Handleiding voor het gebruik van de GRADE-methode voor het effect van interventies

Trudy Bekkering[1,2,3], Nicolas Delvaux[3,4], Anne-Catherine Vanhove[2,5], Martine Goossens[1,4], Nicole Dekker[4]

1. Cebam (Centrum voor Evidence-Based Medicine), Leuven
2. Cochrane Belgium, Leuven
3. Academisch Centrum voor Huisarts Geneeskunde, KU Leuven
4. WOREL (Werkgroep ontwikkeling richtlijnen eerstelijn), Antwerpen
5. CeBaP (Centrum voor Evidence-Based Practice), Rode Kruis Vlaanderen, Mechelen

Leuven, 31 November 2020

Achtergrond

Deze handleiding beschrijft de belangrijkste stappen van het beoordelen van de kwaliteit van het wetenschappelijk bewijs en het bepalen van de sterkte van aanbevelingen, zoals uitgewerkt door de *Grading of Recommendations, Assessment, Development and Evaluation* (GRADE) Werkgroep. Voor meer details over alle stappen verwijzen wij naar het GRADE Handboek (1).

Het doel van de GRADE Werkgroep was om te komen tot een uniforme, transparante methode om de kwaliteit van het wetenschappelijk bewijs te beoordelen (m.a.w. hoe zeker is men van dat bewijs? Verderop spreken we daarom van "zekerheid van bewijs") en om aanbevelingen te formuleren (m.a.w. beveel je een interventie aan of niet en hoe sterk kan iets worden aanbevolen?). Door het rapporteren van alle beslissingen gemaakt bij het doorlopen van de GRADE-stappen wordt dit proces transparant. Inmiddels gebruiken steeds meer internationale organisaties, die aanbevelingen formuleren over de gezondheidszorg, de GRADE-methode.

De GRADE-methode is ontwikkeld voor systematische reviews en richtlijnen met klinische vragen over het effect van interventies. Inmiddels bestaan er ook GRADE-methoden voor andere onderzoeksvragen, zoals diagnosestelling, prognose, kwalitatieve vragen (zie literatuurlijst). Een andere GRADE-methode is beschikbaar voor systematische reviews waarin de resultaten niet gepoold kunnen worden en is gekozen voor een narratieve manier van synthetiseren (zie literatuurlijst). En worden er nog meer ontwikkeld. In deze handleiding beschrijven we de GRADE-methode voor het effect van interventies.

Het is mogelijk een GRADE-beoordeling te doen op een enkele studie, als bijvoorbeeld maar 1 studie een bepaalde uitkomstmaat presenteert. Zie ook verderop voor meer informatie hierover.

De resultaten van de systematische review worden gepresenteerd in een Summary of Findings (SoF)-tabel Tabel 1). Dit is de basis waarop een GRADE-beoordeling van de zekerheid van het bewijs kan plaatsvinden. Een meer gedetailleerd overzicht van de redenen om het niveau van zekerheid te downgraden of upgraden noemt een GRADE-profiel (Tabel 2). Soms worden beide onderdelen in 1 tabel samengenomen.

Tabel 1. Voorbeeld van een Summary of Findings-(SoF) tabel (bron: Leopoldino et al, 2019)

Paracetamol versus placebo voor artrose van de heup of knie						
Patiënten: mensen met artrose van heup of knie Setting: verscheidene revalidatie, orthopedische or reuma klinieken Interventie: Paracetamol Comparison: Placebo						
Uitkomsten	Verwachtte absolute effectgrootte (95%CI)		Relative effect (95%CI)	N deelnemers (studies)	Zekerheid bewijs (GRADE)	Opmerkingen
	Placebo	Paracetamol				
Gemiddeld verschil in pijn (0-100 schaal) (3 tot 12 wk)	Gemiddeld verschil was -23	Gemiddeld verschil is 3.2 punten lager (1.0 lager tot 5.4 lager)	MD -3.2 (-5.4 tot -1.0)	2355 (7 RCTs)	Hoog	Klinisch niet belangrijk verschil
Drop-out door neveneffecten (24 wks)	65 per 1000	77 per 1000 (59 tot 100)	RR 1.19 (0.91 tot 1.55)	3023 (7 RCTs)	Matig[a]	Verschil zowel statistisch als klinisch niet relevant
Voetnoten: MD=mean difference; RR=relatieve risk; CI=confidence interval; RCT=randomised controlled trial a. Gebaseerd op score in placebogroep uit studie van Micelli-Richard 2004 b. Gedowngraded met 1 niveau vanwege imprecisie (gering aantal events of 95%BI sluiten een klinisch belangrijk effect niet uit).						

