

Metodehåndbog for udarbejdelse af Nationale Kliniske Retningslinjer 2017-2020

© Sundhedsstyrelsen, 2017. Publikationen kan frit refereres med tydelig kildeangivelse.

Sundhedsstyrelsen
Islands Brygge 1
2300 København S

URL: <https://www.sst.dk/da/nkr>

Sprog: Dansk

Versionsnummer 1.0

Udgivet af Sundhedsstyrelsen, marts 2017.

Indhold

1	Introduktion	5
1.1	Baggrund	5
1.2	Om håndbogen	5
1.3	Definition – hvad er en NKR?	5
1.4	Organisatorisk ramme	5
1.5	Afgrænsning til andre faglige anbefalinger	6
2	Fase I – ansøgning til puljen	7
2.1	Valg af emner og afgrænsning	7
2.2	Organisering	7
2.3	Indsendelse af protokol	8
3	Fase II - Udarbejdelse	10
3.1	Fokuserede spørgsmål	11
3.1.1	Spørgsmål om effekten af en intervention inklusive screening/tidlig opsporing	11
3.1.2	Spørgsmål om diagnostiske tests til udredning	14
3.1.3	Diagnostisk præcision	15
3.1.4	Patientrelevante outcome	15
3.1.5	Hvilken rolle skal den nye test spille?	15
3.1.6	Overdiagnostik	16
3.2	Litteratursøgning	18
3.2.1	Søgestrategi og -beskrivelse	20
3.3	Obligatoriske programmer til udarbejdelse af retningslinjen	20
3.3.1	RefWorks	20
3.3.2	Covidence	21
3.3.3	Review Manager	21
3.3.4	MAGIC	21
3.4	Udvælgelse af evidens	21
3.4.1	Typer af studiedesign	22
3.4.2	Sortering af søgeresultater	23
3.5	Vurdering af evidens	24
3.5.1	Vurdering af retningslinjer	24
3.5.2	Vurdering af systematiske oversigtsartikler	27
3.5.3	Vurdering af primære studier	29
3.6	Sammenfatning af evidens	31
3.6.1	Dataekstraktion	31
3.6.2	Retningslinjer	31
3.6.3	Sammenfatning af resultater fra interventionsstudier	32
3.6.4	Sammenfatning af resultater fra diagnostiske primærstudier	32
3.7	Gradering af evidens (GRADE)	34
3.7.1	Vurdering af tiltroen til det enkelte outcome	34
3.7.2	Nedgradering af evidensen for effekt af interventioner	36
3.7.3	Opgradering	37
3.7.4	Nedgradering af evidens for præcision af diagnostiske tests	38
3.7.5	Evidensprofiler (interventionsstudier)	40
3.7.6	Evidensprofiler (diagnostiske tests studier)	43
3.7.7	Vurdering af tiltroen til det overordnede evidensniveau	44
3.8	Udarbejdelse af anbefalinger	45
3.8.1	Fra evidens til anbefaling ved interventionsstudier	45
3.8.2	Fra evidens til anbefaling ved diagnostiske studier	46
3.8.3	Manglende evidens	47
3.8.4	Evidensbaserede anbefalinger	47

	3.8.5	God praksis anbefalinger	49
	3.8.6	Videre forskning	49
	3.9	Implementering og monitorering	50
4		Fase III – publicering	52
	4.1	Indsendelse af færdigt udkast	52
	4.2	Format	52

1 Introduktion

1.1 Baggrund

Med Finanslov 2017 er der afsat 38 mio. kr. i perioden 2017-2020 til en pulje til videreførelse af arbejdet med nationale kliniske retningslinjer. Formålet er at understøtte en fortsat udvikling af ensartet behandling af høj kvalitet for patienterne, samtidig med at kompetenceudvikling og ejerskabet til udvikling af nationale kliniske retningslinjer baseret på GRADE-metoden uden for Sundhedsstyrelsen styrkes. Puljen etableres i regi af Sundhedsstyrelsen og målrettes faglige selskaber og miljøer til udarbejdelse og opdatering af kliniske retningslinjer. Dele af puljemidlerne anvendes ligeledes til Sundhedsstyrelsens opdatering af de eksisterende nationale kliniske retningslinjer.

1.2 Om håndbogen

I forbindelse med arbejdet med de nationale kliniske retningslinjer 2012- 2016 har Sundhedsstyrelsen fastlagt en model for udarbejdelse af NKR baseret på GRADE-metoden. Metodehåndbogen henvender sig til de faglige selskaber og miljøer som søger midler fra NKR-puljen til at udarbejde puljefinansierede nationale kliniske retningslinjer, men kan også anvendes som inspiration til fagpersoner, som skal udarbejde kliniske retningslinjer i andre sundhedsorganisationer.

1.3 Definition – hvad en er en NKR?

De nationale kliniske retningslinjer er systematisk udarbejdede udsagn, der kan bruges som beslutningsstøtte af fagpersoner, når de skal træffe beslutning om passende og korrekt sundhedsfaglig ydelse i specifikke kliniske situationer, samt af patienter og andre, som ønsker at orientere sig i retningslinjerne. Nationale kliniske retningslinjer dækker udvalgte dele af et patientforløb og er som udgangspunkt tværfaglige.

En national klinisk retningslinje indeholder udvalgte og velafgrænsede kliniske problemstillinger som kan ses som 'punktnedslag' i et patientforløb fra forebyggelse, de tidlige manifestationer, diagnostik, behandling og rehabilitering. De kliniske problemstillinger udvælges af de faglige selskaber og miljøer som de områder, hvor der mangler beslutningsstøtte. Hver af de udvalgte kliniske problemstillinger udmøntes i fokuserede spørgsmål som danner baggrund for en systematisk litteraturgennemgang og evidensvurdering. Formålet er at skabe overblik over gavnlige såvel som skadelige virkninger af specifikke sundhedsfaglige ydelser. På baggrund af denne afvejning udarbejdes anbefalinger.

Udgangspunktet er, at der besvares ca. 5 fokuserede spørgsmål per retningslinje. En national klinisk retningslinje er altså ikke en A til Z lærebog, men et evidensbaseret beslutningsstøtteværktøj målrettet klinikere.

1.4 Organisatorisk ramme

Kommissoriet for udarbejdelse af nationale kliniske retningslinjer sætter følgende organisatoriske ramme for arbejdet 2017-2020:

- En *pulje*, som kan søges af de faglige selskaber og miljøer til udvikling af nationale kliniske retningslinjer udarbejdet efter GRADE-metoden. Puljen opslås årligt.
- Et *sekretariat* i Sundhedsstyrelsen, der udvikler konceptet for NKR, administrerer puljen, kvalitetsvurderer retningslinjerne udarbejdet for puljemidler og betjener det rådgivende udvalg.
- Et *rådgivende udvalg*, der rådgiver vedr. metodevalg, rammer for arbejdet, kriterier for tildeling af midler og prioritering af ansøgninger m.v. Det rådgivende udvalg høres før udvælgelse af ansøgere til modtagelse af puljemidler og før publikation af de enkelte kliniske retningslinjer blandt andet med henblik på understøttelse af implementering i kommuner og regioner.

1.5 Afgrænsning til andre faglige anbefalinger

En national klinisk retningslinje indeholder alene anbefalinger for de udvalgte velafgrænsede problemstillinger i patientforløbet, hvorfor en national klinisk retningslinje ikke kan stå alene, men komplementeres og suppleres af andre retningslinjer – fx. tværfaglige- og tværsektorielle retningslinjer for andre dele af patientforløbet eller andre patientpopulationer, retningslinjer udarbejdet (monofagligt) af selskaber og faglige organisationer samt regionale og kommunale retningslinjer, vejledninger og instrukser. Arbejdet med nationale kliniske retningslinjer er således et supplement til men erstatter ikke den øvrige indsats i videnskabelige selskaber, faglige sammenslutninger m.fl. med at udvikle kliniske retningslinjer. De nationale kliniske retningslinjer kan give input til og supplere lokale retningslinjer. Det kan ikke forventes, at NKR alene opfylder behovet for retningslinjer, så der vil fortsat være behov for lokale og regionale retningslinjer, som bør baseres på NKR, hvor disse foreligger.

Puljefinansierede nationale kliniske retningslinjer udarbejdes helt eller delvist på finansiering fra NKR - puljen, men vil blive udarbejdet uden for regi af Sundhedsstyrelsen. Sundhedsstyrelsen vil kvalitetsvurdere den kliniske retningslinje før publicering baseret på AGREEII instrumentet (Appraisal of Guidelines for Research and Evaluation). AGREEII er et redskab til primært at bedømme den metodologiske kvalitet af kliniske retningslinjer, og den samlede retningslinje vil derfor ikke nødvendigvis være et udtryk for Sundhedsstyrelsens holdninger eller anbefalinger på det givne område. De nationale kliniske retningslinjer er ikke juridisk bindende, og ved uoverensstemmelse mellem forskellige kliniske retningslinjer vil det altid være det faglige skøn i den konkrete kliniske situation, der er afgørende for beslutningen om passende og korrekt sundhedsfaglig ydelse.

En national klinisk retningslinje indeholder alene konkrete handlingsanvisninger indenfor udvalgte, velafgrænsede kliniske problemstillinger (dvs. *hvad* der skal gøres og *hvem* er det relevant for). Den har ikke som primært formål at afklare visitation og organisering af indsatsen (*hvor* skal indsatsen tilbydes, *hvem* der skal tilbyde indsatsen eller *hvornår* den skal tilbydes) eller samfundsøkonomiske konsekvenser (hvad er den afledte effekt på *ressourcerne* og er disse til stede). Disse typer af problemstillinger kan eksempelvis håndteres i en visitationsretningslinje, et pakkeforløb, et forløbsprogram, et referenceprogram eller en medicinsk teknologivurdering.

2 Fase I – ansøgning til puljen

2.1 Valg af emner og afgrænsning

Sundhedsstyrelsen er overordnet ansvarlig for udvælgelse af, hvilke ansøgninger fra de faglige selskaber og miljøer, der skal støttes af puljemidler. Puljeuddelingen vil være baseret på transparente faglige kriterier for emnevalg m.v. med udgangspunkt i den hidtil anvendte NKR model.

Emneforslagene bør være afgrænsede til dele af patientforløbet eller dele af et sygdomsområde. Afgrænsningen betyder, at en NKR kun undtagelsesvis vil indeholde retningslinjer for et emne, der dækker hele patientforløbet fra forebyggelse, de tidlige manifestationer, over diagnostik, behandling og rehabilitering.

Kriterier for emner:

- Betydelig sygdomsbyrde (populationsstørrelse, sygelighed, dødelighed)
- Store variationer i praksis
- Ny risikobetonet teknologi
- Mulige indikationsskred (overbehandling)
- Stort ressourceforbrug (dyr behandling, ventetider m.v.)
- Sundhedspolitisk prioriteret område

Emner for nationale kliniske retningslinjer bør leve op til mindst to af ovenstående kriterier. Der lægges endvidere vægt på, om problemstillingen er egnet til behandling i en national klinisk retningslinje, herunder at problemstillingen er af klinisk snarere end organisatorisk karakter. Der lægges vægt på, at emneforslagene adresserer vigtige problemstillinger, dvs. enten kontroversielle, uafklarede eller påtrængende, og ikke trivielle eller allerede velbeskrevne problemstillinger.

2.2 Organisering

Den overordnede organisering og rammerne for arbejdet med udarbejdelse af den nationale kliniske retningslinje bør være på plads ved ansøgning om midler fra NKR-puljen. Foruden kontaktperson/juridisk ansvarlig behøver øvrige personer ikke være navngivne på ansøgningstidspunktet.

Sundhedsstyrelsen anbefaler følgende organisering, men det er ikke et krav for tildeling af midler:

- Projektleder
- Fagkonsulent, som fungerer som hovedskribent på den nationale kliniske retningslinje
- Metodekonsulent, med særlig viden om evidensvurdering
- Søgesepecialist til litteratursøgning
- Arbejdsgruppe, som skal repræsentere de relevante specialer og fagligheder

- Styregruppe eller referencegruppe, som kan medvirke til at højne kvaliteten og bidrage til implementering

I metodehåndbogen vil disse titler blive brugt ved beskrivelse af de processer og opgaver, der indgår i udarbejdelsen af en national klinisk retningslinje, men anden opgave – og ansvarsfordeling kan anvendes i det konkrete arbejde.

Arbejdsgruppen vil være omdrejningspunkt for udarbejdelsen af den kliniske retningslinje og må forvente at skulle bidrage med en betydelig arbejdsindsats. Arbejdsgruppen og øvrige personer, involveret i arbejdet, bør deltage i Sundhedsstyrelsens metodeundervisning for at kunne leve op til kravene til udarbejdelse af en national klinisk retningslinje.

Arbejdsgruppens opgaver vil være:

- Udvælgelse og fokusering af spørgsmål, som skal danne grundlag for anbefalingerne i retningslinjen (se afsnit 3.1 for mere information om fokuserede spørgsmål).
- Bidrage med søgeord til søgestrategien, således at relevant litteratur ikke overses. Selve søgestrategien udarbejdes af en søgespecialist i samarbejde med fagkonsulenten (se afsnit 3.2 for mere information om søgeprocessen).
- Efter at fagkonsulenten og metodekonsulenten har foretaget den indledende sortering af den fundne litteratur, indgår arbejdsgruppemedlemmerne i den endelige sortering af litteraturen, for at sikre at al relevant litteratur inkluderes (se afsnit 3.4.2 for mere information om sortering).
- AGREE II vurdering af relevante eksisterende retningslinjer (se afsnit 3.5.1 for mere information om AGREE II).
- AMSTAR vurdering af relevante systematiske oversigtsartikler (se afsnit 3.5.2 for mere information om AMSTAR).
- Kvalitetsvurdering af primær litteratur (se afsnit 3.5.3 for mere information om evidensvurdering af primær litteratur).
- Udformning af anbefalinger (se afsnit 3.8 for mere information om udarbejdelse af anbefalinger).
- Kritisk gennemlæsning af udkast og den endelige version af den nationale kliniske retningslinje.

2.3 Indsendelse af protokol

Blandt de ansøgere, der vurderes at leve op til kravene for tildeling af midler, prioriterer Sundhedsstyrelsen i samarbejde med det Rådgivende Udvalg for NKR, på baggrund af de tilgængelige puljemidler til fordeling, hvem der inviteres til at indsende en protokol til Sundhedsstyrelsen.

Protokollen skal nøje beskrive de valgte fokuserede spørgsmål, som skal være formuleret ud fra PICO tilgangen (se afsnit 3.1). Desuden skal retningslinjens målgruppe beskrives, og der skal være en overordnet redegørelse af den videre proces, der er planlagt til udarbejdelsen af retningslinjen. Det er bl.a. vigtigt at få beskrevet den systematiske litteratursøgning og hvordan kvaliteten af evidensen vil blive vurderet, så det bliver tydeligt, at der er en forståelse for brugen af GRADE metoden. Ydermere vil der være krav om at protokollen indeholder overvejelser omkring implementering og monitorering af retningslinjens anbefalinger samt behovet for videre forskning på området.

Sundhedsstyrelsen vil tage udgangspunkt i AGREEII instrumentet (Appraisal of Guidelines for Research and Evaluation) til at godkende protokollen.

