

Klaringsrapporterne 1996-2001

Lave anførte evidensniveauer for den diagnostiske træfsikkerhed af anbefalede klinisk-biokemiske undersøgelser

Ledende overlæge Per E. Jørgensen,
1. reservelæge Jonna Skov Madsen,
lægelig forsker Lone G.M. Jørgensen,
overlæge Torben Bjerregaard Larsen, lektor Jørgen Hilden &
ledende overlæge Erik Magid

*Odense Universitetshospital, Afdeling KKA,
Aalborg Sygehus, Klinisk-epidemiologisk Afdeling,
Vejle Sygehus, Klinisk-biokemisk Afdeling,
Københavns Universitet, Biostatistisk Afdeling, og
Amager Hospital, Klinisk-biokemisk Afdeling

Resumé

Introduktion: Formålet med studiet var dels at undersøge, hvilket evidensniveau der var anført for den diagnostiske træfsikkerhed af anbefalede klinisk-biokemiske undersøgelser, dels at opgøre den generelle anvendelse af referencer i klaringsrapporterne i årene 1996-2001.

Materiale og metoder: Hver af de 54 klaringsrapporter, som omhandlede en klassisk klinisk problemstilling, blev gennemgået for at finde anbefalinger om diagnostisk anvendelse af klinisk-biokemiske undersøgelser. Det anførte evidensniveau for den diagnostiske træfsikkerhed af disse undersøgelser blev bedømt ud fra et internationalt anerkendt evidenshierarki. Desuden blev det noteret, om man i klaringsrapporterne angav referencer, om der i så fald var henvisninger hertil i teksten, og om princippet med angivelse af den evidensbaserede styrke af anbefalinger blev anvendt.

Resultater: I 20 klaringsrapporter fandtes i alt 36 anbefalinger om diagnostisk anvendelse af biokemiske undersøgelser. For tre fjerdedele af disse anbefalinger var der anført det lavest mulige evidensniveau med hensyn til den diagnostiske træfsikkerhed. Opgørelsen af den generelle anvendelse af referencer viste, at man i fem ud af de 54 klaringsrapporter slet ikke anførte referencer, og at man kun i 34 bragte henvisninger til litteraturlisten i direkte relation til udsagn i teksten. Princippet med at angive den evidensbaserede styrke af anbefalinger blev anvendt i fem af klaringsrapporterne.

Diskussion: Internationalt er der stor opmærksomhed på behovet for at sikre, at diagnostiske procedurer anvendes på et solidt evidensgrundlag. Klaringsrapporterne må betragtes som væsentlige nationale kliniske vejledninger, og denne undersøgelse bekræfter, at der er et dokumentationsproblem for klinisk-biokemiske undersøgelser. Den generelle anvendelse af referencer tyder dog på, at det anførte evidensniveau også kan være utilstrækkeligt for andre typer af anbefalinger.

kliniske vejledninger ikke tilfredsstillende [1-4]. Det er vigtigt, at anbefalinger i sådanne vejledninger hviler på et solidt og gennemskueligt evidensgrundlag. Det bør fremgå, hvordan den anvendte litteratur er blevet udvalgt, og centrale anbefalinger bør være ledsaget af en entydig kildeangivelse, gerne med angivelse af evidensniveauet bag eller styrken af anbefalingen [5]. Mens man i dag har velkendte og gode værktøjer til at sikre et godt evidensniveau for anbefalinger om behandlinger, er forholdene vanskeligere, når det drejer sig om diagnostiske procedurer såsom klinisk-biokemiske undersøgelser [6-9].

