



SUNDHEDSSTYRELSEN

# MODEL

# Metodehåndbogen

MODEL FOR UDARBEJDELSE AF  
NATIONALE KLINISKE RETNINGSLINJER



2018

## METODEHÅNDBOGEN

### Model for udarbejdelse af nationale kliniske retningslinjer

© Sundhedsstyrelsen, 2018.

Publikationen kan frit refereres med tydelig kildeangivelse.

#### Sundhedsstyrelsen

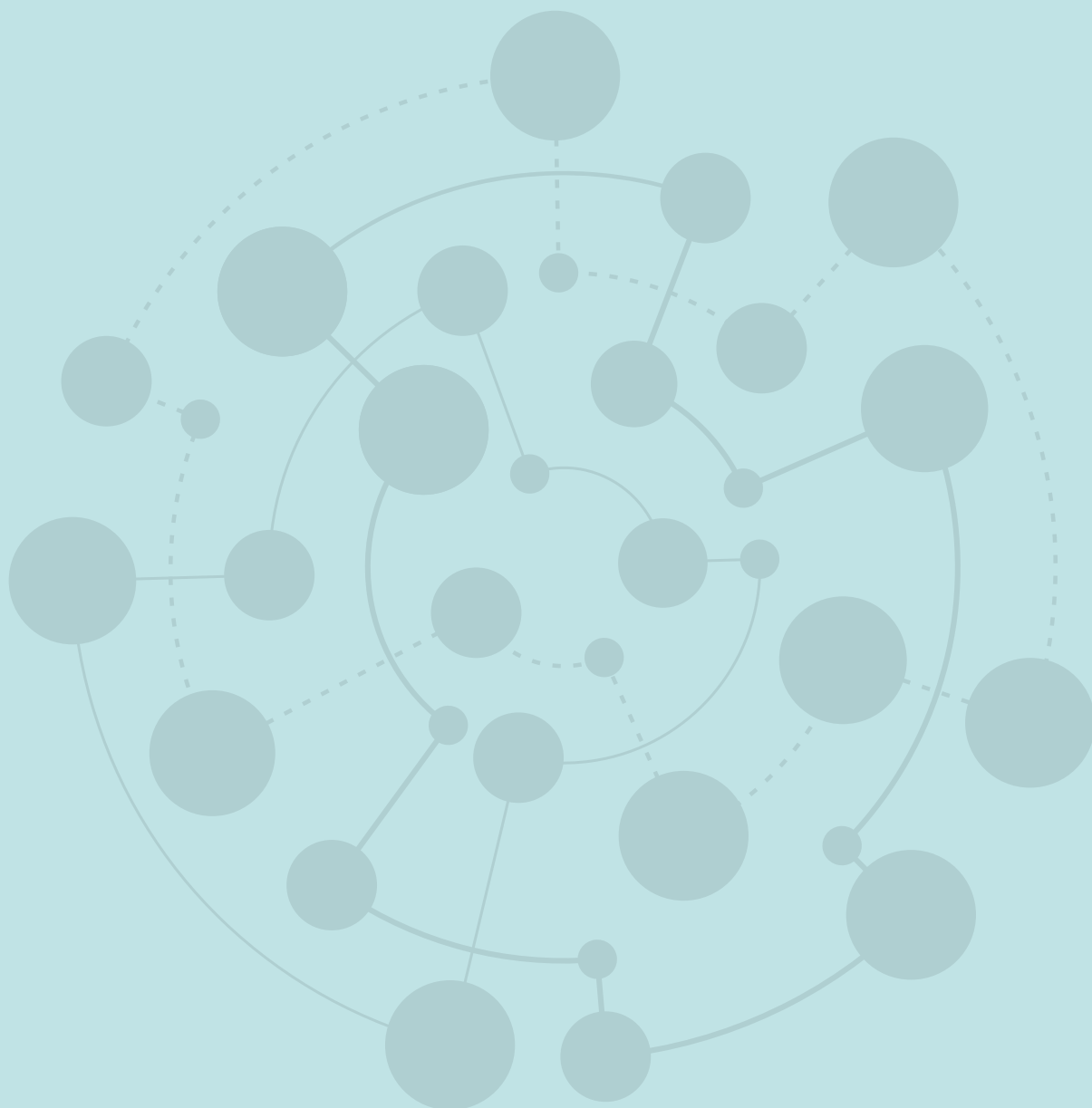
Islands Brygge 67  
2300 København S

URL: <http://www.sst.dk>

Sprog: Dansk

Versionsnummer 3

Udgivet af Sundhedsstyrelsen, december 2017.



# Indhold

<b>1.</b>	<b>INTRODUKTION</b>	<b>5</b>
1.1	Om håndbogen	7
1.2	Definition – hvad en er en NKR?	7
1.3	Organisatorisk ramme	8
1.4	Tidsrammen	8
1.5	Afgrænsning til andre faglige anbefalinger	8
	<b>FASE I</b>	
<b>2.</b>	<b>FORBEREDELSE</b>	
2.1	Valg af emner og afgrænsning	11
2.2	Kontrakt med fag- og metodekonsulent samt søgespecialist	11
2.3	Nedsættelse af arbejds- og referencegruppe	12
2.3.1	Arbejdsgruppen	12
2.3.2	Referencegruppen	13
2.3.3	Formand	13
2.4	Projektsekretariatet	13
2.5	Habilitet	13
	<b>FASE II</b>	
<b>3.</b>	<b>UDARBEJDELSE</b>	<b>15</b>
3.1	<b>Fokuserede spørgsmål</b>	<b>18</b>
3.1.1	Spørgsmål om effekten af en intervention inklusive screening/tidlig opsporing	18
3.1.2	Spørgsmål om diagnostiske tests til udredning	21
3.1.3	Diagnostisk præcision	22
3.1.4	Patientrelevante outcome	22
3.1.5	Hvilken rolle skal den nye test spille?	22
3.1.6	Overdiagnostik	23
3.2	<b>Litteratursøgning</b>	<b>25</b>
3.2.1	Referencehåndtering	27
3.2.2	Afgrænsning vedrørende sprog	27
3.2.3	Søgebeskrivelse	27
3.3	<b>Udvælgelse af evidens</b>	<b>27</b>
3.3.1	Typer af studiedesign	28
3.3.2	Sortering af søgeresultater	29
3.4	<b>Vurdering af evidens</b>	<b>29</b>
3.4.1	Vurdering af retningslinjer	29
3.4.2	Vurdering af systematiske oversigtsartikler	32
3.4.3	Vurdering af primære studier	35
3.5	<b>Sammenfatning af evidens</b>	<b>36</b>
3.5.1	Dataekstraktion	36
3.5.1.1	Retningslinjer	36
3.5.1.2	Systematiske oversigtsartikler	36
3.5.1.3	Primære studier	37
3.5.2	Sammenfatning af resultater fra interventionsstudier	37
3.5.3	Sammenfatning af resultater fra diagnostiske primærstudier	37

<b>3.6</b>	<b>Gradering af evidens (GRADE)</b>	<b>38</b>
3.6.1	Vurdering af tiltroen til det enkelte outcome	38
3.6.2	Nedgradering af evidensen for effekt af interventioner	40
3.6.3	Opgradering	41
3.6.4	Nedgradering af evidens for præcision af diagnostiske tests	42
3.6.5	Evidensprofiler (interventionsstudier)	44
3.6.6	Evidensprofiler (diagnostiske tests studier)	47
3.6.7	Vurdering af tiltroen til det overordnede evidensniveau	48
<b>3.7</b>	<b>Udarbejdelse af anbefalinger</b>	<b>48</b>
3.7.1	Fra evidens til anbefaling ved interventionsstudier	49
3.7.2	Fra evidens til anbefaling ved diagnostiske studier	49
3.7.2.1	Rationale	50
3.7.3	Sundhedsøkonomiske analyser	50
3.7.4	Manglende evidens	50
3.7.5	Evidensbaserede anbefalinger	51
3.7.6	God praksis anbefalinger	53
3.7.7	Videre forskning	53
<b>3.8</b>	<b>Implementering og monitorering</b>	<b>53</b>
	<b>FASE III</b>	
<b>4.</b>	<b>PUBLICERING</b>	<b>55</b>
4.1	Indledning	56
4.2	Format	56
4.3	Rapportskabelon	56
4.4	Høring, peer reviews og godkendelse	57

# INTRODUKTION

1



# NKR

NATIONALE KLINISKE RETNINGSLINJER

Med Finanslov 2017 er der afsat 38 mio. kr. i perioden 2017-2020 til videreførelse af arbejdet med de 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. Administration og sekretariat for ordningen varetages af Sundhedsstyrelsen. NKR er således betegnelsen for de nationale kliniske retningslinjer, der publiceres af Sundhedsstyrelsen eller af puljeprojekter i dette regi.