Ansøgere, der får godkendt deres protokol, vil modtage midler fra puljen, og kan fortsætte udarbejdelsen af den nationale kliniske retningslinje.

Skabelon for protokol vil snarest blive tilgængelig.

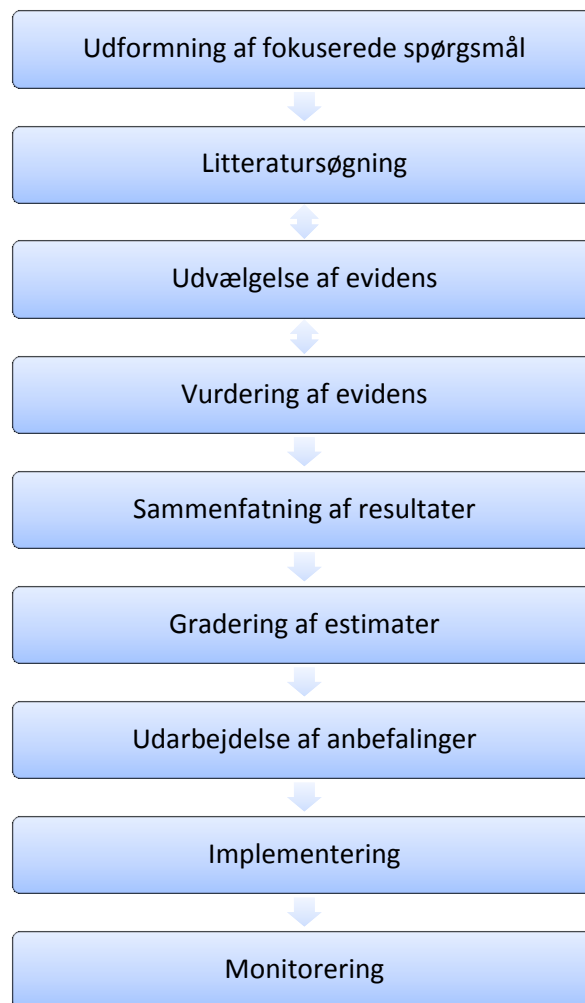
3 Fase II - Udarbejdelse

Nationale kliniske retningslinjer udarbejdes på baggrund af fokuserede spørgsmål, som afspejler kliniske problemstillinger, hvor der savnes retningslinjer.

Metoden for udarbejdelse af en national klinisk retningslinje er baseret på GRADE (Grading of Recommendations Assessment, Development and Evaluation system), der er et internationalt brugt systematisk og transparent system til at vurdere kvaliteten af evidens og til at give anbefalinger. Den følgende tekst vil kort beskrive systemet. For mere detaljeret information henvises til afsnit 3.7 og 3.8 samt GRADE arbejdsgruppens hjemmeside:

<http://www.gradeworkinggroup.org>

GRADE-processen starter med formuleringen af de fokuserede spørgsmål. Herefter foretages en systematisk søgning efter litteratur. Efter at litteraturen er fundet og man har fastlagt evidensgrundlaget for hvert fokuseret spørgsmål, angiver GRADE eksplicitte kriterier, hvormed man kan vurdere kvaliteten af evidensen ud fra fire kriterier fra meget lav til høj. Kriterierne inkluderer: risiko for bias, inkonsistens, unøjagtighed, hvor indirekte evidensen er samt risikoen for publikations bias. Denne proces sikrer, at der systematisk identificeres, vurderes og sammenfattes viden om interventioner eller diagnostiske tests.



Figur 1 Trinene i udarbejdelsen af en NKR

Sundhedsstyrelsen vil udbyde endags kurser i anvendelsen af GRADE, så arbejdsgruppemedlemmerne er klædt på til at anvende de forskellige metoder i forbindelse med udarbejdelsen af nationale kliniske retningslinjer. Sundhedsstyrelsen anbefaler, at alle der skal indgå i processen for udarbejdelsen af en national klinisk retningslinje deltager på et GRADE kursus.

3.1 Fokuserede spørgsmål

Emnet for den pågældende NKR er indledningsvist afgrænset i ansøgningen. Yderligere afgrænsning foretages af arbejdsgruppen, som beslutter, hvilke fokuserede spørgsmål retningslinjen skal indeholde. De fokuserede spørgsmål skal godkendes af Sundhedsstyrelsen i forbindelse med indsendelse af protokol.

De fokuserede spørgsmål er den røde tråd igennem hele retningslinjen. Spørgsmålene skal afspejle centrale kliniske problemstillinger, som man ønsker belyst. De fokuserede spørgsmål præsenteres alle i protokollen, som indsendes til godkendelse i Sundhedsstyrelsen inden processen for dataindsamling (herunder søgning efter litteratur til besvarelse af de endelige fokuserede spørgsmål) påbegyndes. Sundhedsstyrelsen anbefaler, at der formuleres op til 5 fokuserede spørgsmål.

De fokuserede spørgsmål udgør grundlaget for den efterfølgende udarbejdelse af litteratursøgningsstrategi, inklusions- og eksklusionskriterier, fokus ved læsning af litteratur, samt formulering af anbefalinger.

3.1.1 Spørgsmål om effekten af en intervention inklusive screening/tidlig opsporing

De fokuserede spørgsmål specificerer patientgruppen, interventionen, alternativet til interventionen og outcome. Dette afspejles i akronymet PICO (**P**opulation, **I**ntervention, **C**omparison and **O**utcomes)

Population (patientgruppen): Definition af sygdommen/tilstanden og patientmålgruppen med relevante demografiske faktorer (fx alder).

Intervention (interventionen): Definition af interventionen/erne. Hvilke variationer (eksempelvis dosis, administrationsform o.l.) findes der af interventionen? Skal alle variationer inkluderes?

Comparison (alternativet): Hvad er alternativet til den undersøgte intervention? Hvilken sammenligning er relevant i forhold til praksis? I situationer hvor der er usikkerhed om effekten af interventionen kan det være relevant at sammenligne med placebo/ingen behandling. I de situationer, hvor der er bred enighed om at interventionen virker, bør der sammenlignes med standard behandling eller en anden form for intervention.

Outcome (effekter): Hvilke outcome vurderes at være nødvendige at vide noget om for at opveje gavnlige mod skadelige effekter af den undersøgte intervention? Her skal indtænkes relevante outcome for alle beslutningstagere, det vil sige patienten såvel som klinikerens. Det er således vigtigt, at man i arbejdsgruppen udvælger outcome ud fra vigtighed og ikke blot ud fra, hvad det vurderes, at man finder rapporteret i litteraturen.

Outcome kan omfatte kliniske hændelser (fx frakturer, slagtilfælde) og andre patient relevante outcome (fx funktionsevne, symptomer, livskvalitet). Relevante skadevirkninger skal *altid* inkluderes, herunder alvorlige bivirkninger (SAE, serious adverse events), fx dødelighed, blødninger.

Udover at gøre sig klart hvilke outcome, der er relevante for retningslinjen, bør arbejdsgruppen også overveje, hvordan disse outcome kan måles. Hvilke måleredskaber vil fx sandsynligvis blive anvendt? Hvis der er flere måleredskaber, der kan

bruges til at måle den samme effekt, fx smerte, hvilket måleredskab vil så blive foretrukket, såfremt outcome er af rapporteret med flere skalaer i det samme studie?

Derudover er det vigtigt at fastlægge hvilke tidspunkter, der er relevante at se på. Ønsker man at se på både kortsigtede og langsigtede virkninger? Hvorledes vil man definere kortsigtet og langsigtet? Dette er vigtigt, da effekter kan variere betydeligt over tid og det kan således have stor betydning om man vælger at fokusere på effekten efter 6 uger eller efter 12 uger. Det kan også være relevant at man i arbejdsgruppen diskuterer mindste kliniske relevante forskel for hvert enkelt outcome.

Af hensyn til overskueligheden anbefaler Sundhedsstyrelsen, at der højst formuleres 8 outcome.

Outcome skal specificeres og rangordnes efter, hvorvidt de er kritiske for beslutningstagningen eller blot vigtige, men ikke kritiske for beslutningstagningen. Da det vil være vanskeligt at definere de kritiske og vigtige effekter for en del fagområder, kan der være hjælp at hente på denne hjemmeside: <http://www.comet-initiative.org>, hvor meget forskning vedrørende standardisering af relevante outcome (core outcomes) er samlet.

Mens alle relevante kritiske og vigtige outcome bør inkluderes, bør mindre vigtige outcome ikke medtages. Dette fordi retningslinjen ellers ville blive vildledende, da læseren ville blive forelagt data, der er af ringe eller ingen betydning. Man bør desuden udvise forsigtighed ved brugen af surrogat outcome, såsom laboratorieresultater eller radiologiske resultater (fx tab af knoglemineralindhold som et surrogat for frakturer i forbindelse med hormonbehandling). Disse bør undgås eller fortolkes med forsigtighed, da de ikke er præcise mål for klinisk relevante effekter (patient-relevante outcome). Interventioner kan således have en god virkning på et surrogat outcome, men ingen eller endog skadelige indvirkning på det klinisk relevante outcome som surrogat outcomet menes at være et udtryk for.

I tilfælde af, at der kun findes surrogatmål rapporteret, kan man benytte dette til at belyse et kritisk outcome, men man bør så nedgradere kvaliteten af evidensen (pga. indirectness) Man bør aldrig vælge et surrogat outcome (laboratorieoutcome) som et kritisk outcome.

De fokuserede spørgsmål formuleres i protokollens dertil indrettede skema. Skabelon for protokol vil snarest blive tilgængelig.

Nedenfor er der et eksempel på overvejelser i forbindelse med et tænkt fokuseret spørgsmål om, hvorvidt børn med mellemørebetændelse bør tilbydes antibiotika:

Populationen: Børn med mellemørebetændelse. Her kan det være relevant at overveje, om det skal gælde alle børn eller om man for eksempel mener, at der for børn under seks måneder er særlige forhold der kan spille ind. Hvis man mener dette skal man overveje at splitte spørgsmålet op så man har et der går på børn under seks måneder, og et der går på børn over seks måneder.

Interventionen: Antibiotikabehandling. Her skal man overveje, om man vil tage alle former med eller om det kun er specifikke former for antibiotika, man er interesseret i.

Comparison: Her skal man overveje hvad alternativet er. I dette eksempel kunne sammenligningen være 'at vente og se' (Watchful waiting).

Outcome: Når outcome skal fastlægges, kan det være en god ide med en bruttoliste (se eksempel på næste side), hvor alle potentielt relevante outcome bringes op. Herefter kan arbejdsgruppen diskutere, hvilke de vil beholde og hvilke som ikke er så vigtige alligevel. Hvilke ville du vælge?

Eksempel på bruttoliste:

- Smerter dag 1
- Smerter dag 2-3
- Smerter dag 4-7
- Smerter dag 7-14
- Sygdomsvarighed
- Helbredelse/respons
- Andel med mavesmerter
- Andel med kvalme
- Andel med diarre
- Andel med udslæt
- Ingen klinisk bedring dag 3
- Forværring af tilstanden på hvilket som helst tidspunkt
- Spontan perforation af trommehinden
- Svær infektion fx mastoiditis eller meningitis
- Serious adverse event (SAE)
- Stop af behandling
- Stop af behandling pga manglende effekt
- Stop af behandling pga skadevirkninger
- Udvikling af dobbeltsidig mellemørebetændelse (blandt de børn hvor det var ensidigt til start)
- Gentagne mellemørebetændelser 1 år efter behandling
- Hørelse 1 år efter behandling
- Hørelse 5 år efter behandling

3.1.2 Spørgsmål om diagnostiske tests til udredning

De fokuserede spørgsmål specificerer patientgruppen, den eller de diagnostiske test(s) man ønsker at undersøge, reference standarden samt outcome, hvilket afspejles i akronymet PIRO (**P**opulation, **I**ndex test, **R**eference standard and **O**utcomes). Man skal desuden overveje, hvilken rolle den nye test skal spille i den diagnostiske udredning og om der er risiko for overdiagnostisk.

Population: Detaljeret beskrivelse af populationen, herunder sygdommen (sværhedsgrad, symptomer) med relevante demografiske faktorer (fx alder eller køn).. Hvis det er relevant, beskrives hvilken klinisk sammenhæng for testen man er interesseret i (fx almen praksis eller hospital). Præcisionen og balancen mellem gavnlige og skadelige virkninger af tests varierer alt efter om den foretages i den primære, sekundære eller tertiære sektor. Antagelser om prævalensen for den pågældende sygdom i den relevante kliniske sammenhæng skal derfor angives.

Index test: Hvilken test eller eventuelt flere tests ønsker man at vurdere? Herunder specificeres de diagnostiske tærskelværdier (grænsen for adskillelsen mellem syg og rask), hvem der skal udføre og fortolke testen, betingelserne testen udføres under (laboratorium, klinik) kan også specificeres.

Reference standard: Hvad er reference standarden? (Gold standard). Referencestandarden er den test eller de tests som optimalt set anvendes til at bestemme fraværet eller tilstedeværelsen af en given sygdom hos en patient. Referencestandarden vælges ud fra at den er den mest pålidelige procedure til at fastslå tilstedeværelsen af sygdom.

Der SKAL angives en referencestandard, idet fx sensitivitet og specificitet af den undersøgte test ellers ikke kan bestemmes. Hvis der ikke er en referencestandard, eller denne er meget usikker, kan spørgsmålet kun besvares ved hjælp af patientrelevante outcome. I disse tilfælde bruges PICO formatet for formulering af spørgsmål, idet man bliver nødt til at se på patientrelevante outcome, for at opveje gavnlige og skadelige virkninger.

Outcome: Hvilke outcome vurderes at være vigtige for at kunne vurdere gavnlige og skadelige virkninger af den undersøgte diagnostiske test.

Outcome af en diagnostisk test kan bestemmes på to niveauer: ved hjælp af mål for den diagnostiske præcision eller ved patientrelevante outcome. Det mest optimale vil ofte være, hvis man kan finde patientrelevante outcome, men dette er sjældent muligt og man vil da skulle bestemme testens egenskaber ved hjælp af surrogat outcome som sensitivitet og specificitet (diagnostisk præcision). Dette kræver ikke nødvendigvis randomiserede studier, men kan kvantificeres i observationelle studier, hvor patienterne først udsættes for den nye test (index testen) og derefter for den eksisterende test (referencestandard). Hvis det er veldokumenteret, at den eksisterende test forbedrer patientrelevante outcomes, er denne fremgangsmåde ofte den eneste etisk forsvarelige. Se en mere udførlig forklaring af diagnostisk præcision og patientrelevante outcome i de efterfølgende to afsnit.

De fokuserede spørgsmål formuleres i et skema som findes [her](#).

3.1.3 Diagnostisk præcision

De fleste studier om diagnostisk præcision kvantificerer sensitivitet, specificitet, falsk positive og falsk negative værdier af en ny test i forhold til en eksisterende test. Selvom sensitivitet/specificitet osv. er surrogat outcome, kan dette være tilstrækkeligt til at underbygge en anbefaling, hvis referencestandard er meget pålidelig og velunderbygget. Dette er desværre sjældent tilfældet. Hvis referencestandard er usikker, kan de nævnte mål for en tests præcision være vildledende og skal da tolkes med stor forsigtighed. I disse situationer skal man have viden om hvilken effekt sensitiviteten og specificiteten (altså testens evne til at finde de syge og de raske) har på patientrelevante outcome for at kunne give en stærk anbefaling for anvendelse af en given test.

