

Gradering av kvaliteten på dokumentasjonen

Gunn E. Vist¹, Ingvil Sæterdal², Per Olav Vandvik¹ og Signe A. Flottorp¹

1) Seksjon for forebyggende, helsefremmende og organisatoriske tiltak, Nasjonalt kunnskapssenter for helsetjenesten

2) Seksjon for legemidler og helseøkonomi, Nasjonalt kunnskapssenter for helsetjenesten

Korrespondanse: Gunn E. Vist, Epost: gunn.vist@kunnskapssenteret.no

SAMMENDRAG

Hva menes med at noe er ”godt dokumentert”? I hvilken grad kan vi stole på forskningsresultater? Den mest brukte metoden for å vurdere kvalitet på dokumentasjon og styrke på anbefalinger er GRADE (Grading of Recommendations Assessment, Development and Evaluation). GRADE brukes i systematiske oversikter og i retningslinjer der man vurderer effekten av og gir anbefalinger om ulike behandlingsalternativer. I GRADE gjøres et klart skille mellom kvaliteten på dokumentasjonen og styrken på anbefalingene. For å vurdere kvaliteten på dokumentasjonen tar GRADE utgangspunkt i en godt utført systematisk oversikt som omfatter den aktuelle problemstillingen. Både randomiserte forsøk og observasjonsstudier kan inngå. GRADE benytter fem kriterier som vurderes for eventuell nedgradering av kvaliteten på dokumentasjonen: studiekvalitet, konsistens mellom studiene, direktehet, presisjon og rapporteringsskjevheter. I tillegg kan observasjonsstudier vurderes for oppgradering ved følgende tre kriterier: sterke assosiasjoner, dose-responseeffekter og forvekslingsfaktorer. Alle vurderingene noteres og presenteres sammen med resultatene i oppsummeringstabeller.

Vist GE, Sæterdal I, Vandvik PO, Flottorp SA. **Grading the quality of the evidence.**

Nor J Epidemiol 2013; 23 (2): 151-156.

ENGLISH SUMMARY

What is meant by the claim that something is “well documented”? How much confidence can we have in the results? The most commonly used method to grade the quality of the evidence is GRADE (Grading of Recommendations Assessment, Development and Evaluation). GRADE is used in systematic reviews, where the effects of different treatments or options are being compared, and in guidelines with recommendations about the competing options. GRADE makes a clear distinction between the quality of the evidence and the strength of recommendation. When grading the quality of the evidence with GRADE, the starting point is a well conducted systematic review of research on the question that is being assessed. Both randomized controlled trials and observational studies can be included in the evaluation. GRADE uses five criteria to judge for possible downgrading of the quality of the evidence: risk of bias, consistency, directness, precision and reporting bias. Observational studies may in addition be assessed for possible upgrading of the quality of evidence using the following three criteria: strong association, dose response relationship, and plausible confounding. All the judgments are noted and presented together with the results in tables.

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

HVORFOR GRADERE KVALITETEN PÅ DOKUMENTASJONEN?

De fleste av oss gjør oss opp en mening om hvor mye vi stoler på de forskningsresultatene vi blir presentert – bevisst eller ubevisst. Ikke all forskning er like god: noen resultater kan vi stole på, andre bør vi kanskje ikke ha noen tro på i det hele tatt. Graderingsmetoder har til hensikt å hjelpe oss med disse vurderingene. Systematiske og transparente metoder for å utføre disse vurderingene er en fordel. Ikke bare vil det hjelpe oss å huske hvorfor vi konkluderte som vi gjorde, men det vil også gjøre det lettere for andre å vurdere om de er enige i våre vurderinger, og spesifisere hva de eventuelt er uenige i. Systematiske og transparente metoder kan også bidra til å fremme konsistente vurderinger da de samme aspektene blir vurdert fra gang til gang.

HVA ER GRADE?