## 1.1 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. Denne håndbog henvender sig til de arbejds- og referencegrupper, konsulenter samt projektmedarbejdere i Sekretariatet for Nationale Kliniske Retningslinjer, som skal udarbejde NKR, men kan også anvendes som inspiration til fagpersoner generelt i andre sundhedsorganisationer, som skal udarbejde kliniske retningslinjer. Herudover har Sundhedsstyrelsen udarbejdet en separat metodehåndbog, som henvender sig til de faglige selskaber og miljøer som søger midler fra NKR-puljen til at udarbejde puljefinansierede nationale kliniske retningslinjer.

## 1.2 Definition – hvad en er en NKR?

Sundhedsstyrelsens 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 den faglige arbejdsgruppe 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. 10 fokuserede spørgsmål per retningslinje. En national klinisk retningslinje er altså ikke en A til Z lærebog, men et evidensbaseret beslutningsstøtteværktøj målrettet klinikere.**

### 1.3 Organisatorisk ramme

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

- 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.
- 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 *sekretariat* i Sundhedsstyrelsen, der udvikler konceptet for NKR og betjener de forskellige arbejdsgrupper samt det Rådgivende Udvalg.
- En sundhedsfaglig *arbejdsgruppe* for hver klinisk retningslinje bestående af ca. 10 medlemmer udpeget af relevante faglige selskaber
- En *referencegruppe* bag hver klinisk retningslinje udpeget af regioner, kommuner, patientforeninger og andre relevante interessenter på området.

### 1.4 Tidsrammen

Tidsrammen for udarbejdelse af en national klinisk retningslinje er på 10 måneder fra første arbejdsgruppemøde til offentliggørelse af retningslinjen. Dertil kommer ca. to måneders forberedelse af arbejdet i arbejdsgruppen (afgrænsning af emnet, udarbejdelse af kommissorium, sammensætning af projektgruppe, udpegning af medlemmer til arbejdsgruppe og referencegruppe, indledende søgninger og formulering af et første bruttoudkast til fokuserede spørgsmål). Eksempel på generisk tidsplan ses på Sundhedsstyrelsens hjemmeside.

### 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 – f.eks. 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 nuværende 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



opfylde behovet for retningslinjer, så der vil fortsat være behov for lokale og regionale retningslinjer, som bør baseres på NKR, hvor disse foreligger.

Nationale kliniske retningslinjer er faglig rådgivning, hvilket indebærer, at Sundhedsstyrelsen anbefaler relevante fagpersoner at følge retningslinjerne. De nationale kliniske retningslinjer er dog 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 beslutning 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.



# Fase I FORBEREDELSE

# 2



## 2.1 Valg af emner og afgrænsning

Sundhedsstyrelsen er overordnet ansvarlig for valg af emner på baggrund af en indstilling fra det Rådgivende Udvalg for NKR. Emneforslagene bør være afgrænsede til f.eks. 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, rehabilitering. Emneforslagene vil bl.a. blive sagsbehandlet af sekretariatet for nationale kliniske retningslinjer efter de vedtagne kriterier og med udgangspunkt i eksisterende data og evt. supplerende datatræk for aktivitet.

### **Udvælgelseskriterierne:**

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

Emneforslag, der lever op til mindst to af ovenstående kriterier, gennemgår en nærmere vurdering, hvor der lægges 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. Der er i vurderingen desuden lagt vægt på, at emneforslagene har et indhold, der har national relevans og med fordel kan løses på nationalt niveau (fx begrundet i tværfagligt eller tværsektorielt indhold).

Sekretariatet udarbejder et kommissorium for de enkelte nationale kliniske retningslinjer, som ligeledes forelægges det Rådgivende Udvalg for NKR. Kommissoriet skal indeholde den nationale kliniske retningslinjes baggrund, formål, problemstilling, omfang, patientgruppe og målgruppe. Kommissoriet skal desuden angive arbejdsgruppens og referencegruppens opgaver og sammensætning. Endelig skal kommissoriet angive en tidsplan for processen.

## 2.2 Kontrakt med fag- og metodekonsulent samt søgespecialist

Når det Rådgivende Udvalg har tilsluttet sig emneforslagene indgår Sundhedsstyrelsen kontrakt med en fagkonsulent, efter samråd med relevante faglige miljøer. Fagkonsulenten udlånes fra sin aktuelle stilling eller varetager arbejdet sideløbende med sin aktuelle stilling i en sammenlagt periode svarende til minimum 3 måneder i de 9-10 måneder udarbejdelsen af retningslinjen tager. Fagkonsulenten fungerer som hovedskribent på den nationale kliniske retningslinje og indgår som en del af sekretariatet.

Der tilknyttes endvidere en metodekonsulent til sekretariatet for den enkelte NKR med særlig viden om evidensvurdering samt en søgespecialist til litteratursøgning.

## 2.3 Nedsættelse af arbejds- og referencegruppe

### 2.3.1 Arbejdsgruppen

Sundhedsstyrelsen nedsætter gennem henvendelse til de relevante videnskabelige selskaber og faglige organisationer en arbejdsgruppe bestående af ca. 10 medlemmer afhængigt af emnet. Arbejdsgruppens medlemmer skal repræsentere de relevante specialer og fagligheder, som emnet for retningslinjen berører.

Der afholdes normalt 6 møder i arbejdsgruppen. Arbejdsgruppens medlemmer må forvente en betydelig arbejdsindsats med gennemgang og vurdering af eksisterende litteratur på området mellem møderne samt kort frist til at kommentere på tekstudkast til retningslinjen, særligt i slutfasen. Arbejdsgruppemedlemmerne forventes endvidere at deltage i en undervisningsdag i NKR metode og evidensvurdering. Arbejdsgruppemedlemmerne må forvente at skulle yde en særlig stor arbejdsindsats mellem 2. og 5. møde. For en mere detaljeret tidsplan henvises til Sundhedsstyrelsens hjemmeside.

#### ▶ Arbejdsgruppens opgaver:

- 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, forventes arbejdsgruppemedlemmerne at være med i den endelige sortering af litteraturen, for at sikre at al relevant litteratur inkluderes (se afsnit 3.3.2 for mere information om sortering).
- AGREE II vurdering af relevante eksisterende retningslinjer (se afsnit 3.4.1 for mere information om AGREE II).
- AMSTAR vurdering af relevante systematiske oversigtsartikler (se afsnit 3.4.2 for mere information om AMSTAR).
- Kvalitetsvurdering af primær litteratur såfremt dette ikke vurderes at være en for stor arbejdsbyrde (se afsnit 3.4.3 for mere information om evidensvurdering af primær litteratur).
- Udformning af anbefalinger (se afsnit 3.7 for mere information om udarbejdelse af anbefalinger).
- Kritisk gennemlæsning af udkast og den endelige version af den nationale kliniske retningslinje.

## 2.2.2 Referencegruppen

Sundhedsstyrelsen nedsætter endvidere gennem henvendelse til relevante myndigheder og organisationer en referencegruppe bestående af en bredere kreds af interessenter såsom regionerne, Danske Regioner, KL, patientforeninger mv. som inviteres til at kommentere på arbejdet med retningslinjerne hhv. relativt tidligt og sent i arbejdsprocessen. Der afholdes 2 møder i referencegruppen.