		Referencestandard	
		+	-
Index test	+	Sandt positiv	Falsk positiv
	-	Falsk negativ	Sandt negativ

Sensitivitet: Andel som er syge (jf. referencestandard) og som index testen identificerer som syge (detekterer)

Specificitet: Andel som ikke er syge (jf. referencestandard) og som index testen identificerer som ikke syge.

Hvis anbefalinger baseres på mål for diagnostisk præcision skal gruppen overveje hvad de psykiske, behandlingsmæssige og prognostiske konsekvenser af falsk positive og falsk negative svar er.

Baserer man sig udelukkende på diagnostisk præcision kan man ikke undersøge om index testen faktisk er bedre end referencestandard da idet resultaterne af referencestandard betragtes som "facitlisten". Finder den nye test flere sygdomstilfælde end den eksisterende, vil dette derfor per definition være falsk positive. Ønsker man at undersøge om index testen er bedre end referencestandard er det derfor nødvendigt med patientrelevante outcome.

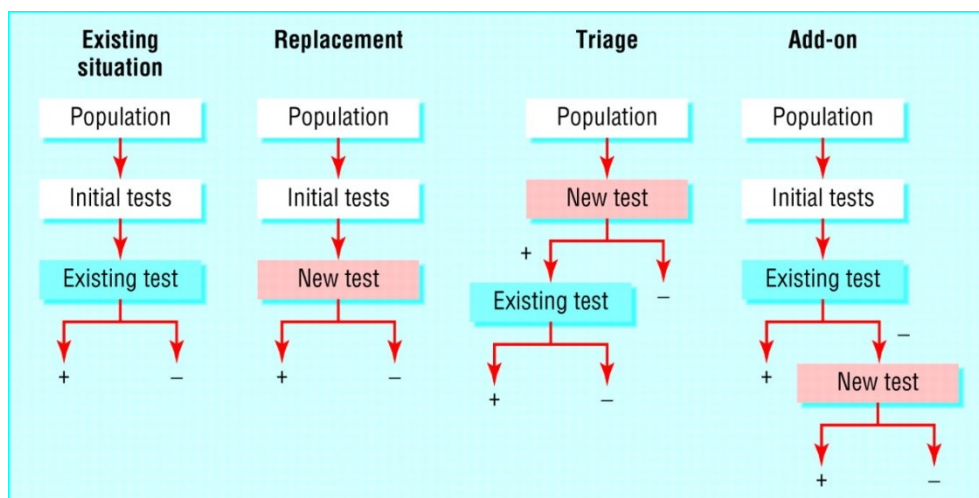
3.1.4 Patientrelevante outcome

Hvilke outcome vurderes at være vigtige for at opveje gavnlige mod skadelige virkninger af den undersøgte intervention? Her skal overvejes relevante outcome for alle beslutningstagere fx klinikerne såvel som patienten. Se afsnit 3.1.1. for mere information

3.1.5 Hvilken rolle skal den nye test spille?

Udover population, indextest, referencestandard og outcome skal man overveje den nye tests rolle i den diagnostiske udredning samt risikoen for overdiagnostik. Det er vigtigt at fastlægge formålet med testen, da dette er med til at bestemme, hvordan man vil vægte betydningen af de forskellige resultater. Der kan være tale om: erstatning, triage eller add-on.

Figur 2 Testens rolle i den diagnostiske udredning



Fra: Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ*. 2006;332:1089-92.

- **Erstatning for eksisterende test:** Den nye test kan være billigere, nemmere eller hurtigere at udføre, have færre gener/skadevirkninger for patienten, etc. Her kan det være tilstrækkeligt at dokumentere, at sensitivitet og specificitet er sammenlignelige med den eksisterende referencetest.
- **Triage:** En test der anvendes til at afgøre hvilke patienter, der skal tilbydes opfølgende diagnostiske tests. Dette kan være relevant hvis den opfølgende test har skadevirkninger, er dyr eller svært tilgængelig, eller besværlig at udføre. Her er det særligt vigtigt at overveje konsekvenserne af falsk negative resultater.
- **Add-on:** En supplerende test der skal forbedre præcisionen af eksisterende metoder til diagnostik. Her er det ligeledes vigtigt at overveje konsekvenserne af falsk negative og falsk positive resultater af den nye test.

3.1.6 Overdiagnostik

En væsentlig udfordring ved diagnostiske test er, at man risikerer at diagnosticere sygdom, som aldrig ville have udviklet sig eller givet symptomer i personens resterende levetid. Da man for en specifik person med et positivt testresultat ikke kan vide, om vedkommende er overdiagnosticeret, bliver alle oftest tilbudt behandling. Denne behandling er dermed per definition overflødig, og da alle behandlinger har skadevirkninger, betyder overdiagnostik derfor, at man vil skade nogle af disse personer.

Ud over de fysiske konsekvenser af overbehandling, som kan optræde på både kort sigt (fx komplikationer til kirurgi) og lang sigt (fx skadevirkninger til medicin), kan det ofte have alvorlige psykiske konsekvenser at få en diagnose, særligt hvis der er tale om overdiagnostik af alvorlige sygdomme som kræft eller udposninger på hovedpulsåren eller arterier i hjernen. Særligt hyppig er overdiagnostik ved tests, der skal finde risikofaktorer til sygdom (fx forhøjet kolesterol, overvægt eller blodtryk), da andelen af personer som udvikler sygdom som følge af disse risikofaktorer ofte er lav. Mål for præcision som sensitivitet og specificitet kan altså sjældent stå alene, da overdiagnostik ikke kan kvantificeres ud fra de gængse mål for en

tests præcision. Ofte rummer en referencestandard mulighed for overdiagnostik, for eksempel patologisvar ved kræftdiagnostik, eller ved psykiatriske lidelser, hvor kriterierne for en diagnose kan ændres over tid og hviler på subjektive bedømmelser med observatørvariation.

Man kan vurdere om overdiagnostik ved en ny test kan være et væsentligt problem ved at forholde sig til 3 forhold:

1: Er den relevante sygdom klart defineret? Dette er tilfældet hvis definitionen ikke har ændret sig over tid, ikke i betydelig grad hviler på subjektive bedømmelser, og at der ikke er betydelig inter-observatør variation i forbindelse med testen.

2: Referencestandard skal være pålidelig, hvilket bedømmes vha. QUADAS-2 (afsnit 3.5.3).

3: Er der fundet mål for diagnostisk præcision (sensitivitet/specificitet) eller patientrelevante outcome? Hvis kriterium 1 og 2 er opfyldt, kan det være tilstrækkeligt med mål for diagnostisk præcision, for at kunne give en stærk anbefaling. Hvis kriterium 1 og 2 ikke er opfyldt, kræves der evidens i form af patientrelevante outcome for at kunne give en stærk anbefaling. Hvis disse krav ikke er opfyldt, kan man kun give en svag anbefaling.

Videre læsning

Cochrane handbook for systematic reviews of interventions:

<http://handbook.cochrane.org/>

Cochrane handbook for systematic reviews of diagnostic test accuracy:

<http://srdta.cochrane.org/handbook-dta-reviews>

GRADE arbejdsgruppens hjemmeside <http://www.gradeworkinggroup.org>

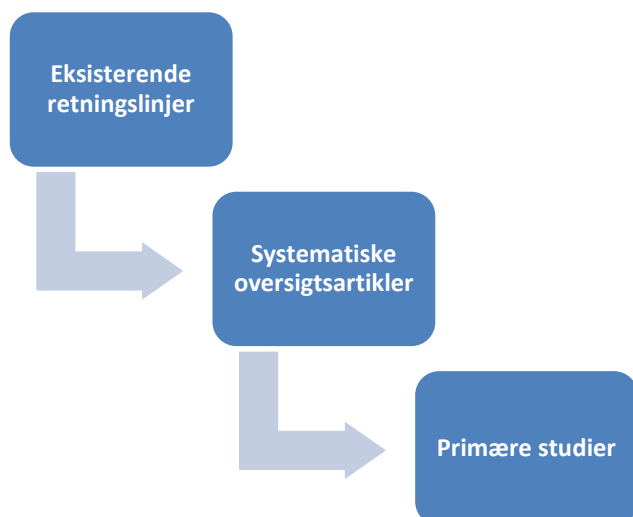
GRADE guidelines: 2. Framing the question and deciding on important outcomes. Guyatt G et al. Journal of Clinical Epidemiology 2011; 64:395-400

COMET (Core Outcome Measures in Effectiveness Trials) <http://www.comet-initiative.org/>

ROBINS-I vejledning <https://sites.google.com/site/riskofbiastool/welcome/home>

3.2 Litteratursøgning

Nationale kliniske retningslinjer skal om muligt bygge på eksisterende viden fra publicerede guidelines, systematiske oversigtsartikler eller randomiserede studier. Derfor er denne triade model valgt.



Figur 3 Trin i litteratursøgningsprocessen

Litteratursøgning efter evidens på de udvalgte PICO's skal ske ved systematiske søgninger, der er dokumenteret ved en søgestrategi, søgeprotokoller og en overordnet søgebeskrivelse. Sundhedsstyrelsen anbefaler, at der indgås aftale med en søgespecialist i forbindelse med søgning efter litteratur. Søgeprocessen planlægges i tæt samarbejde mellem søgespecialisten, fagkonsulenten og arbejdsgruppen.

Det er ikke et krav, at der søges på guidelines eller systematiske oversigtsartikler, men det er en mulighed, hvis man ønsker at følge den triade som Sundhedsstyrelsen har benyttet i de NKR, der er udgivet mellem 2012-2016. Udover at bidrage med estimater kan retningslinjer og systematiske oversigtsartikler bidrage med referencer og evidensvurderinger. Dog kan man risikere at gentagne søgninger og at screening af oversigtsartikler ikke fører til brugbar evidens. Det er således op til den enkelte arbejdsgruppe at vurdere om en sådan triade er relevant.

Hvis man følger triaden søges der først efter eksisterende retningslinjer. Hvis der ikke findes nogen, der ikke besvarer de stillede spørgsmål eller de er af dårlig kvalitet søges dernæst efter systematiske oversigtsartikler. Findes ingen oversigtsartikler, eller de er af dårlig kvalitet, søges efter primærlitteratur. Findes brugbare retningslinjer eller systematiske oversigtsartikler som dog er forældede opdateres deres søgning, det vil sige, der søges efter primærlitteratur publiceret siden deres søgning er afsluttet. Den model er valgt for at sikre, at der ikke udføres arbejde der allerede er gjort, og som vil kunne bruges til at besvare det ønskede PICO. Dog, som nævnt ovenfor, er det ikke altid, at der er brugt GRADE i de identificerede guidelines, som Sundhedsstyrelsen kræver for udarbejdelse af en national klinisk retningslinje. I så fald er det ikke muligt at bruge de pågældende estimater og man må selv i gang.

Da der ofte er en afgrænset tidshorisont for en national klinisk retningslinje, er det vigtigt at afgrænse søgningen på en relevant og hensigtsmæssig måde. Det skal

dog være afgrænsninger der giver mening i forhold til emne og det pågældende PICO eksempelvis:

- Tidsafgrænsning – f.eks. 10 år
- Population – f. eks. kun børn, ældre over 65 år.
- Publikations type - f.eks systematiske reviews, metaanalyser, RCT
- Sprog – f. eks. engelsk, svensk, norsk

Med hensyn til tidsafgrænsningen, skal der dog tages højde for eventuelle udviklinger i interventioner eller diagnostiske tests så der ved behov kan søges længere tilbage. Dette diskuteres individuelt for hvert enkelt fokuseret spørgsmål. Se i øvrigt vejledningen om [den praktiske litteratursøgningsproces](#).

Derefter vælges databaser, som minimum søges der altid i Medline/PubMed og Embase, se listen med de udvalgte databaser på [Sundhedsstyrelsens hjemmeside](#).

Alt efter emne tilføjes relevante databaser. Nogle databaser kan der være adgang til via den institution man tilhører, andre ligger gratis på nettet, og andre kræver medlemskab. Fremskaffelsen af fuldttekst referencerne kan være gebyrbelagte. Vær derfor opmærksom på at søgning og fremskaffelse af referencer kræver et budget.

Upublicerede data kan vanskeligt medtages i evidensvurdering, og i en national klinisk retningslinje vil der som udgangspunkt kun blive medtaget referencer til studier, der er publiceret i et offentligt, peer-reviewed forum, herunder elektroniske og trykte tidsskrifter. For at afdække evt. publikationsbias kan der, hvor relevant, medtages søgning på upublicerede studier i anerkendte databaser som fx clinicaltrials.gov.

Søgetermer og strategi aftales endeligt mellem fagkonsulenten, metodekonsulenten og søgespecialisten.

Der udarbejdes en søgeprotokol med søgestrategier for hvert enkelt PICO. Følger man triaden er der tre søgeprotokoller, én for søgningen på guidelines, én for sekundærlitteratur (systematiske reviews, meta-analyser) og én for primær litteratur (randomiserede kontrollerede forsøg /RCT).

Søgeprotokollen har til formål at skabe overblik og gennemskuelighed i processen vedrørende indsamling af information. Søgeprotokollen dokumenterer på en detaljeret og transparent måde hvad der er søgt efter, hvor der er søgt og hvordan der er søgt og udvalgt vha. eksplicitte og prædefinerede selektionskriterier. Søgeprotokollen er med til at sikre konsistens ved opfølgning eller gentagelse af søgningen, og skal være så detaljeret, at man ved at følge beskrivelsen kan udføre søgningen og få omtrentligt samme resultat. Se skabelon for søgeprotokol på [Sundhedsstyrelsens hjemmeside](#).

Søgeprotokollen bør indeholde følgende elementer:

- Baggrund og problemstilling
- Fokuserede spørgsmål
- Prædefinerede inklusions- og eksklusionskriterier (sprog, tid, population)
- Informationskilder

- Søgestrategi (kombination af søgetermer) og resultat for hver enkelt informationskilde
- Kontaktoplysninger

Der evalueres ved krydscheck undervejs ved hjælp af relevante artikler for at sikre at søgestrategien får det hele med/rammer rigtigt.

3.2.1 Søgestrategi og -beskrivelse

En detaljeret søgestrategi udarbejdes i det dertil indrettede afsnit i protokollen. Søgestrategien skal sammen med protokollen godkendes af Sundhedsstyrelsen inden processen for dataindsamling påbegyndes.

Når søgestrategien er godkendt i Sundhedsstyrelsen kan søgningen påbegyndes. Den ansvarlige søgespecialist udformer en overordnet søgebeskrivelse, som indeholder beskrivelse af hele søgeprocessen, som den er foregået, fra søgningen efter retningslinjer til søgningen per fokuserede spørgsmål til evt. supplerende søgninger.

Selve søgehistorien fra de forskellige databaser kan ses i søgeprotokollen. Denne skal publiceres som et bilag ved offentliggørelse af retningslinjen.

Når litteratursøgnings- og sorteringsprocessen er afsluttet udfyldes et flowdiagram, der dokumenterer, hvor mange referencer der blev fundet og hvilke der blev sorteret fra. I programmet Covidence, som Sundhedsstyrelsen stiller til rådighed, er det muligt at opsætte hele sorteringsprocessen, så den kan kopieres direkte over i flowdiagrammet. Disse flowdiagrammer skal publiceres som bilag ved offentliggørelsen af retningslinjen. Skabelon for flowdiagram findes i protokolskabelonen.

