfordringene epidemiologer står overfor er nettopp hvordan en best kan justere for systematiske skeivheter mellom eksponerte og ikke-eksponerte grupper (seleksjonsbias), når en forsøker å trekke ut kunnskap om årsak og virkning fra ikke-eksperimentelle studier.

Nødvendigheten av å sammenlikne likt med likt ble, blant andre, påpekt av Isaac Massey i 1723: "To form a just Comparison, and calculate right in this Case, the Circumstances of the Patients, must and ought to be as near as may be on a Par" (3).

Det at sammenlikningene ofte ble gjort retrospektivt gjorde det vanskelig å vurdere hvor sammenliknbare gruppene faktisk var. Eksperimentelle metoder kan langt på vei løse dette problemet, noe som også ble erkjent av mange på 1700-tallet. James Linds eksperiment for behandling av skjorbuk i 1747 regnes som et av de første eksemplene på en klinisk utprøving. Han hadde 12 sjømenn med skjorbuk ombord på skipet der han var skipslege, og han prøvet ut seks forskjellige behandlingsmetoder ved å dele mennene inn i grupper på to. Lind beskriver mennene som omtrent like syke, men hvordan han foretok selve fordelingen til de forskjellige behandlingene er ikke kjent (3).

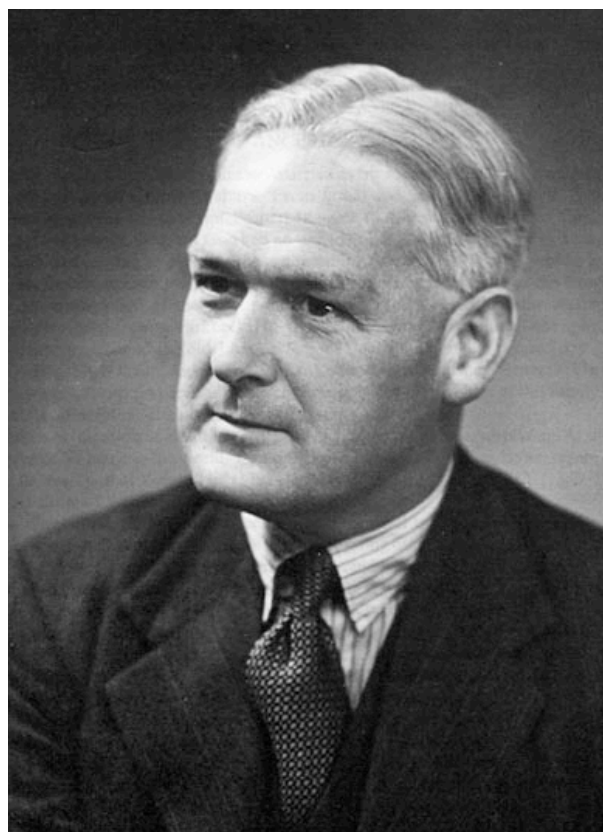
I løpet av 1800-tallet er det mange eksempler på at leger testet ut behandlingsprinsipper ved å la annenhver pasient motta samme behandling, mens de resterende fikk en annen eller ingen behandling. Thomas Graham Balfour brukte denne metoden for å studere om bruk av belladonnaurten kunne forebygge skarlagensfeber. Guttene på barnehjemmet der Balfour arbeidet, deltok i studien. For å avgjøre hvem som skulle ha den aktive behandlingen valgte han annenhver gutt på lista over alle som bodde på barnehjemmet. Han begrunnet framgangsmåten med at han ønsket å unngå seleksjons skeivheter ("to avoid the imputation of selection") (3).

Annenhver-metoden (alternate allocation) var nærmest standard framgangsmåte på begynnelsen av 1900-tallet, og over hele verden ble slike studier utført for å undersøke virkninger av tiltak (3). Forskere i India, for eksempel, utførte en hel serie studier av forskjellige typer serum for behandling av pest, som herjet datidens Bombay (3). I et brev til *The Lancet* i 1900 la

den indiske legen Nasserwanji Hormusji Choksy stor vekt på betydningen av stringent metodikk: "the cases were to be taken for treatment *alternatively* in the order of admission and *without any attempt at exclusion of any sort.*" (4).