### ▶ Referencegruppens opgaver:

- Kommentering af afgrænsning og fagligt indhold i retningslinjerne
- Bidrage med relevant faglig og organisatorisk viden
- Kommentering på udkast til retningslinjer inden de sendes i høring

Referencegruppen skal medvirke til at højne kvaliteten og relevansen af den nationale kliniske retningslinje samt medvirke til at fremme implementeringen

## 2.3.3 Formand

Sekretariatet for NKR udpeger en formand, som enten kan være en person, der er ansat i Sundhedsstyrelsen eller en ekstern fagperson, der i sit formandskab repræsenterer styrelsen.

## 2.4 Projektsekretariatet

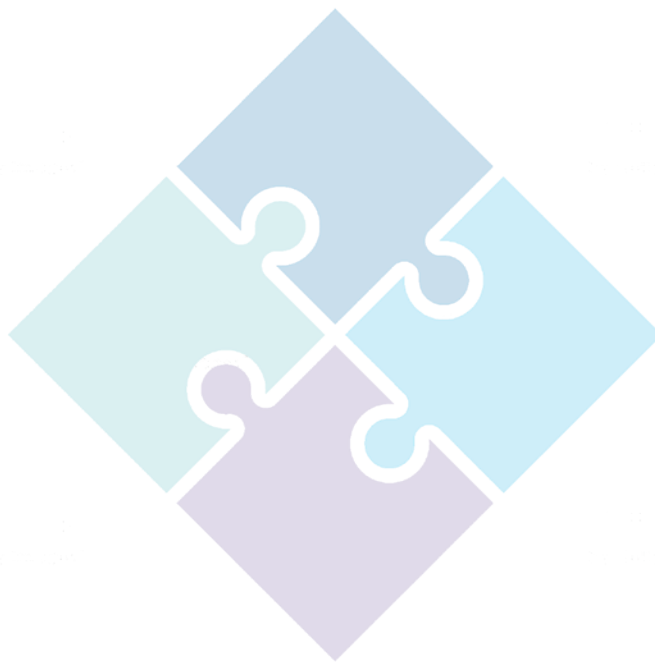
Sundhedsstyrelsen stiller sekretariatsbetjening til rådighed for arbejds- og referencegrupperne. Dette sekretariat består af en formand, en projektleder, en skrivende fagkonsulent, en metodekonsulent og en søgespecialist samt sekretærbistand.

## 2.5 Habilitet

For at sikre gennemsigtighed og håndtere eventuelle habilitetsproblemer anvender Sundhedsstyrelsen habilitetserklæringer, hvor der oplyses om forhold, der kan have betydning for vurderingen af, om arbejdsgruppemedlemmer, konsulenter mv. er inhabil. Udfyldelsen af habilitetserklæringen fritager ikke den enkelte fra i konkrete tilfælde at gøre opmærksom på forhold, der kan medføre inhabilitet, uanset om disse forhold fremgår af habilitetserklæringen eller ej. Den enkelte har også pligt til med det samme at gøre opmærksom på nye væsentlige forhold af betydning for habiliteten samt pligt til at opdatere habilitetserklæringen, således at de nye væsentlige oplysninger fremgår af denne. Udfyldelsen af habilitetserklæringen erstatter ikke lægers og tandlægers pligt til at søge Sundhedsstyrelsen om tilladelse til at være tilknyttet en lægemiddelvirksomhed jfr. Apotekerlovens § 3, stk. 2, ligesom en sådan godkendelse ikke kan erstatte en habilitetserklæring.

Habilitetserklæring for arbejdsgruppen udfyldes på [www.sst.dk](http://www.sst.dk) forud for arbejdets opstart. Dette er et ufravigeligt krav. Habilitetserklæringerne vurderes i Sundhedsstyrelsen vil være offentligt tilgængelige på Sundhedsstyrelsens hjemmeside. Medlemmer af referencegrupperne skal ikke afgive habilitetserklæringer.

Medlemsskab af advisory boards m.v. i en virksomhed, der har aktiviteter på et af Sundhedsstyrelsens fagområder, anses generelt som uforenligt med virke for Sundhedsstyrelsen på det pågældende sagsområde.



# Fase II UDARBEJDELSE



## Nationale kliniske retningslinjer

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

## Sundhedsstyrelsen

- anvender 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 3.6 og 3.7 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, giver 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. Formænd, fagkonsulenter og arbejdsgrupper vil modtage undervisning i GRADE og evidensvurdering. Der afholdes indledningsvist et obligatorisk endags kursus og der vil løbende være undervisning i forbindelse med arbejdsgruppemøderne, således at deltagerne er klædt på til at anvende de forskellige metoder. Igennem hele processen står metodekonsulenten desuden til rådighed til at svare på spørgsmål.



**Figur 1. Trinene i udarbejdelse af en NKR**



## 3.1 Fokuserede spørgsmål

Emnet for den pågældende NKR er indledningsvist afgrænset i kommissoriet. Yderligere afgræsning foretages af arbejdsgruppen, som beslutter hvilke fokuserede spørgsmål retningslinjen skal indeholde. Udarbejdelsen af fokuserede spørgsmål sker med metode-mæssig bistand fra metodekonsulenten. Inden første arbejdsgruppemøde udarbejdes et forslag til fokuserede spørgsmål. Det sker på baggrund af kommissoriet og i et samarbejde mellem sekretariatet, fagkonsulenten, metodekonsulenten og evt. Sundhedsstyrelsens sagkyndige rådgiver(e) inden for området. Udkastet til fokuserede spørgsmål drøftes på 1. møde i arbejdsgruppen. Herunder er det væsentligt at arbejdsgruppen har mulighed for at foreslå andre temaer og konkrete spørgsmål så det sikres at de fokuserede spørgsmål kommer til at afdække de mest relevante kliniske problemstillinger. Mellem 1. og 2. møde i arbejdsgruppen kvalificeres udkastet, og den endelige udvælgelse af de fokuserede spørgsmål foretages på 2. møde i arbejdsgruppen efter drøftelse med referencegruppen.

De fokuserede spørgsmål er den røde tråd igennem hele retningslinjen. Spørgsmålene skal afspejle centrale kliniske problemstillinger, som man ønsker belyst, og de skal formuleres, inden litteraturen gennemgås. Der formuleres op til 10 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)

- P** . . . . . **P**opulation (patientgruppen): Definition af sygdommen/tilstanden og patientmålgruppen med relevante demografiske faktorer (f.eks. alder).
- I** . . . . . **I**ntervention (interventionen): Definition af interventionen/erne. Hvilke variationer (eksempelvis dosis, administrationsform o.l.) findes der af interventionen? Skal alle variationer inkluderes?
- C** . . . . . **C**omparison (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.
- O** . . . . . **O**utcome (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.

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

Udover at gøre sig klart hvilke outcome, der er relevante for retningslinjen, bør arbejdsgruppen også gøre sig klart, hvordan disse outcome kan måles. Hvilke måleredskaber vil f.eks. sandsynligvis blive anvendt? Hvis der er flere måleredskaber, der kan bruges til at måle den samme effekt, f.eks. smerte, hvilket måleredskab vil så blive foretrukket såfremt outcome er afrapporteret 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.

Outcome skal specificeres og rangordnes efter, hvorvidt de er kritiske for beslutningstagningen eller 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 for omfattende og potentielt vildledende, da læseren ville blive forelagt data, der er af ringe eller ingen betydning. Mindre vigtige outcome inkluderer brugen af indirekte eller surrogat outcome, såsom laboratorieresultater eller radiologiske resultater (f.eks. 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. Interventioner kan have en god virkning på et surrogat outcome, men ingen eller endog skadelige indvirkning på den klinisk relevante outcome som surrogat outcomet menes at være et udtryk for. Det kan dog blive nødvendigt at anvende surrogat outcome i tilfælde, hvor der ikke er lavet studier med patient relevante outcome. Dette påvirker styrken af evidensen og anbefalingen (se evidensvurdering via GRADE systemet).

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

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

- P** . . . . . **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 således at man har et der går på børn under seks måneder og et der går på børn over seks måneder.
- I** . . . . . **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.

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

**O** . . . . . **Outcome:** Når outcome skal fastlægges kan det være en god ide med en bruttoliste (se eksempel herunder), hvor alle potentielt relevante outcome bringes op. Herefter kan arbejdsgruppen diskutere, hvilke de ender med at ville beholde og hvilke 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. Dette afspejles i akronymet PIRO (**P**opulation, **I**ndex test, **R**eference standard and **O**utcomes). Man skal desuden oveveje hvilken rolle den nye test skal spille i den diagnostiske udredning og om der er risiko for overdiagnostisk.