3.3 Obligatoriske programmer til udarbejdelse af retningslinjen

Sundhedsstyrelsen stiller følgende netbaserede programmer til rådighed:

- Covidence (www.covidence.org) til sorteringer og ekstrahering af referencerne.
- MAGIC (www.magicapp.org) til udarbejdelse af retningslinjen.

Alle retningslinjer skal udarbejdes i disse systemer.

Sundhedsstyrelsen anbefaler Refworks til referenceoverførelse fra databaser til Covidence samt Review Manager, et værktøj, hvori man kan udarbejde meta-analyser og lave forest plots.

3.3.1 RefWorks

RefWorks bruges som udgangspunkt til referencehåndtering i processen op til at referencerne importeres i Covidence. RefWorks er et web-baseret program til referencehåndtering, hvor man kan opbygge emneinddelte databaser over litteraturreferencer. Referencer fra stort set alle databaser kan eksporteres direkte til RefWorks eller de kan inddateres manuelt. Herfra kan referencerne overføres til Covidence.

Refworks genererer automatisk litteraturlisterne efter det ønskede format og kan trække referencerne ind som kilder i det pågældende dokument. Når retningslinjen skrives i Magic, kan de udvalgte referencer hentes ind direkte fra RefWorks.

NKR-sekretariatet kan oprette adgang via Sundhedsstyrelsens licens til arbejdsgruppen. En praktisk vejledning til søgespecialisten findes på [Sundhedsstyrelsens hjemmeside](#).

3.3.2 Covidence

Covidence (covidence.org) er et web-baseret referenceprogram, der holder styr på alle de referencer, der er tilknyttet et projekt. Oprindeligt blev det designet til at støtte forskerne, når de skulle skrive reviews. Med designstøtte fra Sundhedsstyrelsen er det nu også i stand til at håndtere guidelines. Covidence styrer referencerne, fra de bliver importeret fra Refworks, over grovsorteringen og upload af fuldtæst, til vurderingsprocessen (Risk of bias), dataekstraheringen og efterfølgende eksport til RevMan og 'MAGIC'. I praksis fungerer Covidence ved, at NKR-sekretariatet registrerer fagkonsulenten i Covidence og laver et login. Dette login bruger både fagkonsulenten og søgespecialisten. Herefter kan man oprette mapper for hvert enkelt PICO hvori man kan placere, administrere og dataekstrahere/risk of bias vurdere sine referencer.

3.3.3 Review Manager

Når man har sorteret, dataekstraheret og risk of bias vurderet sine referencer i Covidence, eksporteres disse til Review Manager til analysebearbejdning, Review Manager er udarbejdet af Cochrane og programmet kan downloades gratis på nettet [her](#). Sundhedsstyrelsen anbefaler, at man benytter Review Manager til udarbejdelse af meta-analyser, da dette værktøj arbejder sammen med både Covidence og Magic. Revman filen kan omformateres til pdf-format.

3.3.4 MAGIC

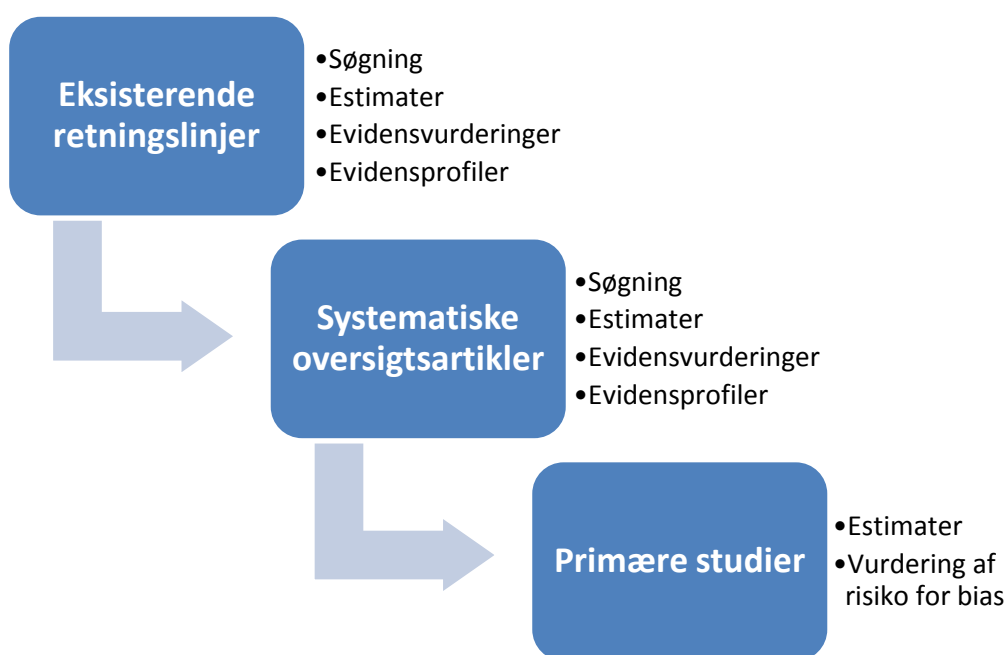
Magic er et netbaseret værktøj, udarbejdet af norske forskere i samarbejde med GRADE. I dette program kan man opsætte og skrive sin retningslinje, samt udarbejde evidensstabeller på baggrund af analyser importeret fra Review Manager. NKR-sekretariatet kan oprette adgang via Sundhedsstyrelsens licens til arbejdsgruppen. Magic er opbygget, så det gøres nemt at følge processen fra evidens til formulering af anbefaling som GRADE anbefaler det. Den færdige retningslinje udgives på Magic's hjemmeside: magicapp.org, som er frit tilgængelig. Der kan desuden printes en autogeneret pdf af retningslinjen.

3.4 Udvælgelse af evidens

Der er flere kilder til evidens. Som nævnt i afsnit 3.2 om søgninger er første skridt at undersøge, om der er eksisterende retningslinjer, der besvarer de fokuserede spørgsmål. Findes der ingen relevante eksisterende retningslinjer søges efter systematiske oversigtsartikler. Findes ingen relevante systematiske oversigtsartikler søges efter primærstudier. Det kan desuden være nødvendigt at søge efter primærlitteratur for at opdatere de fundne retningslinjer eller systematiske oversigtsartikler. Efter hver søgning skal den fundne litteratur sorteres og vurderes. Det skal vur-

deres om den fundne litteratur er af tilstrækkelig høj kvalitet og om de elementer, der skal bruges for at kunne udarbejde en anbefaling er til stede. Retningslinjer og systematiske oversigtsartikler kan kun bruges til at formulere anbefalinger ud fra, hvis de har anvendt GRADE. Retningslinjer/systematiske oversigtsartikler, der ikke har anvendt GRADE, men indeholder estimater og/eller risk of bias vurderinger kan bruges som udgangspunkt for en GRADE vurdering.

- For sekundær litteratur vil dette især sige: Er der søgt systematisk?
- Præsenteres der estimater?
- Er de studier der ligger til grund for estimaterne vurderet med hensyn til risiko for bias (evidensvurdering)?
- Foreligger der evidensprofiler?



Figur 5 Trin i udvælgelse af evidens

En gennemgang af, hvornår en retningslinje er af tilstrækkelig kvalitet findes i afsnit 3.5.1, en tilsvarende gennemgang for systematiske oversigtsartikler findes i afsnit 3.5.2.

Der er udarbejdet et skema, der skal hjælpe til at bevare et løbende overblik over hvilken evidens, der er fundet til at besvare de forskellige fokuserede spørgsmål. Skemaet findes [her](#).

3.4.1 Typer af studiedesign

Det fokuserede spørgsmål afgør hvilke studiedesigns, der skal inkluderes for at besvare spørgsmålet.

Tabel 1 Typer af spørgsmål og eksempler på primære studier, der kan svare på dem

Hvor mange har sygdommen?	Tværsnitsstudier
Hvad er årsager til sygdommen?	Kohortestudier eller case-control studier
Hvad er den bedste måde at fastslå om en person har sygdommen (diagnostik)?	Kohortestudier, tværsnitsstudier, eller randomiserede studier, med en reference standard
Hvad er gavnlige og skadelige virkninger af forebyggelse, screening og behandling (effekt)?	Randomiserede forsøg
Hvordan vil det sandsynligvis gå en patient med en bestemt sygdom (prognose)?	Kohortestudier

Tabel 1 illustrerer de studiedesigns, der som tommelfingerregel skal anvendes når man undersøger forskellige spørgsmål. Det vil sige at hvis man for eksempel har et fokuseret spørgsmål om, hvorvidt man bør bruge intensiv glykæmisk kontrol hos patienter med type 2 diabetes eller konventionel glykæmisk kontrol, vil man først søge efter retningslinjer, der belyser dette systematisk ved at have søgt og vurderet de randomiserede forsøg, der findes på området.

Tabellen dækker ikke alle scenarier. Randomiserede forsøg kan for eksempel have det problem at opfølgingsperioden ikke er lang nok til, at man kan fastslå langtidsvirkninger af interventionen. I sådanne tilfælde kan det være nødvendigt at supplere med information fra kohortestudier eller andre kilder. Dette gælder også ved for eksempel meget sjældne skadevirkninger.

Sundhedsstyrelsen har i udarbejdelsen af de Nationale Kliniske Retningslinjer prioriteret RCT studier, da dette design, er det eneste der kan forebygge selektionsbias og confounding. Man bør således overveje om ens PICO kan besvares med et RCT studiedesign. Er dette tilfældet, vil en evt. videre søgning på observationelle studier sjældent påvirke evidensens styrke. Søgning på observationelle studier, kan dog være relevant hvis det pågældende PICO, f.eks. af etiske årsager, ikke kan besvares med et RCT design.

3.4.2 Sortering af søgeresultater

Når søgeresultater foreligger, skal litteraturen grovsorteres med henblik på at udvælge de retningslinjer, oversigtsartikler eller primære studier, som skal kvalitetsvurderes i arbejdsgruppen.

Sorteringen består af to faser:

1. Grovsortering på emne og indhold (titel og resumé): Hvis det fundne studie åbenlyst ikke er relevant for NKR emnet, frasorteres den.
2. Grovsortering på metodisk kvalitet: Studier af for ringe kvalitet frasorteres med sparring fra metodekonsulenten. Kriterierne for vurdering kan ses i afsnittet 'Vurdering af evidens' under de relevante studiedesigns. For sekun-

dær litteratur skal der i første omgang fokuseres på, om søgningen var systematisk.

De studier der er tilbage efter grovsorteringen bestilles hjem i fuld tekst og finsorteres af to personer uafhængigt af hinanden. Dette skal sikre, at der ikke overses relevante referencer. For at mindske arbejdsbyrden på det enkelte arbejdsgruppemedlem kan opgaven deles mellem flere medlemmer.

3.5 Vurdering af evidens

Hvordan evidensen skal vurderes afhænger af studiets design. I de efterfølgende afsnit gennemgås vurderingen af retningslinjer, systematiske oversigtartikler, randomiserede forsøg, kohorte studier, case-control studier, tværsnitsstudier samt diagnostiske studier.

3.5.1 Vurdering af retningslinjer

Retningslinjer som er blevet udvalgt via fuldtekstsorteringen, vurderes af to personer uafhængigt af hinanden ved hjælp af AGREE II værktøjet (<http://www.agreetrust.org>).

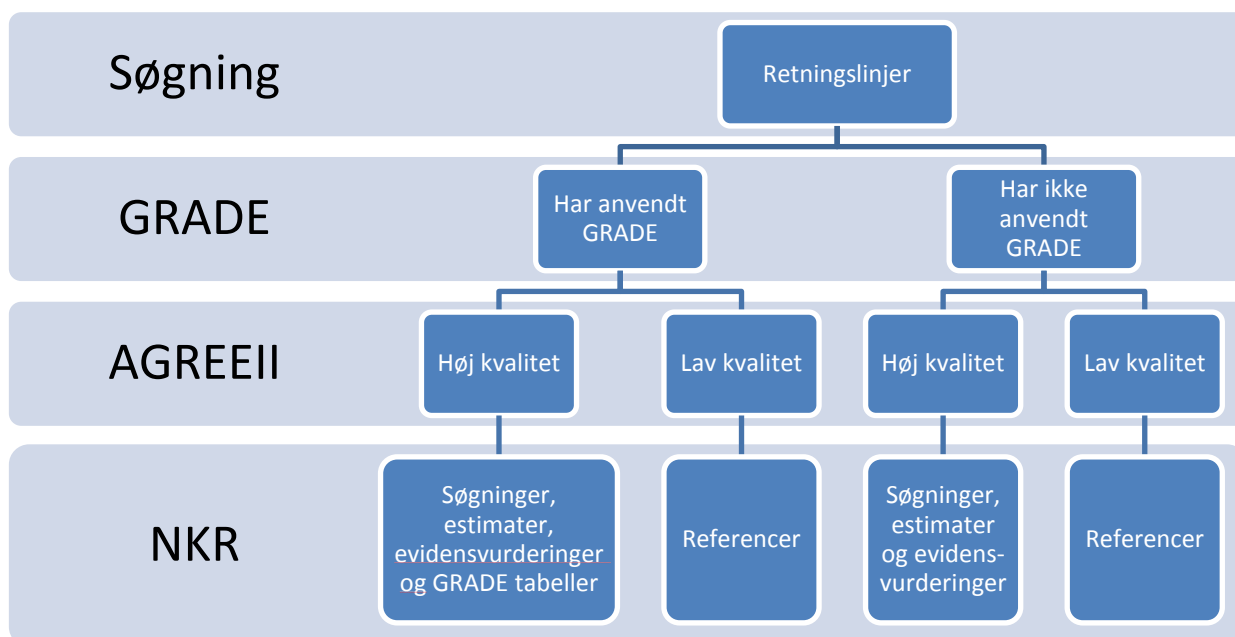
Fagkonsulenten vurderer alle udvalgte retningslinjer og koordinerer hvilke arbejdsgruppemedlemmer, der vurderer hvilke retningslinjer. I tilfælde af uenighed om vurderingen kontakter fagkonsulenten arbejdsgruppemedlemmet og konsensus søges opnået ved hjælp af diskussion. I de tilfælde hvor dette ikke er muligt, inddrages metodekonsulent og/eller arbejdsgruppe.

AGREE II har udviklet My AGREE PLUS, et online værktøj, der bidrager til en systematisk gennemgang af retningslinjens kvalitet. Der findes hjælp og videoer her: (<http://www.agreetrust.org/help/>). Fagkonsulenten får hjælp af værktøjet til at koordinere processen og har mulighed for fx at sende reminders.