STREPTOMYCINFORSØKET, 1946

Det britiske Medical Research Council (MRC) gjennomførte en studie i 1931-1933 for å evaluere virkningene av anti-pneumokokk serum til behandling av lobær pneumoni. I alt 773 pasienter deltok i studien som benyttet den vanlige fordelingsmetoden: annenhver pasient som kom til sykehusene fikk behandling med injeksjon av serum – de andre ikke (5). En av MRCs statistikere, Austin Bradford Hill, bekymret seg over hvor forskjellige de to gruppene endte opp med å bli. Det var, for eksempel, langt flere unge pasienter blant dem som ble behandlet med serum enn blant dem som ikke fikk behandlingen (6). Bradford Hill mente at fordelingsmetoden hadde sviktet, og at dette måtte skyldes at prinsippet om å gi behandling til annenhver pasient ikke var fulgt til punkt og prikke. Det at alle involverte parter visste hvilken behandling neste pasient ville få, gjorde det mulig – bevisst eller ubevisst – å manipulere fordelingen i studien. For eksempel, dersom forrige pasient fikk serumbehandling, visste alle at neste pasient ikke skulle få behandlingen. Hvis legen mente at den neste pasienten burde få slik behandling,



Bilde 2. Austin Bradford Hill (1897–1991). Hentet fra James Lind Library.

kunne han definere pasienten til ikke å oppfylle kriteriene for deltakelse i studien. Dersom dette skjedde flere ganger ville det føre til forskjeller mellom gruppene. Dermed ville det blir svært vanskelig å tolke resultatene: Skyldtes den observerte effekten selve behandlingen, eller var den et resultat av feilkilder på grunn av skeivfordelingen av pasienter?

Bradford Hill var oppsatt på å forhindre dette problemet da MRC noen år senere planla utprøvingen av streptomycin for behandling av tuberkulose. Han mente det avgjørende for å sikre så sammenliknbare grupper som mulig var å forhindre at den eller de som meldte pasienter inn i studien ikke visste hvilken gruppe den neste pasienten ville havne i. Løsningen Bradford Hill foreslo var å lage en liste med tall i tilfeldig rekkefølge, som så ble brukt til å bestemme fordelingsrekkefølgen – hvem som skulle og hvem som ikke skulle få streptomycin (7). Dessuten – og det var det viktigste – måtte denne rekkefølgen holdes skjult for alle involverte parter. Sykehusene måtte derfor, for hver pasient som oppfylte kriteriene for deltakelse i studien, melde pasienten inn via telefon til studiesekretariatet. Der lå koden til som bestemte hva slags behandling pasienten skulle få: streptomycin eller ordinær behandling (sengeleie) (8).

Det var stor etterspørsel etter streptomycin i Storbritannia, men lite å oppdrive. Den amerikanske produsenten solgte 50 kg til de britiske myndighetene, som videresendte nesten alt til MRC for bruk i studien. Dermed lå det godt til rette for stor oppslutning om deltakelse i studien, og for at legemiddelet skulle bli fordelt til pasientene på en så rettferdig måte som mulig – hvilket var nettopp det Bradford Hill søkte å oppnå gjennom randomisering.

Streptomycinstudien regnes for å være det første randomiserte kontrollerte forsøket innen medisinsk forskning, og for å ha satt standarden for hvordan kliniske utprøvinger bør utføres. Men det finnes enkeltstående tidligere eksempler på bruk av randomisering innenfor andre fagfelt, blant annet utdanningsforskning (9). Uttrykket "randomized" ble antakelig først brukt av statistikeren Ronald Fisher i hans artikkel fra 1926 om eksperimentell landbruksforskning (10). Fishers begrunnelse for randomisering var fundert på statistiske argumenter, mens Bradford Hills begrunnelse var langt mer praktisk og enkel: Randomisering med skjult fordelingsnøkkel er den enkleste måten for å sikre tilfeldig fordeling av pasienter til de to gruppene som skal sammenliknes (7).