**P** . . . . . **P**opulation: Detaljeret beskrivelse af populationen, herunder sygdommen (sværhedsgrad, symptomer) med relevante demografiske faktorer (f.eks. 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.

**I** . . . . . **I**ndex 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.

**R** . . . . . **R**eference 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 f.eks. 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.

**O** . . . . . **O**utcome: 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 outcomes 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 forsvarlige. 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 referencestandarden er meget pålidelig og velunderbygget. Dette er desværre sjældent tilfældet. Hvis referencestandarden 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. referencestandarden) og som index testen identificerer som syge (detekterer)

**Specificitet:** Andel som ikke er syge (jf. referencestandarden) 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 referencestandarden da idet resultaterne af referencestandarden 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 referencestandarden er det derfor nødvendigt med patientrelevante outcome.

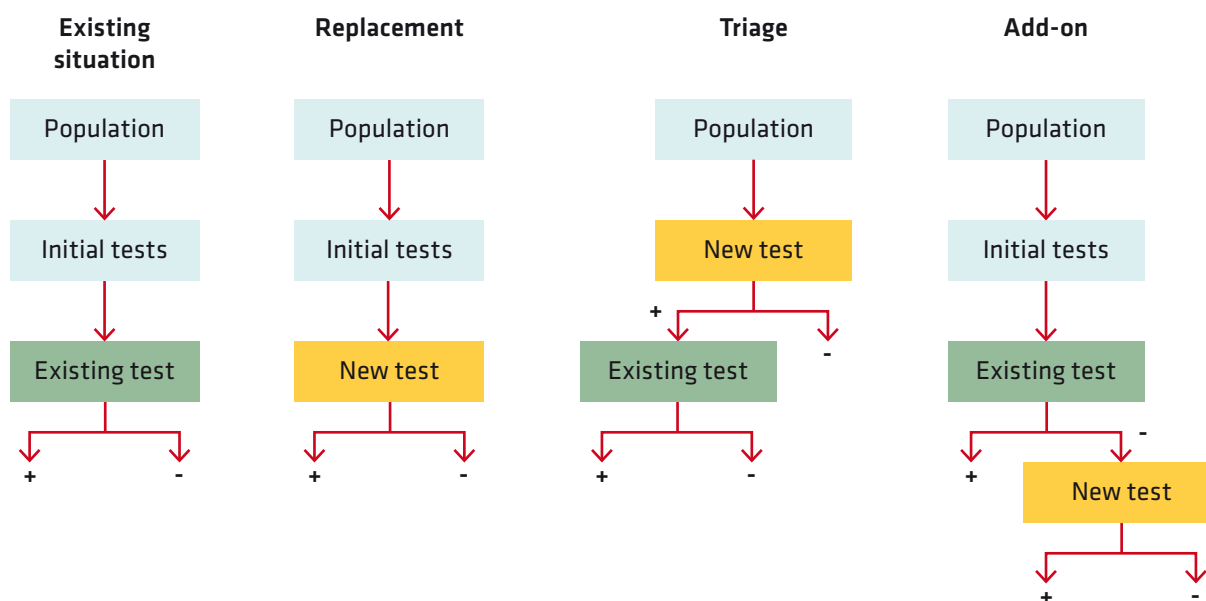
### 3.1.4 Patientrelevante outcome

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 f.eks. 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 er 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 3.4.3).
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://www.cochrane-handbook.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/>



## 3.2 Litteratursøgning

Nationale kliniske retningslinjer skal om muligt bygge på eksisterende viden fra publicerede guidelines eller systematiske oversigtsartikler. Efter behov suppleres med søgninger i primærlitteratur (randomiserede forsøg eller observationelle studier). Dette gælder i de tilfælde, hvor guidelines og systematiske oversigtsartikler ikke besvarer de fokuserede spørgsmål eller ikke er opdaterede. Den foreliggende litteratur søges og vurderes på systematisk vis, og der anvendes anerkendte metoder til søgning og vurdering.

Søgningen efter evidens forløber i tre trin. Først søges efter eksisterende retningslinjer, hvis der ikke findes nogen eller de er af dårlig kvalitet (se afsnit 3.4.1) søges dernæst efter systematiske oversigtsartikler. Findes ingen oversigtsartikler eller de er af dårlig kvalitet (se afsnit 3.4.2) søges efter primærlitteratur. Findes retningslinjer eller systematiske oversigtsartikler som er forældede opdateres deres søgning, det vil sige, der søges efter primærlitteratur publiceret siden deres søgning.

**Figur 3. Trin i litteratursøgningsprocessen**



Søgeprocessen planlægges i tæt samarbejde mellem søgespecialisten, fagkonsulenten, arbejdsgruppen og sekretariatet. Forud for første arbejdsgruppe møde søges efter retningslinjer, der søges som udgangspunkt 10 år tilbage efter retningslinjer medmindre det vurderes at perioden skal afgrænses yderligere, for eksempel på grund af teknologiske udviklinger. En eventuel søgningen efter systematiske oversigtsartikler sættes i gang så snart, det er afklaret at eksisterende retningslinjer ikke svarer på de fokuserede spørgsmål. Systematiske oversigtsartikler søges som udgangspunkt 10 år tilbage, medmindre det vurderes at perioden skal afgrænses yderligere. Konkluderes det at heller ikke systematiske oversigtsartikler besvarer de fokuserede spørgsmål søges efter primær litteratur. Primær litteratur søges som udgangspunkt også med tidsafgrænsningen på 10 år, der skal dog tages højde for eventuelle udviklinger i interventioner eller diagnostiske tests således, at der ved behov kan søges længere tilbage. Dette diskuteres individuelt for hvert enkelt fokuseret spørgsmål. Se i øvrigt vejledningen 'NKR litteratursøgningsprocessen', som findes [her](#).

Upublicerede data kan vanskeligt medtages i søgning og 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 i søgningen efter primærlitteratur kan der, hvor relevant, medtages søgning på upublicerede studier i anerkendte

databaser som f.eks. clinicaltrials.gov. Derudover kan søgningen udvides til at omfatte f.eks. posters, abstracts m.v. fra kongresser.

Hvor relevant kan der suppleres med søgninger i Sundhedsstyrelsens og EMA's databaser for bivirkninger til lægemidler og medicinsk udstyr, ligesom evidensgrundlaget for markedsføringsgodkendelse af lægemidler kan konsulteres via Sundhedsstyrelsens Enhed for Lægemiddelgodkendelse eller hentes direkte via EMA's hjemmeside. Den første litteratursøgning efter retningslinjer er en bred litteratursøgning, der giver arbejdsgruppens medlemmer et billede af eksisterende evidens i form af kliniske retningslinjer indenfor emnet. Søgningen kan bidrage til at afgrænse problemstillingen yderligere og tjene som input til diskussionen af de fokuserede spørgsmål.

Søgetermer og strategi aftales endeligt mellem fagkonsulenten, metodekonsulenten og søgespecialisten. Der er på forhånd vedtaget en liste med de obligatoriske databaser for søgning af retningslinjer, se [her](#) for en liste over hvilke databaser der anvendes.

Eventuel litteratursøgning efter systematiske oversigtsartikler og primærlitteratur foretages efter 2. møde i arbejdsgruppen i relevante databaser. Herudover kan der suppleres med mere fagspecifikke databaser (fx PEDRO, OTSEEKER). Skabelonen til søgeprotokol for søgningen vedr. retningslinjer og skabelonen til søgeprotokol for systematiske oversigtsartikler og primærlitteratur kan ses [her](#).

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. eksplicite 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.

#### ▶ **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 Referencehåndtering

RefWorks bruges som udgangspunkt til referencehåndtering medmindre retningslinjen er udarbejdet i MAGICapp, i de tilfælde bruges referencehåndteringsfunktionen i MAGICapp. 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. Refworks genererer automatisk litteraturlisterne efter det ønskede format og kan trække referencerne ind som kilder i det pågældende dokument. Hensigten er at oprette databaser for hver enkelt NKR, som fagkonsulent og arbejdsgruppen kan råde over under udarbejdelse af den enkelte NKR. En praktisk vejledning til RefWorks findes [her](#).