Grading of Recommendations Assessment, Development and Evaluation (GRADE) er en metode som hjelper oss å vurdere kvaliteten på dokumentasjonen (hvor mye tillit vi kan ha til forskningsresultater) og styrken på anbefalinger (1,2). GRADE er utviklet over flere år av en bredt sammensatt internasjonal gruppe av forskere, klinikere, forfattere av systematiske oversikter og retningslinjeprodusenter (3). GRADE har nå blitt den mest brukte graderingsmetoden i verden. GRADE skiller seg fra de fleste andre graderingsmetoder ved følgende momenter:

- Klart skille mellom kvaliteten på dokumentasjonen og styrken på anbefalinger
- Eksplicit vurdering av hvert av de viktige utfallsmålene

- Klare kriterier for vurdering av kvaliteten på dokumentasjonen som også inkluderer muligheten for oppgradering av kvaliteten på dokumentasjonen fra observasjonsstudier

Denne artikkelen omhandler gradering av kvaliteten på dokumentasjonen ved hjelp av GRADE. En annen artikkel i dette nummeret av Norsk Epidemiologi omhandler bruk av GRADE i utvikling av kliniske retningslinjer for å sikre en systematisk og transparent prosess fra dokumentasjon til anbefalinger (4).

NÅR PASSER DET Å GRADERE KVALITETEN PÅ DOKUMENTASJONEN MED GRADE?

GRADE er utviklet for å kunne brukes i systematiske oversikter og retningslinjer der en behandling eller en strategi sammenliknes med annen behandling, eventuelt med ingen behandling. For å vurdere effekten av ulike alternativer benyttes systematiske oversikter og metodevurderinger utviklet av organisasjoner som Kunnskapssenteret, Cochrane- og Campbell- og Health Technology Assessment (HTA)-samarbeidene (5). I utviklingen av GRADE har det ofte vært benyttet eksempler fra klinisk praksis, men GRADE kan også benyttes for folkehelseiltak og tiltak som omhandler organisering av helsetjenestene. Utgangspunktet for å gradere kvaliteten på dokumentasjonen med GRADE er en systematisk oversikt av høy kvalitet som har vurdert effekten av de tiltakene som vi ønsker å sammenlikne, og som inkluderer endepunktene/utfallene vi mener er viktige. Kvaliteten på en systematisk oversikt og kvaliteten på dokumentasjonen er to atskilte vurderinger. En systematisk oversikt av høy kvalitet kan inneholde dokumentasjon av lav kvalitet, for eksempel. Gradering av kvaliteten på dokumentasjonen ved hjelp av GRADE gjøres i økende grad som del av arbeidet med å utvikle en systematisk oversikt. I noen systematiske oversikter finnes ingen inkluderte studier ("tomme" oversikter), som betyr at vi mangler dokumentasjon om denne sammenlikningen. Noen ganger finnes kun én studie som kan inkluderes og dermed utgjør denne ene studien hele dokumentasjonsgrunnlaget, og graderingen må baseres på denne ene studien.

HVA MENES MED KVALITETEN PÅ DOKUMENTASJONEN?

Med kvaliteten på dokumentasjonen mener vi i hvor stor grad vi kan stole på at effektestimater er korrekte, eller hvor stor tillit vi har til resultatene. Selv om grad av tillit er en kontinuerlig størrelse, er kvaliteten på dokumentasjonen av praktiske grunner delt inn i fire kategorier i GRADE: høy, middels, lav og svært lav (se tabell 1).

HVORDAN BRUKE GRADE?

Utgangspunktet for GRADE er altså en systematisk oversikt av høy kvalitet som omhandler effekt av aktu-

ell behandling med relevant sammenlikning, der resultater for de viktigste utfallene er rapportert.

For hver sammenlikning utvikles en såkalt evidensprofil der hvert av de viktige utfallene vurderes etter gitte kriterier. Evidensprofilen viser tydelig (med bruk av fotnoter) hvilke vurderinger som er utført for hver gradering.

Kvaliteten på dokumentasjonen, eller hvor stor tillit vi har til effektestimater, vurderes for hvert av de viktige utfallene. Vi starter med å notere hvilke *studie-design* som er benyttet i de enkelte studiene som inngår i den systematiske oversikten. Randomiserte kontrollerte studier er mindre utsatt for påvirkninger og feilkilder (6) og starter på høy kvalitet. Studier som ikke er randomiserte (observasjonsstudier) starter på lav kvalitet. Dersom den systematiske oversikten har inkludert studier av ulikt design for de ulike utfallene, graderes kvaliteten på dokumentasjonen fra ulike studiedesign hver for seg.