TALIDOMIDSKANDALEN

I løpet av 1950- og 1960-tallet ble det gjennomført flere hundre randomiserte forsøk innen klinisk medisin. I 1970 innførte de amerikanske legemiddelmyndighetene (FDA) nye, strengere krav til dokumentasjon for legemidler. Det reviderte regelverket innebar i praksis at alle nye legemidler måtte gjennom utprøving i randomiserte kontrollerte studier for å bli godkjent for bruk i USA. Bakgrunnen for dette var blant annet

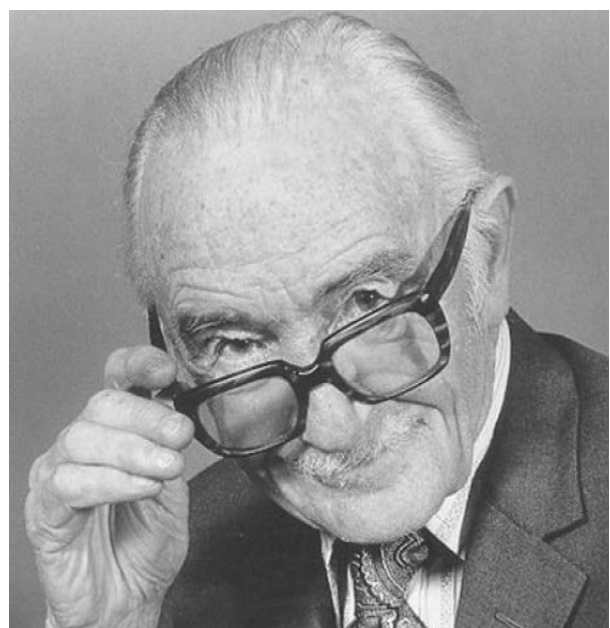
talidomidskandalen: I 1961 kom de første rapportene om en mulig sammenheng mellom mors bruk av legemiddelet talidomid under svangerskapet, og medfødte misdannelser. Tusenvis av barn viste seg å være rammet. Talidomidskandalen førte til at dokumentasjonskravene for nye legemidler ble kraftig innskjerpet over hele verden (11).

I boka "Effectiveness and Efficiency. Random Reflections on Health Services" (1972) presenterte briter Archie Cochrane en kraftfull kritikk av mangelen på dokumentasjon for en stor del av tiltakene i helsetjenesten (12). Han tok til orde for bruk av randomiserte kontrollerte studier til evaluering av både kliniske tiltak og organisatoriske endringer. Cochranes bok hadde stor innflytelse og regnes som en medvirkende årsak til at randomisering stadig oftere ble benyttet i evalueringsstudier, spesielt i Storbritannia.

Selv om det er de tradisjonelle legemiddelutprøvingene som ofte får mest oppmerksomhet, benyttes randomiserte kontrollerte forsøk til å evaluere effekt av tiltak på en rekke områder innen helsefagene. For eksempel, databasen Physiotherapy Evidence Database (PEDro) inneholder nå nesten 20 000 randomiserte studier om virkningene av tiltak innen fysioterapi.

SYSTEMATISKE OVERSIKTER

Men det Archie Cochrane kanskje er mest kjent for er at han tidlig så behovet for en oversikt over den store mengden kunnskap som akkumulerte som følge av de stadig flere randomiserte forsøkene som ble utført: "It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials" (13).



Bilde 3. Archie Cochrane (1908–1988). Gjengitt med tillatelse fra Cardiff University Library, Cochrane Archive, University Hospital Llandough.

Blant dem som tok utfordringen fra Cochrane på alvor var briten Iain Chalmers. Han og noen kolleger satte seg fore å samle inn og holde oversikt over alle utførte og pågående randomiserte forsøk innen obstetrik og nyfødtdisin (14). I 1989 kunne de, etter flere års arbeide med å identifisere studier, presentere "The Oxford Database of Perinatal Trials", som ble utgitt på disketter og oppdatert to ganger i året. Basert på studier fra dette registeret publiserte de flere systematiske oversiktsartikler. For eksempel kunne de, på basis av resultater fra 12 studier, vise at det å gi kortikosteroider til mor ved truende for tidlig fødsel bidro både til redusert sykkelighet og dødelighet hos for tidlig fødte barn (15). Det hadde ikke vært mulig å konkludere sikkert om dette basert på resultatene fra hver av de enkelte studiene som var utført, men da alle studiene ble slått sammen i én kvantitativ analyse – såkalt meta-analyse – ble effekten av behandlingen åpenbar.