Siden 1996 har mange lægevidenskabelige selskaber fået udarbejdet en række klaringsrapporter, der må betragtes som vigtige nationale kliniske vejledninger. Hovedformålet med denne undersøgelse er at belyse, hvilket evidensniveau der er anført for den diagnostiske træfsikkerhed af de klinisk-biokemiske undersøgelser, som man i klaringsrapporterne anbefaler anvendt i diagnostikken. Ved diagnostisk træfsikkerhed forstås undersøgelsens evne til at adskille patienter med og uden en given sygdom. Den belyses bedst ved at sammenligne resultaterne af den pågældende undersøgelse med resultaterne af en eksisterende diagnostisk guldstandard [10]. Under gennemgangen af klaringsrapporterne registrerede vi desuden den generelle brug af referencer, og i hvilken udstrækning man i rapporterne anvendte princippet med at anføre styrken af de givne anbefalinger.

Materiale og metoder

I perioden 1996-2001 blev der udsendt 60 klaringsrapporter. Følgende seks, som skønnedes ikke at være en gennemgang af en klassisk klinisk problemstilling, blev ekskluderet: Den gode medicinske afdeling, Klinisk genetik et nyt tværgående speciale, Kvalitetssikring på arbejdsmedicinske afdelinger, Overvågning og forbedring af lægefaglige kerneydelser, Audit - en metode til kvalitetsudvikling i klinisk praksis samt Den højt specialiserede enhed i det danske sygehusvæsen. Hver af de resterende 54 klaringsrapporter blev gennemgået af denne artikels forfattere i grupper på to. Der blev opnået konsensus om at vurdere følgende forhold:

- Indeholdt klaringsrapporten en graderet angivelse af evidensniveauet eller styrken bag anbefalingerne?
- Blev der anvendt referencer i klaringsrapporten? Hvis ja, var der så litteraturhenvisninger i teksten, så man uden videre kunne se, hvilke referencer de enkelte udsagn var baseret på?
- Indeholdt klaringsrapporten egentlige anbefalinger vedrørende diagnostisk anvendelse af specifikke biokemiske undersøgelser i forbindelse med udredning af den sygdom,

Både lokalt, nationalt og internationalt lægges der stor vægt på gode kliniske vejledninger, der kan sikre en ensartet og rationel patientbehandling. Desværre er kvaliteten af mange

VIDENSKAB OG PRAKSIS | ORIGINAL MEDDELELSE

Tabel 1. Evidenshierarki for studier, der belyser den diagnostiske træfsikkerhed af diagnostiske undersøgelser.

Evidens-niveau	Styrke	Diagnostisk undersøgelse
1a	A	Systematisk oversigt (uden inhomogenitet ^a) af diagnostiske studier, som selv er evidensniveau 1b-studier.
1b	A	En konsekutiv og klinisk relevant patientpopulation, hvor alle er undersøgt med både den evaluerede test og referencetesten, og hvor de to test er tolket uafhængigt af hinanden.
2a	B	Systematisk oversigt (uden inhomogenitet ^a) af diagnostiske studier, som selv er mindst evidensniveau 2b.
2b	B	En ikkekonsekutiv eller en speciel og ikkerepræsentativ patientpopulation, hvor alle er undersøgt med både den evaluerede test og referencetesten, og hvor de to test er tolket uafhængigt af hinanden.
3	B	En klinisk relevant patientpopulation, hvor ikke alle er undersøgt med referencetesten. Den evaluerede test og referencetesten er tolket uafhængigt af hinanden.
4	C	Referencetesten anvendes eller tolkes ikke uafhængigt af den evaluerede test.
5	D	Ekspertvurdering uden anført evidens samt anbefalinger på baggrund af laboratorieforskning eller biokemisk/fysiologisk viden.

