

## Kort indføring i evidens

Claus Thomsen<sup>1</sup>, lægefaglig chef, dr.med., phd; Jens F. Dahlerup<sup>2</sup>, overlæge, dr.med.

<sup>1</sup>Projektafdelingen for Det Nye Universitetshospital, Region Midtjylland; <sup>2</sup>Medicinsk afdeling V, Århus Universitetshospital, Århus Sygehus.

### Evidenshierarkier

I den medicinske verden har man tradition for, at undersøge og bedømme især behandlingseffekt ud fra videnskabelige metoder. Man inddeler styrken af videnskabelig dokumentation bl.a. ud fra den metode, der har dannet grundlag for vurderingen. Der er således dannet forskellige former for evidenshierarkier (tabel 1 og tabel 2), hvor det randomiserede kontrollerede forsøg (RCT'et) betragtes som det mest valide forsknings- og evalueringdesign og derfor anses for "guldstandard". I den medicinske verden anses RCT også for højeste evidens, hvorfor vi i det følgende kort vil gennemgå RCT:

*Et RCT indebærer opdeling af undersøgelsessubjekter i mindst en interventionsgruppe og en kontrolgruppe. Subjekterne (som ofte forstås som individer, men som også kan være organisationer eller andre enheder) i interventionsgruppen modtager den indsats, man ønsker at undersøge, mens kontrolgruppens subjekter modtager sammenligningsgrundlaget, som fx kan være placebo eller den hidtidigt tilbudte indsats. Centralt i forskningsdesignet er, at allokeringen til disse to grupper er randomiseret (tilfældig), således at man eliminerer selektionsbias og sikrer, at indsats- og kontrolgruppen har de samme karakteristika (inden for statistisk tilfældige grænser). Med andre ord søger man ved randomiseringen at sikre, at kun ét karakteristikum er forskelligt i indsats- og kontrolgruppen, nemlig den intervention, man ønsker at undersøge.*

*Guldstandard indebærer således, at indsats- og kontrolgruppen består af to ens grupper, der behandles samtidig, det såkaldte parallelle design. En undersøgelse, der er designet som et RCT, kan afdække, om en indsats har ført til resultater, men den kan ikke i sig selv belyse, hvorfor eller hvorfor ikke indsatsen giver resultater.*

*Man kan endvidere opdele RCT efter graden af blinding, altså i hvilken grad forsøgssubjekterne og forsøgslederne er klar over, hvilke subjekter, der befinder sig i henholdsvis interventions- og kontrolgruppen. Hvis alle involverede parter har kendskab til allokeringen, tales der om ublindede forsøg. Hvis kun én af parterne kender tilskrivningen, kaldes det et enkeltblindet forsøg. Endelig kalder man de tilfælde, hvor allokeringen er skjult for begge parter, for dobbeltblindede (1), hvilket anses for det stærkeste design indenfor RCT. En særlig form for RCT er de såkaldt cross-over-studier, hvor forsøgspersonerne udgør deres egen kontrol ved at skifte gruppe undervejs.*

*Det randomiserede kontrollerede forsøgs styrke er, at randomiseringen i en interventions- og en kontrolgruppe tager højde for tilstedeværelsen af ukendte faktorer, som kan have indflydelse på forsøgsresultatet (konfoundere). Man kontrollerer derved »automatisk« for alle faktorer og ikke alene for de faktorer, som er forskerne bekendt, hvorfor RCT betragtes som højere rangerende end fx økonometriske metoder.*

*Anvendelsen af RCT-design sikrer den bedst mulige dokumentation af årsags-virknings-forholdet mellem indsats og effekt. Den interne validitet karakteriseres på denne baggrund som værende meget høj. Kritikere har heroverfor fremført, at den eksterne validitet til gengæld kan være lav - at der med andre ord kan være problemer med at generalisere resultaterne til andre situationer og sammenhænge.*

Evidenshierarkiet kan beskrives i op til 9 niveauer (tabel 1 og tabel 2). RCT betragtes klart som det undersøgelsesdesign, der producerer viden af størst validitet, mens kvalitative design og aktørmodeller (ekspertvurderinger og brugervurderinger) betragtes som placeret i bunden af hierarkiet.

Mellem disse yderpunkter rangeres en række design frem for alt ikke-randomiserede eksperimenter og forløbsstudier, herunder kohorteundersøgelser. På trods af at evidenshierarkiet indeholder op til 9 niveauer, er betoningen af RCT ofte så markant, at det

næsten udelukkende er denne guldstandard, der associeres med evidens. I dagligdagens kliniske verden er det dog kun en mindre del af aktiviteterne, der er baseret på RCT–evidens.

I DNU-sammenhæng skal vi arbejde på at indsamle og bruge viden i metaanalyse/review-form. Det afgørende er, at validiteten af reviews afhænger af kvaliteten af de primære arbejder, der indgår i analysen. Det betyder for evidensarbejdet i DNU, at man skal være kritisk i opsamlingen af viden, men også erkende, at den aktuelle viden omkring Evidensbaseret Design ikke kan forventes at være af højeste evidensniveau.