### 3.2.2 Afgrænsning vedrørende sprog

Der søges som udgangspunkt efter litteratur på engelsk og de skandinaviske sprog. Et "subsidiaritetsprincip" anvendes endvidere, således at litteratur prioriteres ud fra nærhed i tid og rum, det vil sige prioritering af nyere kliniske retningslinjer, systematiske oversigtsartikler og primærlitteratur fra lande, der står os nært geografisk, kulturelt og med hensyn til klinisk tradition og struktur i sundhedsvæsenet.

### 3.2.3 Søgebeskrivelse

Søgespecialisten 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, som lægges på Sundhedsstyrelsens hjemmeside [www.sst.dk](http://www.sst.dk) ved offentliggørelse af retningslinjen.

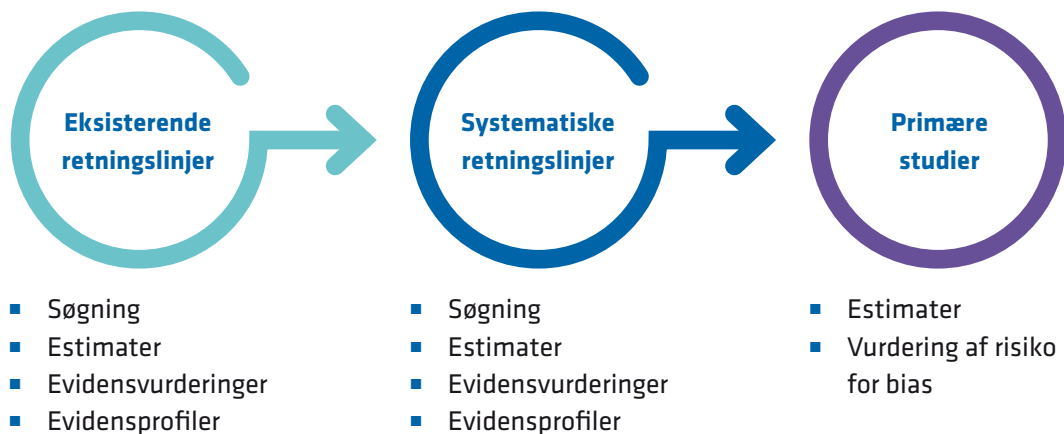
Når litteratursøgnings- og sorteringsprocessen er afsluttet udfyldes et flowdiagram, der dokumenterer hvor mange referencer der blev fundet og hvilke der blev sorteret fra. Se [her](#) for skabelonen til flowchartet.

## 3.3 Udvalgelse 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 vurderes 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. 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 4. Trin i udvælgelse af evidens**



En gennemgang af hvornår en retningslinje er af tilstrækkelig kvalitet findes i afsnit 3.4.1, en tilsvarende gennemgang for systematiske oversigtsartikler findes i afsnit 3.4.2.

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](#).

### 3.3.1 Typer af studiedesign

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

<b>Tabel 1. Typer af spørgsmål og eksempler på primære studier, der kan svare på dem</b>	
<b>Hvor mange har sygdommen?</b>	▶ Tværsnitsstudier
<b>Hvad er årsager til sygdommen?</b>	▶ Kohortestudier eller case-control studier
<b>Hvad er den bedste måde at fastslå om en person har sygdommen (diagnostik)?</b>	Kohortestudier, tværsnitsstudier, eller ▶ randomiserede studier, med en reference standard
<b>Hvad er gavnlige og skadelige virkninger af forebyggelse, screening og behandling (effekt)?</b>	▶ Randomiserede forsøg
<b>Hvordan vil det sandsynligvis gå en patient med en bestemt sygdom (prognose)?</b>	▶ 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 retnings-

linjer, 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.

### 3.3.2 **Sortering af søgeresultater**

Når søgeresultater foreligger, er det fagkonsulentens opgave at grovsortere heri 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: 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 sekundær litteratur skal der i første omgang fokuseres på, om søgningen var systematisk.

Fagkonsulenten er hovedarbejdskraft bag grovsortering. Med hensyn til den metodiske kvalitet af studierne kan metodekonsulenten bruges som sparringspartner, hvis der er studier fagkonsulenten er tvivl om.

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

I Sundhedsstyrelsen bruges online værktøjet Covidence til at sortere studier. Der er danske vejledninger til Covidence som findes [her](#).

## 3.4 **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.

### 3.4.1 **Vurdering af retningslinjer**

Udvalgte retningslinjer 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 arbejds-

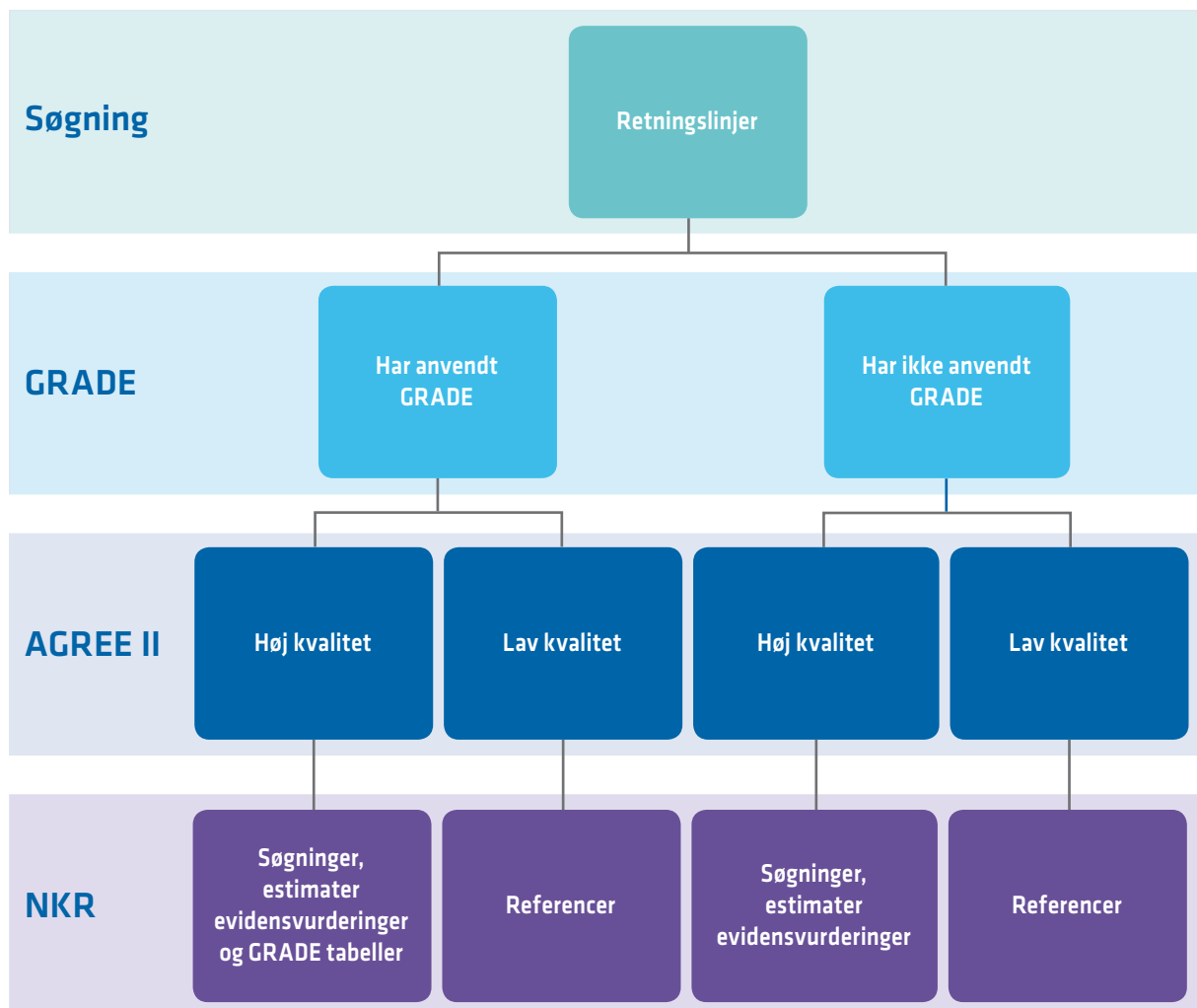
gruppemedlemmet og konsensus søges opnået ved hjælp af diskussion. I de tilfælde hvor dette ikke er muligt inddrages metodekonsulenten.