En eksisterende retningslinje kan potentielt bruges som kilde til:

- Supplerende referencer
- Søgninger
- Evidensvurderinger
- Estimer

Figur 6 Diagram, der illustrerer processen i vurderingen af retningslinjer



De udvalgte retningslinjer gennemgås af fagkonsulenten med henblik på at vurdere, hvilke af fokuserede spørgsmål, de kan bruges til at besvare. For at kunne vurdere, hvorvidt retningslinjen er af høj eller lav kvalitet, bør der fokuseres på hvordan retningslinjen vurderes i domæne 3 'Rigour of development' i AGREE II værktøjet. Domænet består af følgende komponenter:

AGREE II Domain 3. Rigour of development

- 7. Systematic methods were used to search for evidence.
- 8. The criteria for selecting the evidence are clearly described.
- 9. The strengths and limitations of the body of evidence are clearly described.
- 10. The methods for formulating the recommendations are clearly described.
- 11. The health benefits, side effects, and risks have been considered in formulating the recommendations.
- 12. There is an explicit link between the recommendations and the supporting evidence.
- 13. The guideline has been externally reviewed by experts prior to its publication.
- 14. A procedure for updating the guideline is provided.

Nedenstående tabel viser, hvad der kræves for at en retningslinje kan bruges som kilde til de forskellige elementer (supplerende referencer, søgninger, evidensvurderinger, estimer):

<p>Kilde til supplerende referencer</p> <p>Kan bruges så snart retningslinjen dækker relevante fokuserede spørgsmål</p>
--

Kilde til søgninger

Kan bruges når følgende områder er tilfredsstillende dækket:

7. Systematic methods were used to search for evidence.
8. The criteria for selecting the evidence are clearly described.

Kilde til evidensvurderinger

Kan bruges når følgende områder er tilfredsstillende dækket:

7. Systematic methods were used to search for evidence.
8. The criteria for selecting the evidence are clearly described.
9. The strengths and limitations of the body of evidence are clearly described.
12. There is an explicit link between the recommendations and the supporting evidence.

Kilde til estimer

Kan bruges når, hvis meta-analyser er udført og følgende områder er tilfredsstillende dækket:

7. Systematic methods were used to search for evidence.
8. The criteria for selecting the evidence are clearly described.
9. The strengths and limitations of the body of evidence are clearly described.
12. There is an explicit link between the recommendations and the supporting evidence.

Hvis der anvendes retningslinjer i forbindelse med udarbejdelse af den nationale kliniske retningslinje, kan man benytte skemaet 'Overblik over retningslinjer', skemaet findes på [Sundhedsstyrelsens hjemmeside](#). Skemaet hjælper med at bibeholde overblikket over hvilke retningslinjer, der endeligt anvendes og hvad der anvendes fra den enkelte retningslinje. Der skal for hver retningslinje afrapporteres hvilken AGREE II score, der blev opnået, både samlet og særskilt for domæne 3 'Rigour of development'. Det skal anføres om retningslinjen er fundet egnet til anvendelse i NKR'en og om retningslinjen har anvendt GRADE samt dato for den seneste litteratursøgning i retningslinjen.

Såfremt retningslinjen kan bruges til andet end blot supplerende referencer, tager fagkonsulenten nu stilling til hvilke outcome, der er dækket i den pågældende retningslinje. Dette kan med fordel noteres i skemaet: 'Overblik over evidens' som Sundhedsstyrelsen selv anvender som procesværktøj i forbindelse med udarbejdelse af nationale kliniske retningslinjer. Skemaet findes på [Sundhedsstyrelsens hjemmeside](#). Processen gentages for alle egnede retningslinjer. Dette giver et overblik over, hvilke fokuserede spørgsmål der er dækkede og for hvilke, det er nødvendigt at gå videre i søgeprocessen.

Videre læsning:

AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: The AGREE project. QualSaf Health Care. 2003(1):18-23.

Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 1: Performance, usefulness and areas for improvement. CMAJ. 2010(10):1045-52.

Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 2: Assessment of validity of items and tools to support application. CMAJ. 2010(10):E472-8.

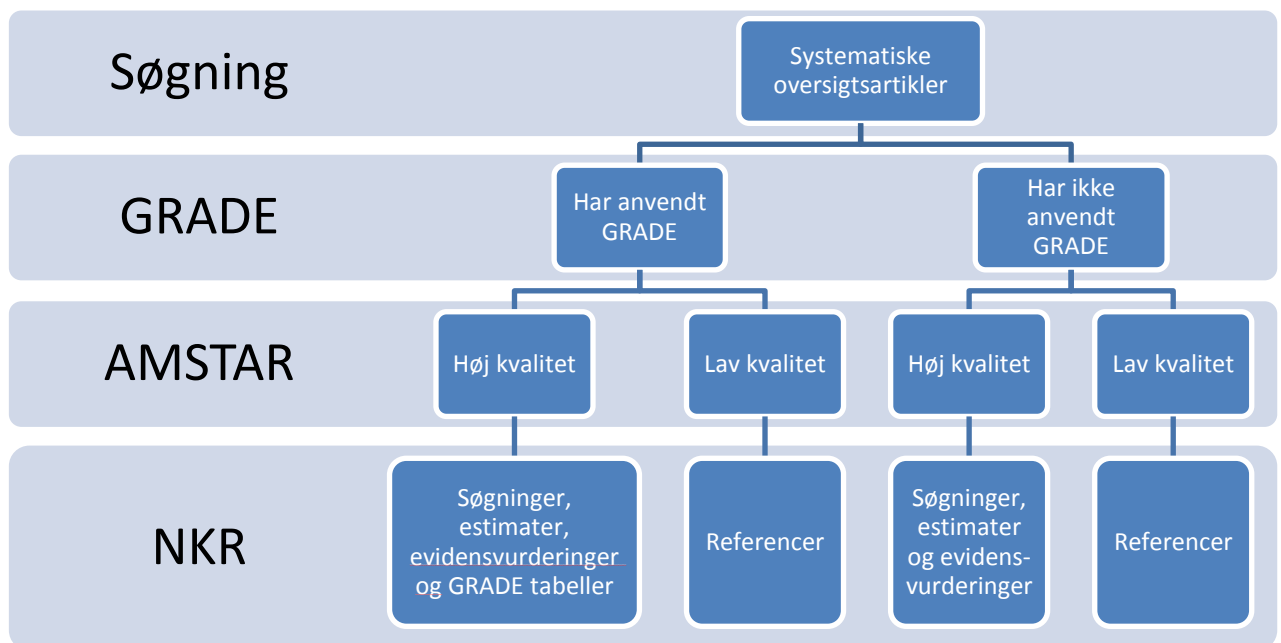
3.5.2 Vurdering af systematiske oversigtsartikler

Systematiske oversigtsartikler vurderes af to personer uafhængigt af hinanden ved hjælp af AMSTAR værktøjet (<http://amstar.ca/>). I tilfælde af uenighed om vurderingen søges konsensus opnået ved hjælp af diskussion. I de tilfælde hvor dette ikke er muligt inddrages en tredje person eller arbejdsgruppen. AMSTAR er endnu ikke onlinebaseret og en vurdering udfyldes derfor i et excel ark. Arket findes på [Sundhedsstyrelsens hjemmeside](#).

En systematisk oversigtsartikel kan potentielt bruges som kilde til:

- Supplerende referencer
- Søgninger
- Evidensvurderinger
- Estimer

Figur 7 Diagram, der illustrerer processen i evidensvurdering af systematiske oversigtsartikler



De egnede systematiske oversigtsartikler gennemgås af fagkonsulenten med henblik på, hvilke af fokuserede spørgsmål de kan bruges til besvare. For at kunne

vurdere dette skal der fokuseres på, hvordan retningslinjen vurderes i forskellige komponenter af AMSTAR værktøjet:

AMSTAR

1. Was an ‘‘a priori’’ design provided?
2. Was there duplicate study selection and data extraction?
3. Was a comprehensive literature search performed?
4. Was the status of publication (i.e., grey literature) used as an inclusion criterion?
5. Was a list of studies (included and excluded) provided?
6. Were the characteristics of the included studies provided?
7. Was the scientific quality of the included studies assessed and documented?
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
9. Were the methods used to combine the findings of studies appropriate?
10. Was the likelihood of publication bias assessed?
11. Were potential conflicts of interest included?

Nedenstående tabel viser, hvad der kræves for at en systematisk oversigtartikel kan bruges som kilde til de forskellige elementer:

Kilde til supplerende referencer

Kan bruges så snart den systematiske oversigtsartikel dækker relevante fokuserede spørgsmål

Kilde til søgninger

Kan bruges når der er klare in- og eksklusionskriterier og følgende område er tilfredsstillende dækket:

3. Was a comprehensive literature search performed?

Kilde til evidensvurderinger

Kan bruges når der er klare in- og eksklusionskriterier og følgende områder er tilfredsstillende dækket:

3. Was a comprehensive literature search performed?
7. Was the scientific quality of the included studies assessed and documented?

Kilde til estimer

Kan bruges når, hvis meta-analyser er udført og følgende områder er tilfredsstillende dækket:

3. Was a comprehensive literature search performed?
9. Were the methods used to combine the findings of studies appropriate?

Hvis der anvendes systematiske oversigtsartikler i forbindelse med udarbejdelse af den nationale kliniske retningslinje, kan man benytte skemaet 'Overblik over systematiske oversigtsartikler', skemaet findes på [Sundhedsstyrelsens hjemmeside](#). Skemaet hjælper med at bibeholde overblikket over hvilke systematiske oversigts-

artikler, der endeligt anvendes og hvad der anvendes fra den enkelte systematiske oversigtsartikel. Der skal for hver systematiske oversigtsartikel afrapporteres hvilken AMSTAR score, der blev opnået. Det skal anføres om oversigtsartiklen er fundet egnet til anvendelse i NKR'en og om den har anvendt GRADE samt dato for den seneste litteratursøgning i oversigtsartiklen.

Såfremt oversigtsartiklen kan bruges til andet end blot supplerende referencer, tager fagkonsulenten nu stilling til hvilke outcome, der er dækket i den pågældende oversigtsartikel. Dette kan med fordel noteres i skemaet: 'Overblik over evidens' som Sundhedsstyrelsen selv anvender som procesværktøj i forbindelse med udarbejdelse af nationale kliniske retningslinjer. Skemaet findes [her](#), og processen gentages for alle egnede oversigtsartikler. Dette giver et overblik over, hvilke fokuserede spørgsmål der er dækkede og for hvilke, det er nødvendigt at gå videre i søgeprocessen

3.5.3 Vurdering af primære studier

Individuelle studiers risiko for bias kan vurderes ved hjælp af checklister. I tabel 2 ses de checklister der anvendes i arbejdet med NKR.

Tabel 2 Risk of bias checklister for primære studier

Studiedesign	Checkliste	Reference
Randomiserede forsøg	Cochrane Risk of Bias Tool	http://handbook.cochrane.org/
Observationelle studier (kohorte, case-control)	ROBINS-I	https://sites.google.com/site/riskofbiastool/
Diagnostiske primære studier	QUADAS-2	http://www.bris.ac.uk/quadas/quadas-2/
Prognostiske primære studier	QUIPS	https://www.researchgate.net/publication/235658860_Assessing_Bias_in_Studies_of_Prognostic_Factors

Evidensvurdering bør altid foretages af to personer uafhængigt af hinanden. Se afsnit 3.6.1 for flere detaljer om dataekstraktionen og for skemaer. Efter at data er ekstraheret og der er opnået konsensus imellem de to observatører overføres beskrivelser af de inkluderede studier, risiko for bias vurderingen og resultaterne fra studierne til bearbejdning mhp. udførelse af en meta-analyse. Sundhedsstyrelsen har i disse sammenhænge anvendt Review Manager (RevMan). RevMan bruges til at samle information om primære studier, lave risk of bias figurer og meta-analyser. Den planlagte metode for ekstrahering af beskrivelser fra de enkelte studier, data og udførelsen af meta-analyser skal tydeligt være beskrevet i protokollen som indsendes til godkendelse i Sundhedsstyrelsen før opstart af databehandlingen (her under søgning af litteratur).

RevMan er anvendt af blandt andet Cochrane til udarbejdelsen af Cochrane Reviews. Programmet er gratis og kan downloades [her](#)

Videre læsning:

Lundh A, Gøtzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol. 2008 Apr 21;8:22.

[Bai A, Shukla VK, Bak G, Wells G. Quality Assessment Tools Project Report. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2012.](#)

QUADAS-2 [Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang](#)

[MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155\(8\):529-36.](#)

Cochrane handbook: Julian PT Higgins and Sally Green
<http://handbook.cochrane.org/>

3.6 Sammenfatning af evidens

Ved hjælp af litteratursøgnings- og sorteringsprocessen beskrevet i de foregående afsnit, har man identificeret de studier, der danner evidensgrundlaget for anbefalingerne. I dette afsnit gennemgås hvorledes evidensen sammenfattes.

3.6.1 Dataekstraktion

Dataekstraktion vil sige, at man henter relevant information ud af studier. Hvad der er relevant information defineres i de fokuserede spørgsmål. Der bruges forskellige systemer til dataekstraktion baseret på om der er tale om retningslinjer, systematiske oversigtsartikler, randomiserede forsøg, kohorte studier, case-kontrol studier eller diagnostiske studier. Dataekstraktionen skal foretages af to personer uafhængigt af hinanden. Uenigheder løses ved diskussion eller ved at inddrage en tredjepart

3.6.2 Retningslinjer

Såfremt den fundne retningslinje er opdateret kan estimater direkte anvendes til brug i den nationale kliniske retningslinje. Hvis en supplerende søgning har afsløret nyere studier som nødvendiggør en opdatering af estimatet foretages en vurdering af hvordan dette bedst gøres. Studierne, der blev fundet ved den supplerende søgning, skal evidensvurderes og al relevant data og information skal ekstraheres. Dernæst skal retningslinjens meta-analyse opdateres med de nye data.

3.6.2.1 Systematiske oversigtsartikler

Såfremt den fundne systematiske oversigtsartikel er opdateret kan estimater direkte anvendes til brug i den nationale kliniske retningslinje. Hvis en supplerende søgning har afsløret nyere studier som nødvendiggør en opdatering af estimatet foretages en vurdering af hvordan dette bedst gøres. Studierne, der blev fundet ved den supplerende søgning, skal evidensvurderes og al relevant data og information skal ekstraheres. Dernæst skal retningslinjens meta-analyse opdateres med de nye data.

3.6.2.2 Primære studier

For randomiserede forsøg og observationelle studier benyttes online værktøjet [Covidence](#). Der er to danske vejledninger på [Sundhedsstyrelsens hjemmeside](#). Til ekstraktion af andre studiedesigns anvendes et andet værktøj eller dataekstraktions ark (fx i Excel).

(se nedenstående skema for links til dataekstraktion af diverse studiedesigns).

Tabel 3 Oversigt over skemaer/værktøjer til dataekstraktion af primære studier

Studiedesign	Skema
Randomiserede forsøg	Link
Observationelle studier (kohorte, case-control)	Link
Diagnostiske primære studier	Link

Prognostiske primære studier	Link
------------------------------	----------------------

3.6.3 Sammenfatning af resultater fra interventionsstudier

Resultaterne fra de primære studier kan præsenteres deskriptivt (narrativt) eller om muligt i en meta-analyse (kvantitativ opsummering). Det er ikke altid tilrådeligt at lave en meta-analyse. Studier kan være for uensartede til at det giver mening at sammenfatte dem i en meta-analyse. Dette kan skyldes at for eksempel populationerne eller interventionerne i de enkelte studier er for forskellige. Der kan også være tale om at studierne design er for forskellige. I sådanne tilfælde frarådes det også at lave en simpel optælling over hvor mange studier der støtter en given intervention og hvor mange der fraråder den, da denne form for analyse antager at studier er fuldstændigt ens med hensyn til deres størrelse (antal patienter eller events) og deres metodologiske kvalitet. Det er derfor nødvendigt mere detaljeret at beskrive hvad de enkelte studier fandt med hensyn de inkluderede outcome, hvor store de er og hvad deres metodologiske kvalitet er. Et eksempel på en narrativ analyse findes på [Sundhedsstyrelsens hjemmeside](#). Meta-analyser kan laves i [RevMan](#) eller lignende statistiske værktøjer.