Tabell 2 viser en oppsummering av de åtte kriteriene vi bruker i GRADE, fem som kan resultere i nedgradering og tre som kan oppgradere kvaliteten på dokumentasjonen. Hver gang vi opp- eller nedgraderer blir begrunnelsen notert. Både randomiserte kontrollerte studier og observasjonsstudier vurderes for nedgradering.

Fem kriterier for nedgradering av kvaliteten på dokumentasjonen

Studiekvalitet (risiko for systematiske feil/ skjevheter, engelsk: "risk of bias"). Her vurderer vi studiens planlegging og utførelse særlig med tanke på randomiseringsprosess, blinding av deltakere og forskere, frafall fra studien og analysemetode. Disse vurderingene bør foretas av de som har utført den systematiske oversikten. Dersom det mangler en tilstrekkelig grundig beskrivelse av randomiseringsprosedyren eller av hvordan allokeringen ble holdt skjult for deltakere og forskere, nedgraderes kvaliteten. Dersom det, for eksempel er stort frafall av pasienter i den ene gruppen og dette ikke er tatt hensyn til ved intention-to-treat-analyse, nedgraderes kvaliteten for dette utfallet. Om det både er begått feil ved randomiseringen og det er stort frafall i den ene gruppen som det ikke er justert for, graderes det ned med to nivåer. Slik kan dokumentasjon fra randomiserte forsøk ende opp som "lav kvalitet" allerede etter vurdering av det første kriteriet.

Siden GRADE graderer kvaliteten på dokumentasjonen for hvert viktig utfall på tvers av studiene som inngår i den systematiske oversikten, vil alle de inkluderte studiene som har rapportert på utfallet bli vurdert for hvert av kriteriene. Ofte er det slik at studiekvaliteten til de forskjellige studiene som har rapportert et utfall varierer. Da må det gjøres en vurdering av hvilke studier man tar mest hensyn til. Studiekvaliteten på de studiene som har flest hendelser er utslagsgivende.

Konsistens eller samsvar mellom studiene. Dersom resultatene fra flere forskjellige studier peker i samme retning har vi større tillit til resultatene enn om de

Tabell 1. Kategorier av kvaliteten på dokumentasjonen etter GRADE.

Høy kvalitet ⊕⊕⊕⊕	Vi har stor tillit til at effektestimater ligger nær den sanne effekten.
Middels kvalitet ⊕⊕⊕○	Vi har middels tillit til effektestimater: effektestimater ligger sannsynlig nær den sanne effekten, men effektestimater kan også være vesentlig ulik den sanne effekten.
Lav kvalitet ⊕⊕○○	Vi har begrenset tillit til effektestimater: effektestimater kan være vesentlig ulikt den sanne effekten.
Svært lav kvalitet ⊕○○○	Vi har svært liten tillit til at effektestimater ligger nær den sanne effekten.

Tabell 2. Oppsummering av kriteriene for gradering med GRADE.

Kvaliteten på dokumentasjonen	Studiedesign	Nedgrader ved	Oppgrader ved*
Høy	Randomisert kontrollert forsøk	Begrensninger ved studie kvaliteten (risiko for systematiske feil) -1 Alvorlig -2 Veldig alvorlig	Sterk sammenheng +1 Sterk +2 Veldig sterk
Middels		Mangel på samsvar -1 Alvorlig -2 Veldig alvorlig	Dose-responseeffekt +1 holdepunkter for dose-responseeffekt
Lav	Observasjonsstudier	Mangel på direktet -1 Alvorlig -2 Veldig alvorlig eller flere faktorer	Forvekslingsfaktorer +1 Alle kjente forvekslingsfaktorer ville ha redusert effekten
Svært lav		Mangel på presisjon -1 Alvorlig -2 Veldig alvorlig	
		Rapporteringsskjvhet -1 Sannsynlig -2 Veldig sannsynlig	

*gjelder kun der det er flere samsvarende studier som ikke er nedgradert.

spriker (heterogenitet). Der det er uforklarlig heterogenitet nedgraderer vi kvaliteten. Er det gode forklaringer på heterogeniteten, for eksempel at resultatene fra alle studiene med én pasientgruppe var forskjellige fra alle resultatene med en annen pasientgruppe, kan det vurderes å analysere disse (pasientgruppene) hver for seg. Dersom det kun foreligger én studie som har målt det aktuelle utfallet, kan vi ikke si noe om grad av samsvar med andre studier, og vi vurderer denne usikkerheten når vi vurderer presisjon (se nedenfor).