AGREE II har udviklet My AGREE PLUS, et online værktøj, der hjælper arbejdsgrupper og fagkonsulenter igennem processen. Der findes hjælp og videoer her: (<http://www.agreetrust.org/help/>). Fagkonsulenten får hjælp af værktøjet til at koordinere processen og har mulighed for f.eks. at sende reminders.

En eksisterende retningslinje kan potentielt bruges som kilde til:

- Supplerende referencer
- Søgninger
- Evidensvurderinger
- Estimer

**Figur 5. Diagram, der illustrerer processen i vurderingen af retningslinjer**



De egnede retningslinjer gennemgås af fagkonsulenten med henblik på, hvilke af fokuserede spørgsmål, de kan bruges til besvare. For at kunne vurdere hvorvidt retningslinjen er af høj eller lav kvalitet 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):

▶ **Kilde til supplerende referencer**

Kan bruges så snart retningslinjen dækker relevante fokuserede spørgsmål.

**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.

Baseret på denne gennemgang udfyldes skemaet 'Overblik over retningslinjer', skemaet findes [her](#). 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 noteres i skemaet: 'Overblik over evidens' som findes [her](#). 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 til at søge efter systematiske oversigtartikler samt for hvilke fokuserede spørgsmål, der skal foretages en opdateret søgning. For de fokuserede spørgsmål, hvor der ikke findes retningslinjer skal der igangsættes søgninger efter systematiske oversigtartikler. For de fokuserede spørgsmål, hvor der findes egnede retningslinjer skal det afgøres om deres søgninger er opdaterede eller om der skal søges efter systematiske oversigtartikler for den manglende periode.

#### **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.

### **3.4.2 Vurdering af systematiske oversigtsartikler**

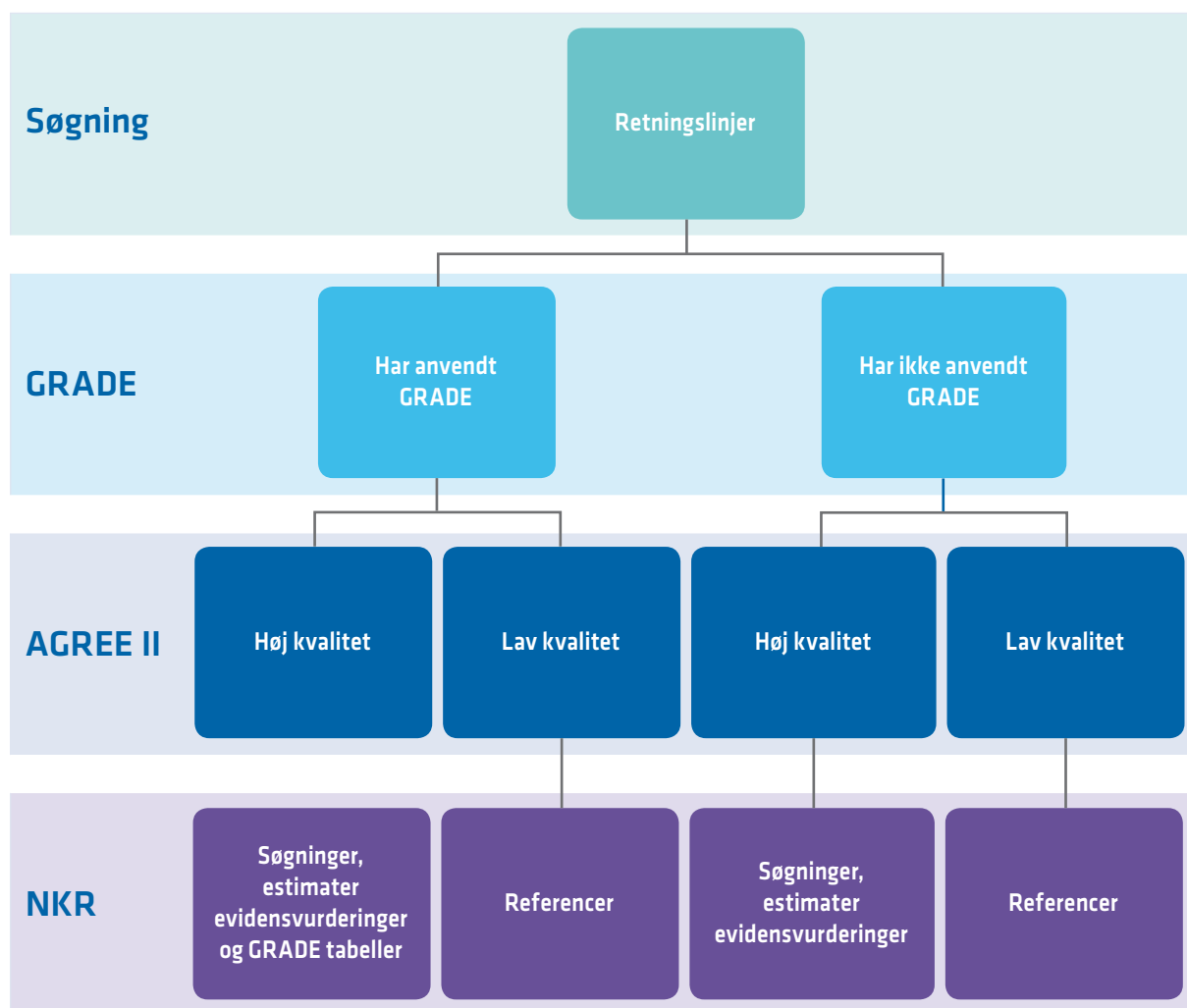
Systematiske oversigtsartikler vurderes af to personer uafhængigt af hinanden ved hjælp af AMSTAR-værktøjet (<http://amstar.ca/>). Fagkonsulenten vurderer alle udvalgte systematiske oversigtsartikler og koordinerer hvilke arbejdsgruppemedlemmer, der vurderer hvilke systematiske oversigtsartikler. I tilfælde af uenighed om vurderingen kontakter fagkonsulenten arbejdsgruppemedlemmet og konsensus søges opnået ved hjælp af diskussion. I de tilfælde hvor dette ikke er muligt inddrages metodekonsulenten. AMSTAR udfyldes i et excel ark. Arket findes [her](#).

En systematisk oversigtsartikel kan potentielt bruges som kilde til:

- Supplerende referencer
- Søgninger
- Evidensvurderinger
- Estimer



**Figur 6. 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 estimater**

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?

Baseret på denne gennemgang udfyldes skemaet 'Overblik over systematiske oversigtsartikler' Skemaet findes [her](#). 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 noteres i 'Overblik over evidens' skemaet, som 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 fortsætte til næste trin i litteratursøgningsprocessen (primær litteratur). For de fokuserede spørgsmål, hvor der ikke findes systematiske oversigtsartikler skal der igangsættes søgninger efter primære studier. For de fokuserede spørgsmål, hvor der findes egnede oversigtsartikler skal det afgøres om deres søgninger er opdaterede eller om der skal søges efter primære studier for den manglende periode.

### 3.4.3 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	<a href="http://handbook.cochrane.org/">http://handbook.cochrane.org/</a>
Observationelle studier (ko-horte, case-control)	ACROBAT-NRSI	<a href="https://sites.google.com/site/riskofbiastool/">https://sites.google.com/site/riskofbiastool/</a>
Diagnostiske primære studier	QUADAS-2	<a href="http://www.bris.ac.uk/quadas/quadas-2/">http://www.bris.ac.uk/quadas/quadas-2/</a>
Prognostiske primære studier	QUIPS	<a href="http://annals.org/article.aspx?articleid=721332">http://annals.org/article.aspx?articleid=721332</a>

Evidensvurdering bør altid foretages af to personer uafhængigt af hinanden. Se afsnit 3.5.1 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 RevMan. RevMan bruges til at samle information om primære studier, lave risk of bias figurer og meta-analyser. Metodekonsulenten vil gennemgå RevMan sammen med fagkonsulenten, hvis det er nødvendigt at inkludere primære studier. RevMan 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.
- 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.