Tabellen er let modificeret i forhold til *Sackett et al's* evidenshierarki [10]. Evidensniveauet er en klassifikation baseret på en videnskabelig undersøgelsesdesign. Styrken angiver den vægt, man tillægger en anbefaling, der er baseret på undersøgelser med et givet evidensniveau.

a) Inhomogenitet mellem forskellige studier betyder, at disse studier er så forskellige (f.eks. med hensyn til patientpopulationer), at forskelle mellem deres resultater ud over rent statistiske tilfældigheder er forventelige.

som klaringsrapporten handler om? Biokemiske undersøgelser, som selv er den diagnostiske guldstandard (f.eks. fastende glukose til diagnostik af diabetes, kolesterol samt triglycerider ved dyslipidæmi og koronarenzymer ved akut myokardieinfarkt), blev ikke medtaget. Anbefalinger vedrørende diagnostik af sygdomme, som var perifere i forhold til klaringsrapportens hovedemne, blev heller ikke medtaget. Vurdering af, om en given undersøgelse skulle betragtes som en klinisk biokemisk undersøgelse, skete efter konsensus i forfattergruppen.

- For hver anbefaling blev både de biokemiske undersøgelser og det diagnostiske formål registreret. Endvidere blev det noteret, hvorvidt anbefalingen var støttet af en eller flere referencer. I klaringsrapporter uden direkte litteraturhenvisninger i teksten blev det vurderet, om nogen af litteraturlistens referencer oplagt måtte formodes at støtte anbefalingen.

For hver anbefaling vedrørende diagnostisk anvendelse af klinisk biokemiske undersøgelser blev de ovenfor anførte referencer gennemgået af en af denne artikels forfattere for at fastlægge det anførte evidensniveau for den diagnostiske træfsikkerhed af den anbefalede undersøgelse. Evidensniveauet blev fastlagt ud fra et evidenshierarki, der er let modificeret i forhold til *Sackett et al's* [10] (Tabel 1). Dette evidenshierarki er valgt, fordi det er specifikt rettet mod studier af diagnostiske undersøgelser diagnostiske træfsikkerhed. Niveau 1 er det højeste/bedste evidensniveau, mens niveau 5 er det laveste/dår-

ligste. I de tilfælde, hvor der ikke var anført nogen referencer, blev anbefalingen betragtet som niveau 5 (ekspertvurdering uden anført evidens). I de tilfælde, hvor der var opgivet referencer, blev evidensniveauet fastlagt ud fra disses metodologiske design. Hvis referencen var en almindelig oversigtsartikel eller en anden klinisk vejledning, blev relevante artikler blandt dennes referencer fremskaffet. Hvis oversigtsartiklen eller den kliniske vejledning udelukkende refererede til andre ikkeoriginalartikler blev evidensniveauet sat som niveau 5. Anbefalinger, som omfattede flere biokemiske undersøgelser, blev tildelt evidensniveauet svarende til den undersøgelse, der havde det højeste niveau.

Resultater

Anbefalinger om diagnostisk anvendelse af klinisk-biokemiske undersøgelser

Tyve (37%) af klaringsrapporterne indeholdt anbefalinger vedrørende diagnostisk anvendelse af specifikke klinisk-biokemiske undersøgelser. De 20 klaringsrapporter indeholdt i alt 36 sådanne anbefalinger. Disse 36 anbefalinger, og de anførte evidensniveauer for den diagnostiske træfsikkerhed er anført i Tabel 2. To af anbefalingerne hvilede på evidensniveau 1, tre på niveau 2, en på niveau 3, to på niveau 4 og 28 på niveau 5. Således var det anførte evidensniveau det lavest mulige for 78% af anbefalingerne.

Generel brug af referencer og angivelse af evidensniveau i de 54 klaringsrapporter

I fem klaringsrapporter (9%) var der ingen referencer, og i 15 (28%) var der en litteraturliste, men ingen litteraturhenvisninger i direkte tilslutning til udsagnene i teksten. De sidste 34 klaringsrapporter (63%) havde en litteraturliste, hvortil der blev henvist i teksten. I fem klaringsrapporter (9%) anvendte man princippet med at angive evidensniveauet eller styrken af anbefalinger i relation til de givne anbefalinger.