Begrepet "meta-analyse" ble introdusert av Gene Glass noen år tidligere, i 1976 (16). I følge han selv var det personlige grunner til at han begynte å interessere seg for å slå sammen resultatene fra mange studier i én analyse. Det var nemlig en økende skepsis til nytten av psykoterapeutisk behandling på denne tida, blant annet basert på noen publikasjoner der forskning på psykoterapi var oppsummert. Glass hadde selv gått flere år til psykoterapeutisk behandling og opplevd det som svært nyttig. Dessuten mente han at de oppsummeringene som var gjort var preget av "arbitrariness, idiosyncrasy and high-handed dismissiveness" (17). Han bestemte seg for å gjøre jobben bedre selv. Sammen med en kollega forsøkte han å finne så mange relevante studier som mulig, og å slå resultatene fra disse sammen på en matematisk robust måte. Til hans tilfredsstillelse var resultatene ganske entydig i favør av psykoterapi (18).

Iain Chalmers klarte å overbevise de britiske helsemyndighetene om behovet for å samle randomiserte studier for alle typer tiltak i helsetjenesten, og å bruke en slik database for å utarbeide systematiske oversikter om effekt av behandlingsmetoder i helsetjenesten. Da et nytt senter for dette arbeidet ble lagt til Universitetet i Oxford, fikk det navnet The Cochrane Centre (19). Det var stor interesse for dette også utenfor Storbritannia, og innen et år var det etablert Cochrane-sentre i Canada, Danmark (The Nordic Cochrane Centre) og i USA. Den første internasjonale Cochrane-konferansen ble avholdt i 1993, og her ble The Cochrane Collaboration (Cochrane-samarbeidet) offisielt lansert. Gjennom 1990-tallet ble det opprettet rundt to nye Cochrane-sentre hvert år, og antallet systematiske oversikter publisert i The Cochrane Library steg hurtig. Den norske forgreningen av det nordiske Cochrane-senteret, ligger i Nasjonalt kunnskapssenter for helsetjenesten.

MEDISINSK METODEVURDERING

I løpet av 1990-tallet etablerte mange land sentre for metodevurdering (Health Technology Assessment – HTA). Disse miljøene har som hovedoppgave å bistå helsemyndighetene med dokumentasjon i beslutningsprosesser om bruk av teknologi (20). Det kan for eksempel være spørsmål om refusjon av et legemiddel, eller hvorvidt et nytt teknisk hjelpemiddel bør tas i bruk. Nødvendigheten av å innhente, vurdere og analysere all eksisterende forskning om effektene av den medisinske eller helsefaglige metoden som vurderes, har vært erkjent i HTA-miljøene fra første stund. Systematiske oversikter har derfor en helt sentral plass i metodevurdering. Nasjonalt kunnskapssenter for helsetjenesten har ansvar for metodevurdering i Norge.

HVOR STOR TILLIT HAR VI TIL RESULTATENE?

Innen kunnskapsbasert medisin regnes gjerne en systematisk oversikt over randomiserte kontrollerte studier som det beste en kan få av dokumentasjon om effektene av et tiltak – toppen av "evidenshierarkiet". Men selv en godt utført systematisk oversikt som omfatter flere randomiserte studier kan ende opp med resultater som ikke er tillitverdige. Hvis, for eksempel, resultatene fra studiene spriker i alle retninger, eller det er store svakheter ved måten de enkelte studiene er utført på, er det tvisomt å trekke en sikker konklusjon om hva effekten av tiltaket faktisk er. Det er derfor nødvendig å supplere det kvantitative resultatet fra en meta-analyse med en vurdering av hvor sikre vi er på at resultatet ligger nært opptil den reelle effekten av tiltaket.