## **3.5 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.

### **3.5.1 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 skemaer 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 for eksempel metodekonsulenten eller formanden.

#### **3.5.1.1 Retningslinjer**

Såfremt den fundne retningslinje er opdateret kan estimer direkte indtastes i MAGICapp. Hvis en supplerende søgning har afsløret nyere studier som nødvendiggør en opdatering af estimatet foretages en vurdering af hvordan dette bedst gøres sammen med metodekonsulenten.

#### **3.5.1.2 Systematiske oversigtsartikler**

Såfremt den fundne systematiske oversigtsartikel er opdateret kan estimer direkte indtastes i MAGICapp. Hvis en supplerende søgning har afsløret nyere studier som nødvendiggør en opdatering af estimatet foretages en vurdering af hvordan dette bedst gøres sammen med metodekonsulenten.

### 3.5.1.3 Primære studier

For randomiserede forsøg foretages dataekstraktionen i online værktøjet [Covidence](#). Der er danske vejledninger til Covidence en til arbejdsgruppemedlemmer og en til fagkonsulenter som findes [her](#). For andre typer studiedesigns anvendes Excel ark (se nedenstående skema for links til de relevante ark).

Tabel 3 Oversigt over skemaer/værktøjer til dataekstraktion af primære studier	
Studiedesign	Skema
Randomiserede forsøg	<a href="#">Link</a>
Observationelle studier (kohorte, case-control)	<a href="#">Link</a>
Diagnostiske primære studier	<a href="#">Link</a>
Prognostiske primære studier	<a href="#">Link</a>

### 3.5.2 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 [her](#). Meta-analyser laves i [RevMan](#).

Hvis data vurderes at være tilstrækkeligt homogene på tværs af studier er det muligt at lave meta-analyser. Metodekonsulenten har som hovedregel ansvaret for at lave meta-analyserne. Principperne for meta-analyser som beskrevet i [Cochrane Handbook for Systematic Reviews of Interventions](#) følges.

Der skal også udfyldes en outcome-matrix, der beskriver hvilke studier, der har set på hvilke outcome. Se et eksempel på en outcome-matrix [her](#).

### 3.5.3 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 [her](#).

**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.

**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>

## 3.6 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 groups [hjemmeside](#), 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 fokuserede spørgsmål, hvor også alle de kritiske og vigtige effektmål er specificerede. Efter at evidensen er identificeret, giver GRADE eksplícite 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.

### 3.6.1 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 (f.eks. 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 effekt 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. Estimerer fra observationelle studier har man som udgangspunkt lav tiltro til. Dog kan tiltroen til estimerer fra velgennemførte observationelle studier opgraderes hvis der for eksempel er store effekter eller dosis-respons sammenhænge.

**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"> <li>■ risiko for bias</li> <li>■ inkonsistens</li> <li>■ indirekte evidens</li> <li>■ unøjagtighed</li> <li>■ publikationsbias</li> </ul>	<ul style="list-style-type: none"> <li>■ effektstørrelse</li> <li>■ dosis-respons</li> <li>■ konfounding</li> </ul>	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"> <li>■ risiko for bias</li> <li>■ inkonsistens</li> <li>■ indirekte evidens</li> <li>■ unøjagtighed</li> <li>■ publikationsbias</li> </ul>	Ej relevant	Høj
				Moderat
				Lav
				Meget lav

### 3.6.2 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 kan man bruge Cochrane's Risk of Bias Tool til at vurdere randomiserede forsøg, ACROBAT til observationelle studier og QUADAS-2 til diagnostiske studier (se afsnit 3.4.3). 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 nedgra-



derer 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**

Hvis effekt-estimatet er unøjagtigt, dvs. konfidensintervallet er bredt, foretages en nedgradering. Det vurderes, om konfidensintervallet overlapper den mindste relevante forskel og om man ville komme med forskellige anbefalinger i hver sin ende af konfidensintervallet. Der vil også som udgangspunkt blive nedgraderet, hvis der kun er et studie.

### **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).

## **3.6.3**

### **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 er 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 bevirke, at effekten på mortalitet underestimeres.

#### **3.6.4 Nedgradering af evidens for præcision af diagnostiske tests**

Tværsnitstudier 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 3.4.3). 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 resultatet. 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 estimerterne.

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 referencestandard har effekt på patientrelevante outcomes, og det vil da ikke være nødvendigt at nedgradere. Hvis der er usikkerhed omkring pålideligheden af referencestandard skal anbefalingen normalt være svag.

#### ***Unøjagtighed i studier af diagnostiske test***

Ved diagnostiske tests studier skal man også vurdere bredden af konfidensintervallerne for estimerterne for den diagnostiske præcision – sensitivitet og specificitet – hvis disse er tilgængelige, eller sandt og falsk positive og negative rater, og nedgradere 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) foretages 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.



### 3.6.5 Evidensprofiler (interventionstudier)

Formålet med evidensprofiler er at give et overblik over evidensen for et givent fokuseret spørgsmål. Evidensprofiler laves ved hjælp af software, der findes forskellige alternativer på markedet. Sundhedsstyrelsen benytter [MAGICapp](#) til evidensprofiler af interventionsstudier.

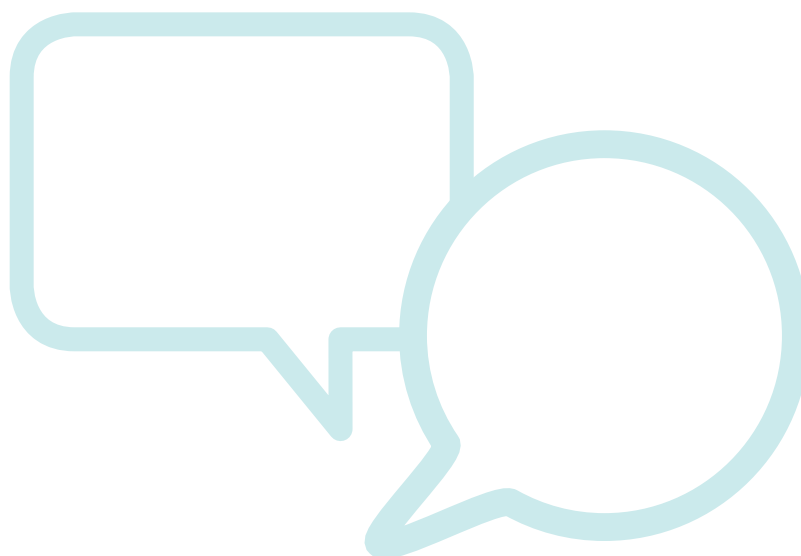
Nedenfor præsenteres et tænkt eksempel på en evidensprofil og en forklaring af de forskellige elementer. Da evidensprofiler kun er ment til at give et overblik er det nødvendigt at læse den systematiske oversigtsartikel og meta-analyse der ligger til grund for estimaterne, hvis man gerne vil vide mere om detaljerne.