Direkthet omhandler sammenliknbarhet og inkluderer vurderinger av hvor generaliserbare forskningsresultatene er. Disse vurderingene om direktet forutsetter kunnskap om hvor og hvordan dokumentasjonen skal benyttes. Denne informasjonen vil ofte være mer relevant i forbindelse med utvikling av anbefalinger for et spesifikt formål enn i utarbeidelsen av en systematisk oversikt for en breiere målgruppe, slik for eksempel en systematisk oversikt fra Cochrane-samarbeidet vil være. I vurderingen av direktet inngår spørsmål som:

Er *populasjonen* (pasientene) i studiene tilstrekkelig lik den populasjonen som dokumentasjonen skal anvendes på? Dersom populasjonene er så ulike at vi er usikre på om vi kan forvente liknende resultater der vi skal bruke dokumentasjonen, for eksempel om vi vurderer et tiltak for barn og alle studiene er utført på godt voksne, så nedgraderer vi kvaliteten på dokumentasjonen. Er *tiltaket* (og sammenlikningen/kontrollen) likt nok sammensatt og likt nok gjennomført i forskningsstudiene til at vi kan forvente tilsvarende effekter hvis det innføres i praksis hos oss? For eksempel vurderer vi om tiltaket er gitt med samme dosering eller intensitet i studiene, som det vil bli i praksis. Er de viktige *utfallene* målt direkte og med valide metoder? Bruk av surrogatutfall (markører for sykdom eller sykdomsrisiko, for eksempel kolesterolnivå) kan gi misvisende resultater, og der det mangler påviste sterke assosiasjoner mellom surrogatutfall og viktige utfall nedgraderer vi kvaliteten for dette utfallet. Dersom det ikke finnes en *direkte sammenlikning* mellom de to tiltakene som

Tabell 3. GRADE evidensprofil for bruk av hjelm for motorsykelkjørere.

Kvalitetsvurderinger							Antall pasienter		Effekt		Kvalitet	Viktig
Antall studier	Design	Studiekvalitet	Konsistens	Direkthet	Presisjon	Andre faktorer	Bruk av hjelm	Ikke brukt hjelm	Relative (95 % KI)	Absolutt		
Død (vurdert med justerte OR)												
4	Observasjonsstudier	Ingen alvorlige begrensninger	Ingen uforklart heterogenitet	Ingen alvorlige forskjeller	Ikke mangel på presisjon	sterk assosiasjon ¹	541/14405 (3,8 %)	618/10546 (5,9 %)	OR 0,58 (0,5 til 0,68)	24 færre per 1000 (fra 18 færre til 28 færre)	⊕⊕⊕○ MODERAT	KRITISK
Hodeskade (vurdert med justerte OR)												
6	Observasjonsstudier	Ingen alvorlige begrensninger	Ingen uforklart heterogenitet	Ingen alvorlige forskjeller	Ikke mangel på presisjon	sterk assosiasjon ¹	1398/5936 (23,6 %)	2365/5899 (40,%)	OR 0,31 (0,25 til 0,38)	229 færre per 1000 (fra 198 færre til 258 færre)	⊕⊕⊕○ MODERAT	KRITISK
Nakkeskade (vurdert med ujustert OR)												
12	Observasjonsstudier	Alvorlig begrensning ²	Ingen uforklart heterogenitet	Ingen alvorlige forskjeller	Ikke presist ³	nei	126/7583 (1,7 %)	211/5784 (3,6 %)	OR 0,85 (0,66 til 1,09)	5 færre per 1000 (fra 12 færre til 3 flere)	⊕○○○ SVÆRT LAV	KRITISK
Ansiktsskade (vurdert med ujustert OR)												
8	Observasjonsstudier	Alvorlig begrensning ²	Uforklarlig heterogenitet ⁴	Ingen alvorlige forskjeller	Ikke mangel på presisjon	nei	1504/10458 (14,4 %)	1856/7145 (26 %)	OR 0,41 (0,32 til 0,52)	134 færre per 1000 (fra 105 færre til 159 færre)	⊕○○○ SVÆRT LAV	KRITISK