Diskussion

I dette studie undersøgte vi, hvilket evidensniveau der var anført for den diagnostiske træfsikkerhed af de klinisk-biokemiske undersøgelser, som i klaringsrapporterne 1996-2001 blev anbefalet anvendt. For tre fjerdedele af anbefalingerne var der anført det lavest mulige evidensniveau, »ekspertvurdering uden anført evidens«. Da vi kun medtog undersøgelserne inden for klaringsrapporternes hovedemner, og da anbefalinger, der omfattede flere biokemiske undersøgelser fik et evidensniveau svarende til den med det højeste niveau, vil dette studie tendere til at overvurdere det anførte evidensniveau bag anbefalingerne om diagnostisk anvendelse af klinisk-biokemiske undersøgelser. Det lave anførte evidensniveau bag disse anbefalinger betyder ikke, at anbefalingerne er forkerte, men det svækker den vægt, læseren kan tillægge dem. Årsagen til det lave anførte evidensniveau er formentlig både utilstrækkelig opmærksomhed på betydningen af at anføre et so-

VIDENSKAB OG PRAKSIS | ORIGINAL MEDDELELSE

Tabel 2. Diagnostiske formål, anbefalede biokemiske undersøgelser og det højeste anførte evidensniveau bag disse anbefalinger i klaringsrapporterne 1996-2001.

Klaringsrapport og diagnostisk formål	Biokemiske undersøgelser	Højeste anførte evidensniveau
1996, nr. 2 Glomerulær lidelse hos patienter med mild til moderat hypertension	Proteinuri og for diabetikere mikroalbuminuri	5
1996, nr. 3 Postoperativ discitis	Sænkingsreaktion og C-reaktivt protein	5
1996, nr. 4 Udredning ved lændesmerter: Infektion/discitis	Sænkingsreaktion, C-reaktivt protein og leukocytal	5
Nyre/urinveje	U-blod, U-leukocytter	5
Længerevarende uforklarede rygsmærter eller mistanke om anden tilgrundliggende sygdom	Sænkingsreaktion, hæmoglobin, basisk fosfatase, calcium, kreatinin, M-komponent, U-blod	5
Nydiagnostiseret osteoporose med hvirvelsammenfald	Sænkingsreaktion, hæmoglobin, basisk fosfatase, ioniseret calcium, thyroideastimulerende hormon	5
1997, nr. 1 Urinblæretumorer	U-blod ved stiks eller mikroskopi	5
1997, nr. 3 Urinvejstumor	U-blod	5
1997, nr. 6 Gummihandskeallergi	RAST-test for latex	5
1997, nr. 7 Væksthormonmangel hos voksne	Væksthormon i forbindelse med stimulationstest (insulintolerancetest eller GHRH + pyridostigmin), <i>insulin-like growth factor-1</i>	5
1998, nr. 1 Diagnostik af eksacerbation ved kronisk obstruktiv lungesygdom	Leukocytter, C-reaktivt protein, hæmatokrit	5
1998, nr. 2 Akut viral hepatitis	Alanin aminotransferase, aspartat aminotransferase, bilirubin, koagulationsfaktorerne 2, 7, 9 og 10	5
1998, nr. 4 Diagnostik af kolorektal cancer	Hæmoglobin	5
Screening for kolorektal cancer	F-hæmoglobin F-hæmoglobin	1b 1b
1999, nr. 4 Udredning ved klinisk mistanke om prostatacancer (PC)	PSA	2b
Diagnostik af PC og PSA <10 ng/ml	Frit/totalt PSA	3
Yderligere diagnostisk information ved frit/totalt PSA <10 ng/ml udelukker knoglemetastaser synlige ved knoglescintigrafi	PSA	2b
PSA <25 ng/ml udelukker lymfeknudemetastaser synlige ved CT	PSA	5
Ved »gunstige« kombinationer af PSA, histologisk grad og T-kategori kan lymfadenektomi udelades	PSA	4
1999, nr. 6 Subarachnoidealblødning hos patienter, hvor CT er negativ	Spektrofotometrisk undersøgelse for xantokromi på centrifugeret spinalvæske udtaget tidligst 12 timer efter blødningen	2b
1999, nr. 9 Hypertensio arterialis, differentialdiagnostik	Kreatinin, aldosteron, renin, U-mikroalbumin, U-blod, U-sedimenter	5
2000, nr. 1 Udredning af urininkontinens	U-blod, U-sukker, U-leukocytter	5
2000, nr. 3 Biokemisk trombofili hos patienter med kliniske risikofaktorer for trombose	Aktiveret protein-C-resistens, protein-C, protein S, antitrombin, lupus antikoagulans, antikardiolipinantistoffer, homocystein, fibrinogen, trombocytter	5
2000, nr. 10 Differentialdiagnostik ved osteoporose:hyper- eller hypoparathyroidisme	Parathyroideahormon, calcium	5
Malignitetsmistanke	Hæmoglobin, leukocytter, trombocytter, sænkingsreaktion	5
Nyrelidelse	Kreatinin	5
Øget knogleomsætning/leverlidelse	Basisk fosfatase	5
Leverlidelse	Alanin aminotransferase	5
Hypogonadisme (hos mænd)	Østradiol (testosteron)	5
Menopausestatus	Follikelstimulerende hormon	5