Hvis data vurderes at være tilstrækkeligt homogene på tværs af studier er det muligt at lave meta-analyser. Principperne for meta-analyser, som beskrevet i [Cochrane Handbook for Systematic Reviews of Interventions](#), følges. Den valgte metode for udarbejdelsen af metaanalyser skal tydeligt være beskrevet i den protokol der inden opstart af arbejdet med den nationale kliniske retningslinje skal indsendes og godkendes af Sundhedsstyrelsen.

Det anbefales at der udfyldes en outcome-matrix, der beskriver hvilke studier, der har rapporteret hvilke outcomes, dette for at holde overblik over evidensen. Se et eksempel på en outcome-matrix på [Sundhedsstyrelsens hjemmeside](#).

3.6.4 Sammenfatning af resultater fra diagnostiske primærstudier

Resultaterne fra studier af diagnostiske tests kan præsenteres deskriptivt (narrativt), som en meta-analyse (kvantitativ opsummering), som en grafisk præsentation eller som en kombination af disse.

Deskriptiv præsentation: Skabelonen til den deskriptive sammenfatning findes i ekstraktionsarket, som findes på [Sundhedsstyrelsens hjemmeside](#).

Grafisk præsentation: Der er to forskellige måder hvorpå man grafisk kan præsentere resultaterne fra studier af diagnostiske tests. Ved meta-analyser af studier af diagnostiske tests præsenteres resultater for sensitivitet og specificitet. Ydermere præsenteres tal for sandt positive, falsk positive, sandt negative og falsk negative for hvert studie.

Meta-analyser: Det er ikke altid tilrådeligt at lave en meta-analyse til at vurdere sensitivitet og specificitet. Observationelle studier er ofte mere uensartede end randomiserede forsøg, og variationen mellem estimerne fra de indgående studier vil oftere være så stor, at det ikke giver mening at sammenfatte dem i en meta-analyse. Den valgte metode for udarbejdelsen af meta-analyser skal tydeligt være beskrevet i protokollen, der skal godkendes af Sundhedsstyrelsen.

Videre læsning:

Cochrane Handbook Systematic Reviews of Interventions

<http://handbook.cochrane.org/>

Cochrane Handbook for Diagnostic Test Accuracy

<http://srdta.cochrane.org/handbook-dta-reviews>

3.7 Gradering af evidens (GRADE)

Når man systematisk har gennemgået den identificerede litteratur, der relaterer sig til de fokuserede spørgsmål, skal man vurdere kvaliteten af evidensen og gradere styrken af anbefalingerne.

GRADE (Grading of Recommendations Assessment, Development and Evaluation system) er et internationalt brugt systematisk og transparent system til at vurdere kvaliteten af evidens og til at give anbefalinger. Den følgende tekst vil gennemgå systemet i nærmere detaljer. For mere information henvises til [GRADE working group](#), hvor der blandt andet er links til en artikelserie i Journal of Clinical Epidemiologi, som gennemgår de forskellige aspekter af GRADE processen.

GRADE-processen starter med de fokuserede spørgsmål, hvor også alle de kritiske og vigtige effektmål er specificerede. Efter at evidensen er identificeret, giver GRADE eksplicite kriterier, hvormed man kan vurdere kvaliteten af evidensen ud fra fire kriterier fra meget lav til høj. Kriterierne inkluderer: risiko for bias, inkonsistens, unøjagtighed, hvor indirekte evidensen er samt risikoen for publikations bias.

Graderingen af evidensen sker i to etaper. Først graderes tiltroen til estimerne af effekten for de enkelte outcome. Herefter graderes det overordnede niveau af evidens for det fokuserede spørgsmål.

3.7.1 Vurdering af tiltroen til det enkelte outcome

For hvert fokuseret spørgsmål foretages en separat vurdering af evidensen. Såfremt der foreligger data for flere relevante outcome (fx død, myokardieinfarkt, apoplexi, blødning) lægges der særligt vægt på de for patienterne mest kritiske outcome. Eksempelvis er død og myokardieinfarkt mere patientrelevante end kolesterolniveau eller carotis intima tykkelse. Der bør altid medtages minimum én skadesvirkning i vurderingen.

For hvert outcome foretages en individuel evidensvurdering. I GRADE opererer man med fire niveauer for kvaliteten af evidensen:

Høj (⊕⊕⊕⊕)

Vi er meget sikre på, at den sande effekt ligger tæt på den estimerede effekt.

Moderat (⊕⊕⊕○)

Vi er moderat sikre på den estimerede effekt. Den sande effekt ligger sandsynligvis tæt på denne, men der er en mulighed for, at den er væsentligt anderledes.

Lav (⊕⊕○○)

Vi har begrænset tiltro til den estimerede effekt. Den sande effekt kan være væsentligt anderledes end den estimerede effekt

Meget lav (⊕○○○)

Vi har meget ringe tiltro til den estimerede effekt. Den sande effekt vil sandsynligvis være væsentligt anderledes end den estimerede effekt.

For behandlingsinterventioner har man som udgangspunkt høj tiltro til evidens fra randomiserede forsøg. Denne tiltro kan nedgraderes efter en gennemgang af studierne risiko for bias, konsistensen af resultaterne på tværs af studier, den samlede

præcision og størrelsen af effektestimater, hvor direkte de inkluderede studier svarer på det fokuserede spørgsmål samt risikoen for publikationsbias. Estimater fra observationelle studier har man som udgangspunkt lav tiltro til. Dog kan tiltroen til estimater fra velgennemførte observationelle studier opgraderes hvis der for eksempel er store effekter eller klare dosis-respons sammenhænge, og der samtidig ikke er foretaget nedgraderinger.

Tabel 4 Ned- og opgradering af evidens for effekt af interventioner (GRADE)

Evidensniveau udgangspunkt	Studiedesign	Nedgradering	Opgradering	Endeligt evidensniveau
Høj	Randomiserede forsøg	<ul style="list-style-type: none"> • risiko for bias • inkonsistens • indirekte evidens • unøjagtighed • publikationsbias 	<ul style="list-style-type: none"> • effektstørrelse • dosis-respons • konfounding 	Høj
				Moderat
Lav	Observationelle studier			Lav
				Meget lav

For diagnostiske tests vurderes observationelle studier dog som udgangspunkt som værende af høj pålidelighed. Dette er fordi det i nogle situationer ikke er nødvendigt med randomiserede studier for at sammenligne en ny test med en referencestandard med velkendte egenskaber

Tabel 5 Nedgradering af evidens for præcision af diagnostiske tests (GRADE)

Evidensniveau udgangspunkt	Studiedesign	Nedgradering	Opgradering	Endeligt evidensniveau
Høj	Kohortestudier eller randomiserede studier, med en reference standard	<ul style="list-style-type: none"> • risiko for bias • inkonsistens • indirekte evidens • unøjagtighed • publikationsbias 	Ej relevant	Høj
				Moderat
				Lav
				Meget lav

3.7.2 Nedgradering af evidensen for effekt af interventioner

Evidensen kan nedgraderes et til to evidensniveauer for hvert af følgende domæner:

- risiko for bias
- inkonsistens
- indirekte evidens
- unøjagtighed
- publikationsbias

For hvert domæne kan en nedgradering foretages. Er problemet mindre nedgraderes et niveau (eksempelvis fra høj til moderat) og er problemet stort nedgraderes det to niveauer (eksempelvis fra høj til lav). Nedenfor gennemgås kort de forskellige domæner, for en nærmere beskrivelse se de forskellige artikler i GRADE serien i Journal of Clinical Epidemiology, artiklerne findes [her](#).

Risiko for bias

Findes der retningslinjer eller systematiske oversigtsartikler, hvor der er foretaget en relevant vurdering af bias i primærstudierne, kan disse overtages. Ellers bruges Cochrane's *Risk of Bias Tool* til at vurdere randomiserede forsøg ROBINS-I til observationelle studier og QUADAS-2 til diagnostiske studier (se afsnit 3.5.3). Eksempelvis vil et randomiseret forsøg med manglende blinding, risiko for selektionsbias grundet manglende skjult allokering og stort frafald nedgraderes to niveauer, hvorimod et randomiseret forsøg alene med moderat frafald vil nedgraderes et niveau. Selektiv rapportering af outcome (kun de mest positive rapporteres) kan også være en bias, som bør vurderes.

Inkonsistens

Hvis der er inkonsistens i resultaterne fra forskellige studier nedgraderes evidensen. Inkonsistens kan skyldes forskelle i populationer, interventioner, behandling i kontrolgruppen eller definition af effekter. Er der en god forklaring, der opfylder subgruppe kriterierne, herunder at de stemmer med få og a priori definerede hypoteser, så nedgraderer man ikke, men vil snarere differentiere de anbefalinger man vil komme med til forskellige subgrupper.

Indirekte evidens

Relaterer evidensen sig ikke direkte til ens kliniske spørgsmål foretages en nedgradering. Indirekte evidens kan have to baggrunde: 1. man ønsker at sammenligne to behandlinger overfor hinanden, men de enkelte behandlinger er kun sammenlignet overfor placebo. 2. Der er forskelle i population, intervention, kontrolgruppe eller den måde effekterne er målt på mellem det kliniske spørgsmål og de tilgrundliggende studier. Eksempler: Spørgsmålet relaterer sig til ældre med multiple komorbiditeter, men evidensen stammer fra midaldrende relativt raske personer. Spørgsmålet relaterer sig til kirurgisk behandling på et provinshospital i et udviklingsland, men evidensen relaterer sig til behandling på et højtspecialiseret universitetshospital i et industrialiseret land. Spørgsmålet relaterer sig til 40 mg simvastatin behandling i kontrolgruppen, men evidensen stammer fra studier med 10 mg simvastatin. Spørgsmålet relaterer sig til diabetiske komplikationer, men evidensen stammer fra effekt på blodsukker.

Unøjagtighed

Unøjagtighed (imprecision) handler om hvorvidt vi på baggrund af de enkelte outcomes effektestimater og tilhørende konfidensinterval, kan formulere en anbefaling vi har tiltro til. Hvis effekt-estimatet er unøjagtigt, dvs. konfidensintervallet er bredt, foretages en nedgradering. Man bør især overveje at nedgradere hvis man ville komme med modsatrettede anbefalinger i hver sin ende af konfidensintervallet. Det vurderes desuden, om konfidensintervallet overlapper den mindste relevante forskel. Der vil også som udgangspunkt blive nedgraderet, hvis der kun er et studie. Der tages højde for alle outcome, men i formulering af anbefalingen tages stadig stilling til om en evt. unøjagtighed er fundet ved de kritiske eller de vigtige outcome.

Publikationsbias

Hvis der er tegn på publikationsbias (manglende publicering af hele studier) eller selektiv rapportering af effekter (kun de mest positive rapporteres) foretages en nedgradering. Det ses eksempelvis, hvis der kun er publiceret studier, der anvender metoder til at måle outcome, som ikke svarer til, hvad man vil forvente at anvende (eksempelvis brug af en atypisk depressionsskala).

3.7.3 Opgradering

Evidensen fra veludførte observationelle studier kan opgraderes et til to niveauer inden for følgende domæner:

- effektstørrelse
- dosis-respons
- konfounding

For hvert domæne kan der foretages en opgradering. Styrker domænet evidensen lidt, opgraderes det et niveau (eksempelvis fra lav til moderat), og styrker det evidensen meget, opgraderes det to niveauer (eksempelvis fra lav til høj).

Effektstørrelse

Hvis effekten i et observationelt studie er stor, kan evidensen opgraderes et niveau, og hvis den er meget stor to niveauer. Der er her tale om helt ekstraordinære tilfælde, hvor der er tale om en faktor fem til ti relativ forskel mellem grupperne.

Dosis-respons

Hvis der ses en dosis-respons sammenhæng kan evidensen opgraderes. Eksempelvis doseringen af AK-behandling og blødningsrisiko.

Konfounding

Eventuelt konfounding vil bidrage til at underestimere effekten. Hvis eksempelvis en given behandling kun gives til de mest syge patienter, vil manglende kontrol for sygdomsgrad påvirke, at effekten på mortalitet underestimeres.

3.7.4 Nedgradering af evidens for præcision af diagnostiske tests

Tværsnittsstudier og kohortestudier blandt patienter med uafklaret diagnose betragtes som udgangspunkt som havende høj pålidelighed. Evidensen kan nedgraderes til moderat, lav eller meget lav efter vurdering af følgende domæner:

- risiko for bias
- inkonsistens
- indirekte evidens
- unøjagtighed
- publikationsbias

Vurderes tiltroen at blive påvirket i mindre grad, nedgraderes et niveau (eksempelvis fra høj til moderat) og vurderes tiltroen at blive påvirket betydeligt, nedgraderes to niveauer (eksempelvis fra høj til lav).

Risiko for bias

Findes der retningslinjer eller systematiske oversigtsartikler, hvor der er foretaget en relevant vurdering af bias i primærstudierne, kan disse overtages. Ellers bruges QUADAS-2 (se afsnit 3.5.3). Eksempelvis vil manglende blinding af dem som fortolker resultaterne af index testen og reference standarden medføre, at der nedgraderes et niveau, da bedømmelsen af testresultaterne ikke foretages uafhængigt og dermed kan påvirke hinanden. Der kan også nedgraderes hvis ikke alle inkluderede patienter modtog både index test og referencestandard, hvis den diagnostiske usikkerhed for de inkluderede patienter ikke er oplyst, eller hvis det vurderes, at referencestandardet ikke pålideligt kan klassificere patienterne, fx fordi den fører til overdiagnostik.

Inkonsistens

Hvis der er inkonsistens mellem resultaterne fra forskellige studier nedgraderes tiltroen til det samlede resultat. Inkonsistens kan skyldes forskelle mellem populationer, den præcise type af test der undersøges, eller hvilke cut-off værdier for testen der er valgt.

Indirekte evidens

Relaterer evidensen sig ikke direkte til PIRO-spørgsmålet, foretages en nedgradering af tiltroen til estimaterne. Der kan være flere årsager, for eksempel at der er klinisk relevante forskelle mellem studiepopulation og populationen der skal udarbejdes anbefalinger for, hvis testen i studierne og i PIRO-spørgsmålet ikke er identiske, eller hvis den der udfører testen i studierne har særlig ekspertise, som ikke kan forventes af klinikere i praksis.

Hvis formålet med det fokuserede spørgsmål er at sammenligne to index tests med hinanden, er det mest pålidelige studiedesign en sammenligning mellem begge index tests med en referencestandard indenfor samme forsøg. Hvis man ikke finder et sådant forsøg, men de to index tests er sammenlignet med samme referencestandard i separate forsøg, skal man betragte sammenligningen som indirekte og nedgradere tiltroen til estimaterne.