¹ Sterk assosiasjon

² Ikke justerte analyser der de burde ha vært justerte

³ Brede konfidensintervaller som inkluderer både store fordeler og mulig skade

⁴ Uforklarlig heterogenitet, I² 66 %

sammenliknes, kun indirekte sammenlikning (for eksempel at tiltak A og B er sammenliknet med placebo hver for seg, men ikke direkte mot hverandre), nedgraderer vi for dette.

Presisjon dreier seg om hvor presise resultatene er, og om hvor mye data som er tilgjengelig. Der vi har meta-analyse av flere store studier med mange hendelser er det mer presise effektestimater enn der vi for eksempel kun har én liten studie eller flere små studier med få hendelser. Disse har ofte brede konfidensintervaller som gjenspeiler stor usikkerhet og kan omfatte vesentlige forskjeller i effekt, og da nedgraderer vi kvaliteten på utfallet.

Rapporteringsskjevheter omhandler både publiserings-skjevhet og selektiv utfallsrapportering. Dette er en kombinasjon av flere faktorer. Det er mer sannsynlig at studier som viser en positiv effekt blir publisert enn at studier som ikke viser noen forskjeller blir publisert (7). Dessuten kan det være fristende for forskerne å vise frem utfallene med signifikante resultater heller enn ikke-signifikante resultater – selv om disse i utgangspunktet var regnet som vel så viktige. Ikke minst kan det være interessekonflikter som gjør at man ikke ønsker å vise frem dårlige resultater av et tiltak man ønsker å promotere. Dersom vi har grunn til å tro at det foreligger rapporteringsskjevheter nedgraderer vi kvaliteten for dette utfallet.

Tre kriterier for oppgradering av kvaliteten på dokumentasjonen

De tre kriteriene som vurderes for oppgradering benyttes kun der det er flere samsvarende studier som

ikke er nedgradert:

- Sterke eller veldig **sterke assosiasjoner**/sammenhenger mellom intervensjon og utfall. Det vil si at den beregnede effekten er så stor at det er usannsynlig at den skyldes tilfeldigheter.
- Store eller veldig store **dose-responseeffekter**.
- Der alle sannsynlige **forvekslingsfaktorer** (confounders) ville ha bidratt til å redusere effektestimater.

Gradering med GRADE innebærer en rekke vanskelige og til dels subjektive vurderinger. Hvor stort fall eller hvor stor heterogenitet skal til for at vi bør gradere ned kvaliteten på dokumentasjonen, for eksempel? Dersom man er usikker på to kriterier, kan en samlet vurdering tilsi å nedgradere én gang (der det å nedgradere for hvert av dem kan synes for strengt).

Av og til er det slik at de viktige utfallene ikke har blitt rapportert i de inkluderte studiene. Disse utfallene er ikke mindre viktige av den grunn, og det bør derfor komme klart fram at vi ikke vet hvordan intervensjonen innvirker på disse. Derfor blir de viktige utfallene presentert i GRADE-tabellene uavhengig av hvor mye eller med hvilken kvalitet det er på dokumentasjonen, også om dokumentasjon mangler helt.

HVA ER GRADEPRO?

Arbeidsgruppen bak GRADE har utviklet en programvare, GRADEpro, som kan brukes når vi vurderer kvaliteten på dokumentasjonen, og som automatisk genererer "evidensprofiler" (se tabell 3). GRADEpro kan lastes ned gratis fra Internett (8).