VIDENSKAB OG PRAKSIS | ORIGINAL MEDDELELSE

2001, nr. 2 Udredning ved demens	Hæmoglobin, erythrocytter, middelcellevolumen, middelcellehæmoglobinkoncentration, leukocyter, trombozytter, sænkingsreaktion eller C-reaktivt protein, kreatinin, natrium, kalium, calcium, albumin, glukose, thyreoideastimulerende hormon, alanin aminotransferase, vitamin-B ₁₂ eller methylmalonat	5
2001, nr. 4 Prostatacancer, hvor biopsi har vist negativt resultat, men hvor der er klinisk mistanke	PSA, frit/totalt PSA, PSA-velocity	4
Mistænkt dissemineret prostatacancer, hvor kurativ behandling er udelukket, hvor den primære biopsi er negativ, og hvor man ikke ønsker at rebiopere	PSA, alkalisk fosfatase, kreatinin, hæmoglobin	5
2001, nr. 10 Udredning ved urininkontinens hos geriatriske patienter	U-blod, U-sukker, U-leukocyter/bakterier, kreatinin, glukose, cobalamin, methylmalonat, folat	5
2001, nr. 11 Differentialdiagnostik ved vandladningsforstyrrelser ved neurologisk sygdom	Urinstiks (blod, sukker, protein og leukocyter), kreatinin	5

Evidensniveau 1 er bedst, 5 dårligst (Tabel 1).

Hvor intet andet er anført, er der tale om blodprøver.

U- og F- angiver, at der er tale om urin- og fæcesprøver.

RAST: *radio-immunosorbent test*

GHRH: *growth hormone releasing hormone*

PSA: prostataspecifikt antigen

lidt evidensniveau bag diagnostisk anvendelse af klinisk-biokemiske undersøgelser, og at mange af de eksisterende originalartikler om emnet er af dårlig kvalitet [7, 9, 11].

I forbindelse med den »kliniske nytteværdi« af en diagnostisk undersøgelse kan man både interessere sig for den diagnostiske træfsikkerhed og for den pågældende undersøgelses direkte konsekvenser for patienten (*outcome*) [12]. *Outcome* belyses bedst ved randomiserede kontrollerede undersøgelser, om end der kan være behov for designmæssige justeringer i forbindelse med diagnostiske undersøgelser [13]. I dette studie har vi fokuseret på diagnostisk træfsikkerhed, fordi det er det aspekt af diagnostiske undersøgelser, der oftest bliver beskrevet i litteraturen. Desuden er der netop nu betydelig fokus på behovet for at forbedre kvaliteten af sådanne artikler. I 2000 blev der nedsat en arbejdsgruppe kaldet Standards for Reporting of Diagnostic Accuracy (STARD), der består af en række tidsskriftsredaktører samt epidemiologer og laboratorielæger [12]. Formålet med STARD-gruppens arbejde er at skabe konsensus for, hvordan studier af diagnostiske undersøgelser diagnostiske træfsikkerhed bør gennemføres og afrapporteres. STARD-gruppens første anbefalinger med checklister og forslag til flowdiagrammer er netop blevet publiceret [13-16]. Checklisten omfatter en detaljeret beskrivelse af den kliniske problemstilling, af patientpopulationen og af de analysetekniske forhold omkring den diagnostiske undersøgelse. For klinisk-biokemiske undersøgelser vil dette kun kunne opnås ved studier, der gennemføres i et tæt samarbejde mellem det klinisk-biokemiske og andre kliniske specialer.