Tabel 1 (ref. 1 og 2). Det "klassisk" medicinske evidenshierarki

Niveau	Studietype
<b>1a</b>	Systematiske metaanalyser af RCT.
<b>1b</b>	Enkeltstående RCT af god kvalitet.
<b>1c</b>	Kontrollerede, men ikke randomiserede forsøg.
<b>2a</b>	Systematiske reviews af kohortestudier.
<b>2b</b>	Enkeltstående kohortestudier. Dårlige RCT.
<b>3a</b>	Systematiske reviews over casekontrolstudier.
<b>3b</b>	Enkeltstående casekontrolstudier.
<b>4</b>	Caseserier eller kohortestudier eller casekontrol af dårlig kvalitet.
<b>5</b>	Ekspertvurderinger, konsensuskonferencer, kvalitative designs mv.

Tabel 2 (ref. 3) præsenterer to eksempler på evidenshierarkier. Eksempel A, der også præsenteres i dansk oversættelse, indeholder 9 niveauer, eksempel B 6 niveauer.

Niveau	Dansk udgave	Eksempel A Hierarchy of evidence in meta-analysis (ref. 5)	Eksempel B Hierarchy of evidence (ref. 6)
<b>1</b>	Randomiserede, kontrollerede eksperimenter (dobbeltblindedede, enkeltblindedede eller ublindedede)	Randomized controlled trials (with concealed allocation)	At least one properly designed RCT of appropriate size
<b>2</b>	Kvasieksperimenter: Kontrollerede forsøg baseret på matching	Quasi-experimental studies (using matching)	Well-controlled trials without randomisation
<b>3</b>	Forløbsstudier	Before-and-after Comparison	Well-designed cohort or case control studies
<b>4</b>	Tværsnitsundersøgelser	Cross-sectional, random sample studies	Multiple time series or dramatic results from uncontrolled experiments
<b>5</b>	Procesevaluering, aktionsforskning o.lign.	Process evaluation, formative studies and action research	Opinions of respected authorities based on clinical evidence, descriptive studies or expert committee
<b>6</b>	Kvalitative casestudier og etnografiske feltstudier	Qualitative case studies and ethnographic research	Small uncontrolled case series and samples
<b>7</b>	Erfaringer og eksempler på god praksis	Descriptive guides and examples of good practice	
<b>8</b>	Ekspertvurderinger	Professional and expert Opinion	
<b>9</b>	Brugervurderinger	User opinion	

### Evidenstypologi – en alternativ forståelse af evidens

Som alternativ til evidenshierarkiet kan man arbejde ud fra en evidenstypologi, hvor forskellige forskningsmetoder ikke rangordnes, men i stedet grupperes efter, hvad de er gode til at sige noget om. Man kan også her vurdere kvaliteten af et givet studie, f.eks ud fra systematikken i indsamlingen af data, om der er åbenhed om inklusionskriterierne, samt ud fra den omhyggelighed, der har været i anvendelsen af den valgte metode.

I en evidenstypologi kan man skematisk angive, hvor gode forskellige undersøgelses-design er til at belyse forskellige forskningsspørgsmål. Evidenstypologien fremhæver således forskellige designs styrker og svagheder (tabel 3).

Ud fra evidenstypologien er det vigtigt, at trække på forskellige evidens typer og viden fra andre fagområder end den medicinske – både når man skal prioritere indsatser, når indsatser skal implementeres, og når man skal måle effekten af indsatserne. Synspunktet er, at man ved at anerkende og benytte andre metoder end de medicinske forskningsmetoder vil kunne lave en bredere og mere nuanceret vurdering af indsatsers kvalitet og deres mulighed for at blive vellykket implementeret.

Tabel 3 (ref. 3). Eksempel på en evidenstypologi

Design	Kvalitative studier	Spørgeskemaundersøgelse	RCT
<b>Forskningsspørgsmål</b>			
Effekter – virker det?	++*)		++
Proces – hvorfor virker det?	++	+	
Sikkerhed – er der skadevirkninger?	+		++
Efterspørgsel – vil borgerne efterspørge den service, som tilbydes?	++	+	+
Omkostningseffektivitet – fås der nok for pengene?			++
Tilpasning – er det det rigtige tilbud til målgruppen?	++	++	
Er brugerne tilfredse?	++	++	
Note: Jo flere krydser, des større udsagnskraft har det pågældende design i forhold til den nævnte problemstilling			

\*) Ikke jf. Petticrew M and Roberts H (2003), men jvf. Flyvbjerg, B. (2000)

## Referencer

- 1) Sackett et al.: British Medical Journal.2000; 312 (7023): 71-72
- 2) Pedersen et al.: Ugeskrift for Læger. 2001: 163: 3769-3772.
- 3) Petticrew M and Roberts H (2003). Evidence, hierarchies, and typologies: horses for courses. J Epidemiol Community Health 57:527-529
- 4) Flyvbjerg, B (2000). Rationalitet og magt. Bind 1. Det konkrete videnskab. København. Akademisk forlag
- 5) Pawson R (2006). Evidence-based Policy. A Realist Perspective. London: Sage
- 6) Clarke A. (2006). Evidence-Based Evaluation in Different Professional Domains: Similarities, Differences and Challenges. In IF Shaw, JC Greene, MM Mark (eds.): Handbook of Evaluation. Policies, Programs and Practices. SAGE Publications.