Sensitivitet og specificitet er surrogat outcome, der ikke direkte udtrykker effekten af en test på patientrelevante outcome. Dette betyder, at det kan være relevant at nedgradere på grund af den indirekte kobling herimellem. Der kan dog være solid evidens for at referencestandarden har effekt på patientrelevante outcomes, og det vil da ikke være nødvendigt at nedgradere. Hvis der er usikkerhed omkring pålideligheden af referencestandarden skal anbefalingen normalt være svag.

Unøjagtighed i studier af diagnostiske test

Ved diagnostiske tests studier skal man også vurdere bredden af konfidensinterval-
lerne for estimerne for den diagnostiske præcision – sensitivitet og specificitet -
hvis disse er tilgængelige, eller sandt og falsk positive og negative rater, og ned-
gradere herfor hvis relevant.

Publikationsbias

Hvis der er tegn på publikationsbias (manglende publicering af hele studier) eller
selektiv rapportering af effekter (ikke alle forventelige outcomes rapporteres) fore-
tages en nedgradering.

Opgradering af tiltroen til estimer i studier af diagnostiske tests.

Det er i øjeblikket omdiskuteret, om og hvornår det kan være relevant at opgradere
tiltroen til studier af diagnostiske tests.

3.7.5 Evidensprofiler (interventionstudier)

Formålet med evidensprofiler er at give et overblik over evidensen for et givent fokuseret spørgsmål. Evidensprofiler udarbejdes i [MAGICapp](#) som stilles til rådighed af Sundhedsstyrelsen.

Nedenfor præsenteres et tænkt eksempel på en evidensprofil og en forklaring af de forskellige elementer. Da evidensprofiler er ment til at give et overblik over resultaterne, estimaterne og evidensniveauet for det enkelte fokuserede spørgsmål.

Tabel 6 Evidensprofil - interventionsstudier

NKR Eksempel PICO 1 Kaffe eller te til studerende							
Population: Studerende							
Intervention: Kaffe							
Sammenligning: Ingen kaffe							
Outcome (Tidsramme)	Absolut effekt* (95% CI)			Relativ effekt 95% CI	Antal deltagere (studier)	Evidensniveau (GRADE)	Kommentarer
	Ingen kaffe	Kaffe	Forskel med kaffe				
Antal dumpet (Failed their exam) (Endt intervention)	184 per 1000	138 per 1000 (108 to 174)	46 færre per 1000 (76 færre til 10 færre)	RR 0.75 (0.59 til 0.95)	4929 (23) ²⁻²⁴	LAV	Brug af kaffe medfører muligvis færre dumpede studerende sammenlignet med ingen brug af kaffe
Tissepauser pr dag (Pee breaks pr day) (Endt intervention)	3.5 stk	5 stk	MD 1.5 (1.38 flere til 1.62 flere)	Lavere værdi =bedre	1303 (1) ²	HØJ	Brug af kaffe medfører sandsynligvis flere tissepauser per studerende sammenlignet med ingen brug af kaffe
Livskvalitet (Quality of life) (Endt intervention)			SMD 0.11 (0.03 færre til 0.26 mere)	Højere værdi =bedre	1015 (6) ³⁻⁸	LAV	Brug af kaffe medfører muligvis højere livskvalitet sammenlignet med ingen brug af kaffe
*Den absolutte effekt i kontrolgruppen er baseret på den gennemsnitlige effekt i de inkluderede studier medmindre andet er angivet i kommentarfeltet. Den absolutte effekt i interventionsgruppen er baseret på den relative effekt og den absolutte risiko i kontrolgruppen.							
CI: Konfidensinterval; RR: Relativ risiko, MD: Gennemsnitlig forskel, SMD: Standardiserede gennemsnitlige forskelle							
GRADE evidensniveauer:							
Høj: Vi er meget sikre på, at den sande effekt ligger tæt på den estimerede effekt.							
Moderat: Vi er moderat sikre på den estimerede effekt. Den sande effekt ligger sandsynligvis tæt på denne, men der er en mulighed for, at den er væsentligt anderledes.							
Lav: Vi har begrænset tiltro til den estimerede effekt. Den sande effekt kan være væsentligt anderledes end den estimerede effekt							
Meget lav: Vi har meget ringe tiltro til den estimerede effekt. Den sande effekt vil sandsynligvis være væsentligt anderledes end den estimerede effekt.							

Tabel 7 Forklaring af evidensprofilen

	Forklaring	Eksempel
Population	Målgruppen for anbefalingen.	Studerende
Intervention	Den intervention man vil komme med en anbefaling vedrørende.	Kaffe
Sammenligning	Alternativet til intervention.	Ingen kaffe
Outcome	De udvalgte outcome, der blev dømt kritiske eller vigtige for at kunne give en anbefaling. De kritiske outcome er styrende for anbefalingen.	Antal dumpet, tissepauser pr dag, livskvalitet, alle målt umiddelbart efter endt intervention.
Absolut effekt	<p>Den absolutte effekt i kontrolgruppen er baseret på den gennemsnitlige effekt i de inkluderede studier medmindre andet er angivet i kommentarfeltet.</p> <p>For dikotome outcome er den absolutte effekt i interventionsgruppen er beregnet ved hjælp af den relative effekt og den absolutte risiko i kontrolgruppen.</p> <p>For kontinuerte outcome bruges enten MD eller SMD.</p> <p>MD: Den gennemsnitlige forskel.</p> <p>SMD: Den standardiserede gennemsnitlige forskel. For outcome, der bliver målt med forskellige skalaer i de inkluderede studier, er det nødvendigt at standardisere. Enheden bliver nu standard afvigelser i stedet for de enheder skaler oprindeligt brugte, eksempelvis point på en given skala. Som en tommelfingerregel kan resultater under 0,3 tolkes som en lille effekt, 0,3 til 0,8 som en mellem effekt og over 0,8 som en stor effekt</p>	<p>For det dikotome outcome 'Antal dumpet' er angivet den gennemsnitlige absolutte effekt i kontrolgruppen (ingen kaffe) 184 pr 1000. Denne kolonne angiver således baseline risikoen for at dumpe hvis man ikke intervenserer. Den absolutte effekt hos dem, der drikker kaffe er beregnet ved at tage den relative risiko 0,75 (beregnet i en meta-analyse) og gange den med den absolutte risiko hos kontrolgruppen: $0,75 * 184 \text{ pr } 1000 = 138 \text{ pr } 1000$. 'Forskel med kaffe' kolonnen angiver den absolutte forskel: $184 \text{ pr } 1000 - 138 \text{ pr } 1000 = 46 \text{ færre pr } 1000$.</p> <p>For det kontinuerte outcome 'Tissepauser pr dag' er angivet hvor mange pauser der i gennemsnit er i henholdsvis kaffe (3,5 pauser) og ingen kaffe (5 pauser) gruppen samt den gennemsnitlige forskel (MD) med kaffe: 1,5 flere tissepauser pr dag.</p> <p>For det kontinuerte outcome 'Livskvalitet' er angivet den standardiserede gennemsnitlige forskel (SMD) med kaffe: 0,11.</p>
Relativ effekt	For dikotome outcome opgives den relative effekt. Dette vil oftest være som relativ risiko, men kan også være som odds ratio eller hasard ratio afhængigt af data.	For det dikotome outcome 'Antal dumpet' er angivet den relative risiko for at dumpe: 0,75. Den relative risiko er beregnet i en meta-analyse. Det vil sige at indtagelse af kaffe reducerer risikoen for dumpe med 25%.

	<p>For kontinuerte outcome beregnes ikke nogen relativ effekt, i stedet bruges kolonnen til at opgive retningen af den anvendte skala</p>	<p>For de kontinuerte outcome 'Tissepauser pr dag' og 'Livskvalitet' er retningen af skalaen angivet. For tissepauser er det bedre med en lavere værdi og for livskvalitet er det bedre med en højere værdi.</p>
Antal deltagere (studier)	<p>Det samlede antal deltagere og studier der ligger grund til estimatet for de enkelte outcome. Her indsættes også referencer så læseren kan se hvilke studier der bidrager til hvilke outcome. Bemærk at det ikke altid vil være de samme studier, der bidrager til de forskellige outcome, da ikke alle studier måler på de samme outcome.</p>	<p>For 'Antal dumpet' er der i alt 23 studier med 4929 personer der har bidraget til analysen, for 'Tissepauser pr dag' er der et studie med 1303 personer og for 'Livskvalitet' 6 studier med 1015 personer.</p>
Evidensniveau	<p>Evidensniveauet ifølge GRADE, fortolkningen er som følger:</p> <p>Høj: Vi er meget sikre på, at den sande effekt ligger tæt på den estimerede effekt.</p> <p>Moderat: Vi er moderat sikre på den estimerede effekt. Den sande effekt ligger sandsynligvis tæt på denne, men der er en mulighed for, at den er væsentligt anderledes.</p> <p>Lav: Vi har begrænset tiltro til den estimerede effekt. Den sande effekt kan være væsentligt anderledes end den estimerede effekt</p> <p>Meget lav: Vi har meget ringe tiltro til den estimerede effekt. Den sande effekt vil sandsynligvis være væsentligt anderledes end den estimerede effekt.</p>	<p>For alle outcome er vurderingen 'Lav', det vil sige at tiltroen til estimererne begrænset. Årsagen til den begrænsede tiltro finder man i kommentarfeltet</p>
Kommentarer	<p>Her angives hovedårsagerne til en eventuel nedgradering, og det angives om det givne outcome er kritisk eller vigtigt for anbefalingen.</p>	<p>'Antal dumpet' er kritisk for udformningen af anbefalingen, 'Tissepauser pr dag' og 'Livskvalitet' er vigtige.</p> <p>Man kan se at for 'Antal dumpede' er der risiko for bias (systematisk skævridding) og resultaterne var inkonsistente på tværs af studier. For 'Tissepauser pr dag' var der kun et ublindt studie og for 'Livskvalitet' var der risiko for bias og estimatet var upræcist.</p>

3.7.6 Evidensprofiler (diagnostiske tests studier)

Evidensprofiler laves ved hjælp af software, der findes forskellige alternativer på markedet. Sundhedsstyrelsen benytter **GDT** til evidensprofiler vedrørende diagnostiske test, da MAGICapp pt ikke kan håndtere dette endnu. Der er flere valgmuligheder når man skal præsentere evidensen, nedenfor er vist den tabeltype, der hedder 'Summary of Findings'. Tabellen giver et overblik over antallet af deltagere og studier, sensitivitet og specificitet, prævalensen af den pågældende sygdom (op til 3 forskellige prævalenser kan angives), sandt positive, falsk positive, sandt negative, falsk negative samt kvaliteten af evidensen.

Tabel 8 Evidensprofil – diagnostiske tests

Bør IPDS (selv rapporteret) anvendes til at identificere mulig borderline personlighedsforstyrrelse i primær sektor (praktiserende læge)? (Referencestandard: SCID-II)

Sensitivitet: 0.77 | Specificitet: 0.85

Effekt mål	Antallet af resultater per 1000 patienter testet			Antal deltagere (studier)	Evidensniveau (GRADE)
	Prævalens 50%*	Prævalens 20%**	Prævalens 5%***		
Sandt positive (Patienter korrekt klassificeret med borderline personlighedsforstyrrelse)	385	154	39	195 (1)	⊕○○○ MEGET LAV ¹
Falsk positive (Patienter ukorrekt klassificeret som ikke havende borderline personlighedsforstyrrelse)	115	46	11		
Sandt negative (Patienter korrekt klassificeret uden borderline personlighedsforstyrrelse)	425	680	808	195 (1)	⊕○○○ MEGET LAV ¹
Falsk negative (Patienter ukorrekt klassificeret som havende borderline personlighedsforstyrrelse)	75	120	142		

- Ikke praktiserende læge, risiko for bias og kun et studie

*Typisk set i psykiatrisk ambulant regi

**Typisk set i primær sektor

***Typisk set hos den praktiserende læge

3.7.7 Vurdering af tiltroen til det overordnede evidensniveau

Når tiltroen til de enkelte effektestimater er vurderet på outcome-niveau, skal den samlede tiltro til evidensen vurderes på tværs af outcome. Dette afgøres af de outcome der er defineret som kritiske for udarbejdelsen af anbefalingen, og det overordnede evidensniveau fastsættes som udgangspunkt efter det lavest vurderede, kritiske outcome.

Når anbefalingen skal formuleres, tager arbejdsgruppen udgangspunkt i tiltroen til hvorvidt effektestimaterne er pålideligt underbyggede og kan støtte en given anbefaling.

Videre læsning:

GRADE arbejdsgruppen (www.gradeworkinggroup.org)

Artikelserie i Journal of Clinical Epidemiology om GRADE
(http://www.gradeworkinggroup.org/publications/JCE_series.htm)

<http://www.bmj.com/content/353/bmj.i2016>

3.8 Udarbejdelse af anbefalinger

Anbefalinger kan være for eller imod en given intervention eller diagnostisk test, hvilket afhænger af retningen på effektestimaterne. Styrken af anbefalingen kan være stærk eller svag, hvilket afhænger af tiltroen til de estimerede effekter, patienternes værdier og præferencer, balancen imellem gavnlige og skadelige virkninger, samt andre eventuelle overvejelser, for eksempel praktiske hensyn ved implementering. En anbefaling bør være svag, når man har lav tiltro til de estimerede effekter, eller når balancen mellem gavnlige og skadelige virkninger gør, at patienternes præferencer forventes at variere. En anbefaling bør være stærk, når man har stor tiltro til de estimerede effekter, og balancen mellem gavnlige og skadelige virkninger samtidig er entydig.

3.8.1 Fra evidens til anbefaling ved interventionsstudier

Det er ikke alene tiltroen til de estimerede effekter, der ligger til grund for en anbefaling. Nedenstående faktorer skal også diskuteres systematisk i arbejdsgruppen når en anbefaling udarbejdes. Arbejdsgruppens væsentligste overvejelser dokumenteres og publiceres sammen med anbefalingen. Overvejelserne beskrives i nedenstående skema:

Kvaliteten af evidensen

Er tiltroen til de estimerede effekter meget lav, lav, moderat eller høj?

Stærke anbefalinger for eller imod kræver som udgangspunkt, at man har moderat eller høj tiltro til effekterne for de kritiske outcome.

Værdier og præferencer

Hvis arbejdsgruppen vurderer, at patienterne vil vælge forskelligt, når de informeres om de gavnlige og skadelige effekter af behandlingen, dvs. at valget er præferencefølsomt, vil anbefalingen som udgangspunkt blive svag, for eller imod.

Det samme gør sig gældende, hvis man er usikker på, hvad patienternes værdier og præferencer er.

Hvis langt de fleste patienter vil træffe den samme beslutning om behandlingen, vil anbefalingen som udgangspunkt være stærk, for eller imod.

Balancen mellem effekt og skadevirkning

Hvis de gavnlige virkninger af behandlingen utvetydigt overstiger de skadelige, kan man give en stærk anbefaling, for eller imod.

Hvis balancen mellem gavnlige og skadelige virkninger er uklar, gives en svag anbefaling, for eller imod.

Andre overvejelser

Her noteres andre overvejelser arbejdsgruppen har gjort sig. Det kan fx være behov for efteruddannelse af personale, eller praktiske udfordringer i forbindelse med behandlingen.