Verktøyet "GRADE" (Grading of Recommendations Assessment, Development and Evaluation) er under 10 år gammelt, men brukes nå av en rekke organisasjoner for å vurdere kvaliteten på den foreliggende dokumentasjonen om effekt av tiltak – blant annet Verdens helseorganisasjon (WHO) og The Cochrane Collaboration (21). I GRADE inngår en rekke vurderinger, blant annet: Er det alvorlige svakheter ved de enkelte studiene som inngår i den systematiske oversikten? Er resultatene konsistente på tvers av studiene? Hvor overførbare er resultatene til vår problemstilling? Er det snakk om stor effekt? Til sist gjøres en helhetsvurdering av kvaliteten på dokumentasjonen (21).

Til syvende og sist kommer en ikke utenom at det alltid vil være et element av skjønn involvert i slike vurderinger. En vesentlig styrke ved GRADE-systemet er at de subjektive vurderingene blir eksplisitte og dermed kan etterprøves av andre.

Hva er god kunnskap? Derom fortsetter de lærde å strides, men kravet om empiri og om gode metoder for generering av empiri har utvilsomt blitt langt strengere enn de var for bare få tiår siden.

REFERANSER

1. Jackson WA. A short guide to humoral medicine. *Trends Pharmacol Sci* 2001; **22**: 487-9.
2. Tröhler U. The introduction of numerical methods to assess the effects of medical interventions during the 18th century: a brief history. *JLL Bulletin: Commentaries on the history of treatment evaluation*, 2010.

3. Chalmers I, Dukan E, Podolsky SH, et al. The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries. *JLL Bulletin: Commentaries on the history of treatment evaluation*, 2011.
4. Choksy N. Professor Lustig's plague serum. *Lancet* 1900; **156**: 291-2.
5. Medical Research Council. The serum treatment of lobar pneumonia. *Br Med J* 1934; **1**: 241-5.
6. Chalmers I. Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers. *JLL Bulletin: Commentaries on the history of treatment evaluation*, 2010.
7. D'Arcy HP. A change in scientific approach: from alternation to randomised allocation in clinical trials in the 1940s. *BMJ* 1999; **319**: 572-3.
8. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *Br Med J* 1948; **2**: 769-82.
9. Forsetlund L, Chalmers I, Bjørndal A. When was random allocation first used to generate comparison groups in experiments to assess the effects of social interventions? *Economics of Innovation and New Technology* 2007; **16**: 371-84.
10. Fisher RA. The arrangement of field experiments. *Journal of the Ministry of Agriculture in Great Britain* 1926; **33**: 503-13.
11. Meldrum ML. A brief history of the randomized controlled trial. From oranges and lemons to the gold standard. *Hematol Oncol Clin North Am* 2000; **14**: 745-60, vii.
12. Cochrane AL. Effectiveness & Efficiency: Random Reflections on Health Services. London: Nuffield Provincial Hospitals Trust, 1973.
13. Cochrane AL. 1931-1971: a critical review, with particular reference to the medical profession. Medicines for the year 2000. London: Office of Health Economics, 1979.
14. Chalmers I, Hetherington J, Newdick M, et al. The Oxford Database of Perinatal Trials: developing a register of published reports of controlled trials. *Control Clin Trials* 1986; **7**: 306-24.
15. Crowley P, Chalmers I, Keirse MJ. The effects of corticosteroid administration before preterm delivery: an overview of the evidence from controlled trials. *Br J Obstet Gynaecol* 1990; **97**: 11-25.
16. O'Rourke K. A historical perspective on meta-analysis: dealing quantitatively with varying study results. *JLL Bulletin: Commentaries on the history of treatment evaluation*, 2006.
17. Glass GV. Meta-Analysis at 25. *gvglass info*, 2000. <http://www.gvglass.info/papers/meta25.html> (13.2.2013).
18. Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977; **32**: 752-60.
19. Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMJ* 1992; **305**: 786-8.
20. Mørland B. Erfaringer fra 5 år med medisinsk metodevurdering i Norge. *Tidsskr Nor Lægeforen* 2003; **123**: 1213-5.
21. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**: 924-6.