Tabell 4. GRADE SoF tabell for bruk av hjelm ved motorsykkelkjøring sammenliknet med å ikke bruke hjelm.**Pasient eller populasjon:** Motorsykkelkjørere**Setting:** Studiene var utført i USA, Sverige, Taiwan, Thailand**Intervensjon:** Bruk av hjelm**Sammenlikning:** Ikke bruk av hjelm

Utfall	Sammenlikning av risiko i de to gruppene (95 % KI)		Relativ effekt (95 % KI)	Antall deltakere (studier)	Kvaliteten på dokumentasjonen (GRADE)	Kommentar
	Antatt risiko Ikke bruk av hjelm	Tilsvarende risiko Bruk av hjelm				
Dødelighet Justerte OR	Gjennomsnittlig antall som døde var 59 per 1000	Gjennomsnittlig antall som døde var 35 per 1000 (30 til 41)	OR 0,58 (0,5 til 0,68)	24591 (4 studier)	⊕⊕⊕⊖ Moderat¹	
Hodeskade Justerte OR	Gjennomsnittlig antall som fikk hodeskade var 401 per 1000	Gjennomsnittlig antall som fikk hodeskade var 172 per 1000 (143 til 203)	OR 0,31 (0,25 til 0,38)	11835 (6 studier)	⊕⊕⊕⊖ Moderat¹	
Nakkeskade	Gjennomsnittlig antall som fikk nakkeskade var 36 per 1000	Gjennomsnittlig antall som fikk nakkeskade var 31 per 1000 (24 til 40)	OR 0,85 (0,66 til 1,09)	13367 (12 studier)	⊕⊖⊖⊖ Svært lav¹	
Ansiktsskade	Gjennomsnittlig antall som fikk ansiktsskade var 260 per 1000	Gjennomsnittlig antall som fikk ansiktsskade var 126 per 1000 (101 til 154)	OR 0,41 (0,32 til 0,52)	17603 (8 studier)	⊕⊖⊖⊖ Svært lav¹	

¹ Sterk assosiasjon² Ikke justerte analyser der de burde ha vært justerte³ Brede konfidensintervaller som inkluderer både store fordeler og mulig skade⁴ Uforklarlig heterogenitet, I² 66 %

HVA KOMMER UT AV EN GRADERING MED GRADE?

Ved at vi systematisk noterer de vurderingene som vi foretar og de effektestimaterne som vi vurderer i GRADEpro, blir resultatet av vurderingene en såkalt evidensprofil. Tabell 3 viser en GRADE evidensprofil for bruk av hjelm når man kjører motorsykkel. Denne evidensprofilen er utviklet basert på informasjon hentet fra en Cochrane systematisk oversikt (9), og omfatter fire utfall: død, hodeskade, nakkeskade og ansiktsskader. De fleste av disse studiene er utført i USA, men også enkelte andre land (Thailand, Taiwan og Sverige). Vi ser at utfallet død var rapportert i fire observasjonsstudier og at disse fire studiene omfattet 24 951 personer. Disse studiene var godt utført, og hadde samsvarende resultater. Vi har vurdert både populasjonen (motorsykkelkjørerne), tiltaket (bruk av hjelm) og sammenlikningen (ikke bruk av hjelm), og kommet til resultatene antakelig er overførbare. Vi hadde heller ingen grunn til å mistenke rapporterings-skjevheter så vi har ikke nedgradert for det. Effekt-estimatet for død viser en stor assosiasjon (justert odds ratio 0,58 (95 % KI 0,50 til 0,68)), og vi har derfor oppgradert for dette utfallet. Det vil si at vi har vurdert (ved hjelp av GRADE) at vi har moderat kvalitet på dokumentasjonen for utfallet død. For utfallet nakkeskade som er rapportert i 12 studier og omfatter 13 367 personer, var det vesentlige forskjeller mellom gruppene som sammenliknes, og dette var det ikke justert for i analysene. Konfidensintervallet rundt effekt-estimatet for nakkeskader var så bredt at verken fordeler eller ulemper ved bruk av hjelm kunne utelukkes (OR 0,85

(95 % KI 0,66 til 1,09). Vår vurdering var derfor at det er svært lav kvalitet på dokumentasjonen for at bruk av hjelm forebygger nakkeskade. Når det gjaldt ansiktsskade, som er rapportert i åtte studier og omfatter 17 603 personer, var det også betydelige forskjeller mellom gruppene som ble sammenliknet, uten at det var justert for i analysene. Derfor har vi ikke kunnet oppgradere dette utfallet selv om resultatene tyder på at det også her kan være en sterk assosiasjon.