3.8.2 Fra evidens til anbefaling ved diagnostiske studier

Det er ikke alene den viden man har indhentet fra litteraturen, der ligger til grund for en anbefaling. Nedenstående faktorer skal også diskuteres i arbejdsgruppen når en anbefaling udarbejdes. Sensitivitet og specificitet er surrogat outcome og giver dermed ofte kun grundlag for svage anbefalinger. De kan dog i nogle situationer danne grundlag for stærke anbefalinger, men det kræver at nogle forudsætninger er opfyldt. For det første skal der være tale om en veldefineret sygdomskategori, hvor definitionen ikke har ændret sig væsentligt over tid. Dette er for eksempel tilfældet for frakturer men ikke for mange psykiatriske diagnoser. Desuden skal arbejdsgruppen have moderat til høj tiltro til estimeret af potentialet for overdiagnostik ved referencestandard. Potentialet skal være enten lille eller med konsekvenser, der klart opvejes af de gavnlige virkninger. Hvis der er solid evidens for, at referencestandard forbedrer patientrelevante outcome, kan anbefalingen også blive stærk alene baseret på sensitivitet og specificitet, hvis potentialet for overdiagnostik og konsekvenserne heraf er veldokumenterede.

Kvaliteten af evidensen

Er tiltroen til de estimerede effekter meget lav, lav, moderat eller høj?

Stærke anbefalinger kræver som udgangspunkt, at man har moderat eller høj tiltro til effekten på de kritiske outcome.

Værdier og præferencer

Hvis præferencer vurderes at variere, det vil sige at patienterne vil vælge forskelligt, når de informeres om de estimerede gavnlige og skadelige effekter af testen, vil anbefalingen som udgangspunkt være svag. Det samme kan gøre sig gældende, hvis man ikke ved, hvad patientens værdier og præferencer er.

Hvis langt de fleste patienter vil træffe den samme beslutning om testen, vil anbefalingen som udgangspunkt være stærk, hvis de estimerede effekter samtidig er velunderbyggede.

Balancen mellem effekt og skadevirkning

Arbejdsgruppen skal både vurdere om testen kan forbedre prognosen, men også konsekvenserne af falsk positive og negative testresultater, samt risikoen for og konsekvenserne af hhv. under- og overdiagnosticering.

Andre overvejelser

Her anføres fx eventuelle behov for efteruddannelse af personale eller logistiske udfordringer i forbindelse med testen.

3.8.2.1 Rationale

Her foretages en afvejning af de forskellige argumenter, der ledte til anbefalingen. Hensigten er, at gøre beslutningsprocessen gennemskelig. Hvilke outcome valgte arbejdsgruppen at lægge vægt på, var patientpræferencer afgørende, etc. Uddyb også, hvis der blev givet en svag anbefaling, selvom der var høj tiltro til evidensen.

Dette kan fx skyldes, at effekterne var små og det blev vurderet at der var præferencefølsomhed. Det samme gælder, hvis der blev givet en stærk anbefaling i tilfælde med lav tiltro til evidensen, men hvor alle outcome pegede i den samme retning og der ikke var præferencefølsomhed.

3.8.3 Manglende evidens

De fokuserede spørgsmål der stilles i en NKR er spørgsmål, som arbejdsgruppens medlemmer har fundet af afgørende vigtighed. Derfor skal de fokuserede spørgsmål besvares ved anbefalinger. Hvis der ikke foreligger evidens har Sundhedsstyrelsens derfor valgt at der skal udformes en ”god praksis” anbefaling, som kan benyttes i dette tilfælde. Disse anbefalinger er således ikke en del af GRADE metoden, og benyttes *udelukkende*, hvis der ikke foreligger evidens.

3.8.4 Evidensbaserede anbefalinger

Følgende fire anbefalinger kan anvendes, hvis der er evidens: Stærk for, stærk imod, svag for eller svag imod. Nedenfor gennemgås formulering og implikationer af de forskellige typer af anbefalinger. Er der fundet evidens for at behandlinger er ligeværdige kan der i stedet gives en anbefaling for begge interventioner. Styrken og retningen vil afhænge af de tidligere nævnte faktorer. Er der ikke fundet evidens vælges i stedet en god praksis anbefaling.

Stærk anbefaling for ↑↑

Ordlyd: *Giv/brug/anvend...*

Arbejdsgruppen giver en stærk anbefaling for, når der er pålidelig evidens, der viser, at de samlede fordele ved interventionen er klart større end ulemperne.

Følgende vil trække i retning af en stærk anbefaling for:

- Høj eller moderat tiltro til de estimerede effekter.
- Stor gavnlige effekt og ingen eller få skadevirkninger.
- Patienternes værdier og præferencer er velkendte og ensartet til fordel for interventionen.

Implikationer:

- De fleste patienter vurderes at ønske interventionen.
- Langt de fleste klinikere vil tilbyde interventionen.

Svag/betinget anbefaling for ↑

Ordlyd: *Overvej at...*

Arbejdsgruppen giver en svag/betinget anbefaling for intervention en, når det vurderes, at fordelene ved interventionen er marginalt større end ulemperne, eller den tilgængelige evidens ikke kan udelukke en væsentlig fordel ved en eksisterende praksis, samtidig med at skadevirkningerne er få eller fraværende.

Følgende vil trække i retning af en svag/betinget anbefaling for:

- Lav eller meget lav tiltro til de estimerede effekter.
- Balancen mellem gavnlige og skadelige virkninger ikke er entydig.
- Patienternes præferencer og værdier vurderes at variere væsentligt, eller de er ukendte.

Implikationer:

- De fleste patienter vurderes at ønske interventionen, men nogen vil afstå.
- Klinikerne vil skulle bistå patienten med at træffe en beslutning, der passer til patientens værdier og præferencer.

Svag/betinget anbefaling imod ↓

Ordlyd: *Anvend kun ... efter nøje overvejelse, da den gavnlige effekt er usikker og/eller lille, og der er dokumenterede skadevirkninger såsom...*

Eller: *Anvend ikke ... rutinemæssigt, da den gavnlige effekt er usikker og/eller lille, og der er dokumenterede skadevirkninger såsom...*

Arbejdsgruppen giver en svag/betinget anbefaling imod interventionen, når ulemperne ved interventionen vurderes at være større end fordelene, men hvor man ikke har høj tiltro til de estimerede effekter. Den svage/betingede anbefaling imod, anvendes også hvor der er stærk evidens for både gavnlige og skadelige virkninger, men hvor balancen mellem dem er vanskelig at afgøre.

Følgende vil trække i retning af en svag anbefaling imod:

- Lav eller meget lav tiltro til de estimerede effekter. Balancen mellem gavnlige og skadelige virkninger ikke er entydig.
- Skadevirkningerne vurderes at være marginalt større end den gavnlige effekt.
- Patienternes præferencer og værdier vurderes at variere væsentligt, eller de er ukendte.

Implikationer:

- De fleste patienter vurderes at ville afstå fra interventionen, men nogen vil ønske den.
- Klinikerne vil skulle bistå patienten med at træffe en beslutning, der passer til patientens værdier og præferencer.

Stærk anbefaling imod ↓↓

Ordlyd: *Giv ikke/brug ikke/anvend ikke/undlad at...*

Arbejdsgruppen giver en stærk anbefaling imod, når der er der er høj tiltro til, der viser, at de samlede ulemper er klart større end fordelene. Det samme gælder, hvis der er stor tiltro til, at en intervention er nyttesløs.

Følgende vil trække i retning af en stærk anbefaling imod:

- Høj eller moderat tiltro til de estimerede effekter.
- Der er stor tiltro til, at interventionen ikke gavner, eller at den gavnlige effekt er lille.
- Der er stor tiltro til, at interventionen har betydelige skadevirkninger.
- Patienternes værdier og præferencer er velkendte og ensartede imod interventionen.

Implikationer:

- De fleste patienter vurderes ikke at ville ønske interventionen.
- Klinikeren vil meget sjældent tilbyde interventionen

For yderligere beskrivelse af de forskellige evidensbaserede anbefalinger se:

<http://www.gradeworkinggroup.org>

3.8.5 God praksis anbefalinger

God praksis bygger på faglig konsensus blandt medlemmerne af arbejdsgruppen, der har udarbejdet den kliniske retningslinje. Anbefalingen kan være enten for eller imod interventionen. God praksis anbefalinger anvendes udelukkende, når der ikke foreligger relevant evidens. Derfor er denne type anbefaling svagere end de evidensbaserede anbefalinger, uanset om de er stærke eller svage. God praksis anbefaling må kun benyttes, hvis der som minimum er søgt efter RCT studier og ikke fundet noget litteratur. Har arbejdsgruppen fundet blot ét studie skal der, uanset dette studies kvalitet, udarbejdes enten en svag eller en stærk anbefaling.

God praksis ✓

Ordlyd for:

Det er god praksis at overveje at...

Ordlyd imod:

Det er ikke god praksis at...

Det er ikke god praksis rutinemæssigt at...

Det er god praksis at undlade at...

Det er god praksis at undlade rutinemæssigt at...

3.8.6 Videre forskning

Under udarbejdelsen af den nationale kliniske retningslinje, støder arbejdsgrupper nogle gange på områder, hvor evidensen er sparsom eller helt mangler. Sådanne områder kan med fordel udgøre fremtidige forskningsfelter. Det anbefales, at arbejdsgrupperne udpeger disse områder og kommer med konkrete forslag til hvilken forskning, de mener, ville være et væsentligt bidrag til feltet.

3.9 Implementering og monitorering

Nationale kliniske retningslinjer indeholder anbefalinger for et antal udvalgte og velafgrænsede kliniske problemstillinger og udgør dermed ikke én samlet behandlingsvejledning for et sygdomsområde og erstatter ikke det faglige skøn i den konkrete kliniske situation, der er afgørende for beslutning om passende og korrekt sundhedsfaglig ydelse

Det betyder, at retningslinjernes anbefalinger ikke kan stå alene som beslutningsstøtte til klinisk praksis, men skal sammenholdes med de landsdækkende, regionale og kommunale retningslinjer og instrukser på området.

Den enkelte arbejdsgruppe skal foreslå aktiviteter eller tiltag, som kan optimere udbredelsen og anvendelsen af netop denne nationale kliniske retningslinje, eksempelvis udbredelse af kendskab til retningslinjens anbefalinger via faglige fora såsom årsmøder, nyhedsbreve, konferencer mv.

Sundhedsstyrelsen har udarbejdet en [implementeringshåndbog](#). Målgruppen for håndbogen er de personer, der lokalt har ansvar for implementeringen eller opgaver i forbindelse hermed, såsom projektmedarbejderen. Konkret indeholder implementeringshåndbogen et bredt udvalg af metoder, som kan inspirere til lokale initiativer og fremme implementeringen, og værktøjer, der kan anvendes direkte i implementeringsarbejdet. Det vil blandt andet sige forslag til, hvordan man kan fastsætte mål, bruge feedback, lave tidsestimater og inddrage patienter og pårørende. Håndbogen er bygget op i to hovedsektioner:

- En sektion med gode råd til gennemførelse af implementeringsaktiviteterne fra start til slut. Denne del trækker på generel implementeringsteori og forskning i implementering af forbedringer i sundhedssektoren.
- En sektion med udvalgte erfaringsbaserede hjælpeværktøjer, som kan understøtte gennemførelsen af specifikke delopgaver. Denne del indeholder detaljerede skemaer og vejledninger.

Det kan være vanskeligt at monitorere implementeringen af nationale kliniske retningslinjer isoleret set. Dels er det vanskeligt at isolere effekten fra andre parametre, og dels har vi på nuværende tidspunkt ikke datamodellerne til det. De nationale kliniske retningslinjer bliver ikke monitoreret systematisk på nationalt niveau.

I udviklingen af enkelte nationale kliniske retningslinjer er erfaringen dog, at arbejdsgruppemedlemmer ofte har kendskab til en database forankret i de faglige miljøer, som kan indeholde relevante indikatorer i monitoreringen af den pågældende NKR. Som eksempelvis her kan nævnes Dansk Fraktur Database, som er under udvikling og endnu ikke har komplet datadækning – men nyder stor opbakning hos klinikerne. Herudover kan de kliniske databaser under regionernes kliniske kvalitetsudviklingsprogram – nogle anvendes i højere grad end andre - tjene som inspiration til monitorering.

Arbejdsgruppen skal derfor forholde sig til, om der er eksisterende databaser og indikatorer, der kan anvendes i monitorering af den nationale kliniske retningslinje, uden at det indebærer krav om yderligere monitorering på området. Findes der ikke anerkendte indikatorer for monitorering af de nævnte områder, er det arbejdsgrup-

pernes opgave at give forslag til få, centrale målepunkter i forhold til ønskede ændringer i praksis.

4 Fase III – publicering

4.1 Indsendelse af færdigt udkast

Sundhedsstyrelsen anbefaler, at der fra start vælges én fagkonsulent, som samler og sammenfatter arbejdet til en endelig version af den nationale kliniske retningslinje. Ved afslutning af arbejdet godkender arbejdsgruppen udkastet, hvorefter det sendes til Sundhedsstyrelsen.

Efter endelig godkendelse hos arbejdsgruppen sendes retningslinjen til kvalitetsvurdering hos Sundhedsstyrelsen.

Sundhedsstyrelsen vil basere den afsluttende kvalitetsvurdering af den kliniske retningslinje før publicering på AGREEII instrumentet (Appraisal of Guidelines for Research and Evaluation). Herudover vil retningslinjen blive sendt til det rådgivende udvalg for NKR til kommentering.

Sundhedsstyrelsen returnerer herefter retningslinjen til ansøger. I tilbagemeldingen fra Sundhedsstyrelsen vil AGREE-vurderingen fremgå, samt hvilke eventuelle ændringer, der skal foretages inden publicering.

Ansøger sender herefter retningslinjen i peer review og foretager eventuelle rettelser på baggrund af kommentarer fra Sundhedsstyrelsen og peer review. Ansøger publicerer herefter retningslinjen på egen hjemmeside. Sundhedsstyrelsen vil linke til denne fra Sundhedsstyrelsens hjemmeside.

4.2 Format

Nationale kliniske retningslinjer udarbejdes i MAGICapp, og følger dennes skabelon. Der skal endvidere benyttes visse standardformuleringer som Sundhedsstyrelsen vil gøre tilgængelig for ansøgerne.

MAGICapp er et redaktionelt online it-værktøj. Alle elementerne i retningslinjen bliver lagt ind efter en fast struktur og bliver herefter publiceret online. Webdesignet er responsivt, det vil sige, at det automatisk tilpasses de skærmstørrelser, som brugerne benytter. Når brugeren skifter fra laptop til tablet eller mobiltelefon, registreres brugerens skærmstørrelse og opløsning, så designet optimeres til den plads, der er til rådighed på skærmen. Indholdet i MAGICapp har også mulighed for på sigt at kunne integreres i f.eks. elektroniske patientjournaler. Det er muligt løbende at generere en version af retningslinjen i PDF format i MAGICapp, ligesom det er muligt at generere kortere versioner af webdokumentet. Disse versioner kan man ikke rette/slette i, men man vil indenfor den nærmeste fremtid kunne kommentere i dem.

Sproget og det faglige niveau i en national klinisk retningslinje skal henvende sig til sundhedsprofessionelle generelt.

Til rapporten udarbejdes en række bilag som vil være tilgængelige på Sundhedsstyrelsens hjemmeside.