Der er en stærkere tradition for at sikre et godt evidensgrundlag for behandlingsmæssige end for diagnostiske procedurer. I en ny dansk undersøgelse fandt man, at omkring halvdelen af de behandlingsmæssige interventioner er baseret på randomiserede kontrollerede forsøg [17]. At anvendelsen af referencer i knap 40% af klaringsrapporterne var af en sådan karakter, at det var vanskeligt eller umuligt at finde relevante

referencer for specifikke udsagn, tyder dog på, at det anførte evidensgrundlag også kan være utilstrækkeligt for andre ikke-diagnostiske anbefalinger.

Korrespondance: *Per E. Jørgensen*, Afdeling KKA, Odense Universitetshospital, DK-5000 Odense C. E-mail: per.jorgensen@ouh.fyns-amt.dk

Antaget: 9. januar 2004

Interessekonflikter: Ingen angivet

*) Dansk Selskab for Klinisk Biokemi arbejdsgruppe til fremme af evidensbaseret anvendelse af klinisk biokemi.

Litteratur

- Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are Guidelines following guidelines? – the methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999;281:1900-5.
- Varonen H, Mäkelä M. Practice guidelines in Finland: availability and quality. *Qual Health Care* 1997;6:75-9.
- Grilli R, Magrini N, Penna A et al. Practice guidelines developed by speciality societies: the need for a critical appraisal. *Lancet* 2000;355:103-6.
- Miller J, Petrie J. Development of practice guidelines. *Lancet* 2000;355:82-3.
- Habour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;323:334-6.
- Jørgensen PE, Madsen JS, Larsen TB et al. Evidensbaseret anvendelse af klinisk biokemi. *Ugeskr Læger* 2001;163:5815-9.
- Lijmer JG, Mol BW, Heisterkamp S et al. Empirical evidence of design-related bias in studies of diagnostic tests *JAMA* 1999;282:1061-6.
- Price CP. Evidence-based laboratory medicine: supporting decision making. *Clin Chem* 2000;46:1041-50.
- Oosterhuis WP, Niessen RWLM, Bossuyt PMM. The science of systematic reviewing studies of diagnostic tests. *Clin Chem Lab Med* 2000;38:577-88.
- Sackett DL, Straus SE, Richardson WS et al. Evidence-based Medicine – how to practice and teach EBM. 2nd edition. Edinburgh: Churchill Livingstone, 2000.
- Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research – getting better but still not good. *JAMA* 1995;274:645-51.
- Magid E, Jørgensen PE, Jørgensen LGM et al. Diagnostik: nye metodologiske initiativer. *Ugeskr Læger* 2004;166:XXXX-X.
- Bossuyt PMM, Lijmer JG, Mol BWJ. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 2000;356:1844-7.
- Bossuyt PM, Reitsma JB for the STARD group. The STARD initiative. *Lancet* 2003;361:71.
- Bossuyt PMM, Reitsma JB, Bruns DE et al. STARD group. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Int Med* 2003;138:40-4.
- Bossuyt PMM, Reitsma JB, Bruns DE et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Int Med* 2003;138:W1-W12.
- Matzen P. Hvor evidensbaseret er medicinen? *Ugeskr Læger* 2003;165:1431-5.