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	Kritisk outcome. Risiko for bias og inkonsistente resultater
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	⊙ ⊙ ○ ○ LAV	Vigtigt outcome. Kun et ublindt studie
Livskvalitet (Quality of life) (Endt intervention)			SMD 0.11 (0.03 færre til 0.26 mere)	Højere værdi =bedre	1015 (6) 3-8	⊙ ⊙ ○ ○ LAV	Vigtigt outcome. Risiko for bias og upræcist estimat
<p>* 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.</p> <p>CI: Konfidensinterval; RR: Relativ risiko, MD: Gennemsnitlig forskel, SMD: Standardiserede gennemsnitlige forskelle</p>							
<p><b>GRADE evidensniveauer:</b></p> <p><b>Høj:</b> Vi er meget sikre på, at den sande effekt ligger tæt på den estimerede effekt.</p> <p><b>Moderat:</b> 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><b>Lav:</b> Vi har begrænset tiltro til den estimerede effekt. Den sande effekt kan være væsentligt anderledes end den estimerede effekt</p> <p><b>Meget lav:</b> Vi har meget ringe tiltro til den estimerede effekt. Den sande effekt vil sandsynligvis være væsentligt anderledes end den estimerede effekt.</p>							

**Tabel 7. Forklaring af evidensprofilen**

	Forklaring	Eksempel
<b>Population</b>	Målgruppen for anbefalingen.	Studerende
<b>Intervention</b>	Den intervention man vil komme med en anbefaling vedrørende.	Kaffe
<b>Sammenligning</b>	Alternativet til intervention.	Ingen kaffe
<b>Outcome</b>	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.
<b>Absolut effekt</b>	<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.</p> <p>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: <math>0,75 * 184 \text{ pr } 1000 = 138 \text{ pr } 1000</math>. 'Forskel med kaffe' kolonnen angiver den absolutte forskel: <math>184 \text{ pr } 1000 - 138 \text{ pr } 1000 = 46 \text{ færre pr } 1000</math>.</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>
<b>Relativ effekt</b>	<p>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.</p> <p>For kontinuerte outcome beregnes ikke nogen relativ effekt, i stedet bruges kolonnen til at opgave retningen af den anvendte skala</p>	<p>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> <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>
<b>Antal deltagere (studier)</b>	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.	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.

	Forklaring	Eksempel
<b>Evidensniveau</b>	<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>
<b>Kommentarer</b>	<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 skævvridning) 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>



### 3.6.6 Evidensprofiler (diagnostiske tests studier)

Evidensprofiler laves ved hjælp af software, der findes forskellige alternativer på markedet. Sundhedsstyrelsen benytter [GDT](#) til evidensprofiler vedrørende diagnostiske test. 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%***		
<b>Sandt positive</b> (Patienter korrekt klassificeret med borderline personlighedsforstyrrelse)	385	154	39	195 (1)	●○○○ MEGET LAV <sup>1</sup>
<b>Falsk positive</b> (Patienter ukorrekt klassificeret som ikke havende borderline personlighedsforstyrrelse)	115	46	11		
<b>Sandt negative</b> (Patienter korrekt klassificeret uden borderline personlighedsforstyrrelse)	425	680	808	195 (1)	●○○○ MEGET LAV <sup>1</sup>
<b>Falsk negative</b> (Patienter ukorrekt klassificeret som havende borderline personlighedsforstyrrelse)	75	120	142		
1) 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					

### 3.6.7 **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 arbejdsgruppens udgangspunkt i tiltroen til hvorvidt effektestimaterne er pålideligt underbyggede og kan støtte en given anbefaling.

#### **Videre læsning:**

- GRADE arbejdsgruppen  
([www.gradeworkinggroup.org](http://www.gradeworkinggroup.org))
- Artikelserie i Journal of Clinical Epidemiology om GRADE  
([http://www.gradeworkinggroup.org/publications/JCE\\_series.htm](http://www.gradeworkinggroup.org/publications/JCE_series.htm))

## 3.7 **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.





### 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 f.eks. være behov for efteruddannelse af personale, eller praktiske udfordringer i forbindelse med behandlingen.

### 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 estimerer 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.

### 3.7.2.1 **Rationale**

Her foretages en afvejning af de forskellige argumenter, der ledte til anbefalingen. Hensigten er, at gøre beslutningsprocessen gennemskuelig. 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.

### 3.7.3 **Sundhedsøkonomiske analyser**

Det er på nuværende tidspunkt desværre ikke muligt for Sundhedsstyrelsen at foretage sundhedsøkonomiske analyser.

### 3.7.4 **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. Disse anbefalinger skal så vidt muligt baseres på evidens, men selv i fravær af éntydig evidens, bør der formuleres en anbefaling med udgangspunkt i de øvrige forhold, der indgår i afvejningen.

### 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 ...*

Sundhedsstyrelsen 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 gavnlig 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 ...*

Sundhedsstyrelsen 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.
- Klinikeren 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 ...*

Sundhedsstyrelsen 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 ...*

Sundhedsstyrelsen 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.
- Klinikerne vil meget sjældent tilbyde interventionen

For yderligere beskrivelse af de forskellige evidensbaserede anbefalinger se venligst: <http://www.gradeworkinggroup.org>

### 3.7.6 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 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

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 ...*

### 3.7.7 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.

## 3.8 Implementering og monitorering

Nationale kliniske retningslinjer indeholder anbefalinger for 8-10 udvalgte og velafgrænsede kliniske problemstillinger og udgør dermed ikke én samlet behandlingsvejledning for et sygdomsområde. 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. Foreligger der eksempelvis allerede retningslinjer på et område, som anbefalingerne i en NKR omhandler, bør disse opdateres i overensstemmelse med anbefalingerne. Der vil i implementeringen af nationale kliniske retningslinjer således pågå et 'oversættelsesarbejde', potentielt gennem flere niveauer, førend anbefalingen 'når' frontlinjepersonalet i klinisk praksis. Dette skal understøtte, at anbefalingerne fungerer som beslutningsstøtte i mødet mellem sundhedsprofessionel og patient.

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. Derfor vil de nationale kliniske retningslinjer ikke blive 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 arbejdsgruppernes opgave at give forslag til få, centrale målepunkter i forhold til ønskede ændringer i praksis.

# Fase III PUBLICERING

# 4



## 4.1 Indledning

Det er fagkonsulenten, som samler og sammenfatter bidrag fra arbejdsgruppemedlemmer og fra foreliggende sekundærlitteratur til den egentlige løbende tekst i den endelige version af den nationale kliniske retningslinje. Formanden samt de øvrige i sekretariatet for den pågældende NKR herunder metodekonsulenten og projektlederen og evt. Sundhedsstyrelsens sagkyndige rådgiver på området, står til rådighed for sparring med fagkonsulenten.

## 4.2 Format

For at fremme brugervenligheden udarbejdes de nationale kliniske retningslinjer i forskellige formater, et langt PDF format, der indeholder dokumentationen og et kort PDF format (Quick guide). Derudover udarbejdes NKR i digitalt HTML5 format (MAGICapp).

Quick guiden er en 1-2 sides summarisk oversigt over de vigtigste anbefalinger i den kliniske retningslinje, som præsenteres i et brugervenligt format. De centrale anbefalinger med tilhørende styrkegraduering gengives summarisk i den korte version.

Den lange PDF version udgør den egentlige nationale kliniske retningslinje, som udgives af Sundhedsstyrelsen. Versionen bliver lagt som PDF på sst.dk inklusive bilag og kan bruges af dem, som ønsker at orientere sig yderligere i dokumentationen bag den nationale kliniske retningslinje. Rammen for den lange version er 30-50 sider inklusive referencer, men uden bilag.

Den digitale version laves i MAGICapp. 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 ind i f.eks elektroniske patient journaler.

Sproget og det faglige niveau i NKR skal henvende sig til sundhedsprofessionelle generelt.

## 4.3 Rapportskabelon

Rapporten udarbejdes i overensstemmelse med rapportskabelonen i MAGICapp.

Til rapporten udarbejdes en række bilag. Skabelon til afrapportering af flowcharts, skabelon til afrapportering af AMSTAR vurderinger samt skabelon til afrapportering af AGREE II vurderinger findes [her](#).



#### 4.4 Høring, peer reviews og godkendelse

NKR sendes til to uafhængige peer reviewere (som udgangspunkt danske eller nordiske eksperter indenfor området). Samtidigt sendes NKR i bred offentlig høring, dvs. retningslinjerne sendes til relevante interessenter vedr. emnet og lægges på den offentlige høringsportal.

Valget af egnede reviewere træffes af NKR sekretariatet efter indstilling fra arbejdsgrupperne, som forelægges emnet så tidligt, at reviewerne kan vælges i god tid samt at tidspunktet for modtagelse af høringsudkast og afgivelse af reviewet kan fastlægges tidligt i processen. Reviewerne honoreres for arbejdet.

Efter færdiggørelse af den nationale kliniske retningslinje, sendes den til godkendelse i arbejdsgruppen og herefter i det Rådgivende Udvalg for NKR. Efter omstændighederne kan denne godkendelse ske enten ved møder eller skriftligt.