Tabell 3 med GRADE evidensprofil viser alle vurderingene og effektestimaterne for bruk av hjelm på motorsykkel. Det er en omfattende tabell som inkluderer mer informasjon enn mange ønsker seg. GRADE Working Group har derfor også utviklet en oppsummeringstabell, en *Summary of Findings table* (SoF tabell). Tabell 4 viser GRADE SoF tabellen for sammenlikningen som vi presenterte i GRADE evidensprofilen i tabell 3.

HVOR FINNER JEG MER INFORMASJON OM GRADERING MED GRADE?

Programvaren GRADEpro har en hjelpefil som beskriver metodikken i detalj og som er rikt illustrert med eksempler. GRADE Working Group skriver jevnlig artikler rettet mot personer som ønsker å benytte GRADE metodikken, og disse publiseres løpende i tidsskriftet *Journal of Clinical Epidemiology* (10). Her gis detaljerte beskrivelser av alle trinnene både for å gradere kvaliteten på dokumentasjonen og å vurdere styrken på anbefalinger, samt prosesser omkring utvikling av retningslinjer.

HVA ER FORDELENE OG ULEMPENE MED Å VURDERE KVALITETEN PÅ DOKUMENTASJONEN MED GRADE?

Sentrale prinsipper for bruk av GRADE er at vurderingene gjøres eksplisitte og transparente. Denne åpenheten er en vesentlig fordel med GRADE, men kravet om at alle vurderinger skal dokumenteres kan by på utfordringer. Det kan være vanskelig å formulere en begrunnelse for alle vurderinger som gjøres, og gradering med GRADE er mer tidkrevende enn mange andre graderingsmåter. På den annen side er denne åpenheten en styrke ved formidling og deling av infor-

masjon, og ikke minst nyttig når den skal vurderes av andre.

For at en metode skal være klar og eksplisitt fordrer en viss grad av detaljinstruksjon. Utviklerne av GRADE har forsøkt å finne en akseptabel balanse mellom enkelhet i bruk, og grundighet. Gradering av kvaliteten på dokumentasjonen ved hjelp av GRADE fordrer at man har tilstrekkelig metodekompetanse til å kunne gjøre de vurderingene som inngår i graderingen. Mange vil også trenge opplæring i bruk av GRADE. Den nevnte artikkelserien (10) og hjelpefilen i programvaren GRADEpro (8) er nyttige ressurser; dessuten holdes det jevnlig kurs i bruk av GRADE.

REFERANSER

1. Atkins D, Best D, Briss PA, Eccles M, Falck Ytter Y, Flottorp S, et al; Grade Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; **328**; 1490.
2. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendation. *BMJ* 2008; **336**: 924-926.
3. Grade Working Group. www.gradeworkinggroup.org (13.08.13)
4. Vandvik PO, Berg R, Vist G. En ny generasjon troverdige kliniske retningslinjer. *Norsk Epidemiologi* 2013; **23** (2): 197-204.
5. Jamtvedt G. Systematiske oversikter om effekt av tiltak. *Norsk Epidemiologi* 2013; **23** (2): 119-124.
6. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann HJ, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database of Systematic Reviews* 2011, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.pub3.
7. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews* 2009, Issue 1. Art. No.: MR000006. DOI: 10.1002/14651858.MR000006.pub3.
8. GRADEpro [dataprogram], Version 3.2 for Windows. Brozek J, Oxman AD, Schünemann H, 2008. <http://ims.cochrane.org/revman/grade> (13.08.13).
9. Liu BC, Ivers R, Norton R, Boufous S, Blows S, Lo SK. Helmets for preventing injury in motorcycle riders. *Cochrane Database of Systematic Reviews* 2008, Issue 1. Art. No.: CD004333. DOI: 10.1002/14651858.CD004333.pub3.
10. Guyatt GH, Oxman AD, Akl E, Kunz R, Vist G, Brozek J, et al. GRADE guidelines 1. Introduction – GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011; **64**: 383-394.