Tabel 2. Voorbeeld van een GRADE-profiel (gebaseerd op SoF hierboven)

Beoordeling van zekerheid							Resultaten		Effect		Zekerheid	Belang
N studies	Design	Risico op bias	Inconsi-stentie	Indirect-heid	Imprecisie	Andere	placebo	paracetamol	Relatieve effect	Absolute effecten		
Pijn, schaal 0-100 (gemiddeld verschil, follow-up tussen 3 en 12 weken)												
7	RCT	niet ernstig	niet ernstig	niet ernstig	niet ernstig	niet ernstig	-23	-26.2	MD -3.2 (-5.4 tot -1.0)	-3.2 punten lager (1.0 lager tot 5.4 lager)	Hoog	Kritisch
Dropout door neveneffecten (24 weken)												
7	RCT	niet ernstig	niet ernstig	niet ernstig	ernstig[a]	niet ernstig	120/1630 (7.4%)	90/1393 (6.5%)	RR 1.19 (0.91 tot 1.55)	22 meer per 1000 (6 minder tot 35 meer)	Matig	Kritisch

MD=mean difference; RR=relatieve risk

a. Gedowngraded met 1 niveau vanwege imprecisie (gering aantal events of 95%BI sluiten een klinisch belangrijk effect niet uit).

In welke fasen van de richtlijnontwikkeling gebruik je GRADE?

De GRADE-methode bevat specifieke instructies voor 3 onderdelen van het richtlijnontwikkelproces:

1. Selecteren en prioriteren van de uitkomstmaten
2. Beoordelen van de zekerheid van het bewijs
3. Komen van bewijs tot een aanbeveling

Hoe pas je GRADE praktisch toe?

1. Selecteren en prioriteren van de uitkomstmaten

Voor een klinische vraag van een richtlijn zijn veel verschillende uitkomstmaten te bedenken. Richtlijnontwikkelingsgroepen moeten echter pragmatisch zijn en die uitkomstmaten selecteren die voor een patiënt het meest belangrijk zijn om over de klinische vraag een beslissing te nemen. Om deze reden is het ook aan te bevelen om patiënten of patiëntvertegenwoordigers al in deze stap van de richtlijnontwikkeling te betrekken, liefst als lid van de richtlijnontwikkelingsgroep.

De richtlijnontwikkelingsgroep moet uitkomstmaten selecteren op basis van het belang voor een patiënt en niet op basis van het beschikbare bewijs. Dus ook als er geen studies zijn over een uitkomstmaat die voor een patiënt heel belangrijk is, dan moet deze uitkomstmaat wel worden geselecteerd. Dat er geen gegevens beschikbaar zijn, zal blijken uit de SoF-tabel die voor deze uitkomstmaat een lege cel presenteert. Dit geeft ook aan dat er nood is om deze uitkomstmaat in toekomstige studies mee te nemen. Bij het selecteren dient een richtlijnontwikkelingsgroep zowel positieve (effecten) als negatieve (neveneffecten) uitkomstmaten mee op te nemen.

Tijdens deze stap is het belangrijk om zo transparant mogelijk te werken. Dit kan door allereerst alle uitkomsten die mogelijk relevant zijn voor patiënten op te lijsten. Vervolgens beoordelen alle leden van de richtlijnontwikkelingsgroep iedere uitkomst op een schaal van 1 tot 9 per uitkomst het belang ervan voor patiënten om tot een beslissing te komen. Een score van 1 tot 3 staat voor weinig belang, een score van 4 tot 6 voor belangrijk en een score 7 tot 9 voor kritisch (zie Tabel 3). Alle leden van de richtlijnontwikkelingsgroep kennen aan iedere uitkomstmaat een score toe. Daarna worden de resultaten samengenomen en gepresenteerd, waarna binnen de groep consensus wordt bereikt over de meest belangrijke uitkomstmaten.

Tabel 3. Indeling van de uitkomstmaten in drie categorieën volgens hun belang voor het maken van een beslissing.

1 2 3	Van beperkt belang
4 5 6	Belangrijk
7 8 9	Van kritisch belang

Alleen kritische en belangrijke uitkomstmaten hebben een impact op de aanbevelingen in de richtlijn. Het rangschikken van de uitkomstmaten helpt de richtlijnontwikkelingsgroep om te focussen op de meest belangrijke uitkomstmaten. Ook de beoordeling van de zekerheid van wetenschappelijk bewijs gebeurt alleen voor de kritische en belangrijke uitkomstmaten.

Het belang van uitkomstmaten zal waarschijnlijk variëren naargelang van het perspectief. Zo kan een patiënt andere belangen hebben dan een arts of een beleidsmaker. Richtlijnontwikkelingsgroepen moeten beslissen vanuit welk perspectief zij de uitkomstmaten beoordelen en dit documenteren in hun richtlijn. Dit draagt ook bij aan de transparantie van het proces.

De voorkeuren van patiënten m.b.t. uitkomsten worden indien mogelijk gebaseerd op wetenschappelijk bewijs. De richtlijnontwikkelingsgroep kan zich hiervoor baseren op studies of in het ideale geval hierover zelf een systematische review doen. Dergelijke studies geven inzicht in de variatie tussen voorkeuren van patiënten, hoe een patiënt de ziekte of de behandeling ervaart met eventuele neveneffecten en de afweging tussen de effecten en de risico's van een interventie. Als dergelijk bewijs niet beschikbaar is, kan de richtlijnontwikkelingsgroep haar ervaring met eerdere patiënten gebruiken om de voorkeuren en waarden van patiënten in te schatten. Het is belangrijk om te documenteren hoe zij de voorkeuren van patiënten voor de uitkomstmaten hebben bepaald.

2. Beoordelen van de zekerheid van het bewijs

De definitie die GRADE hanteert voor zekerheid van bewijs verschilt enigszins voor een richtlijn en voor een systematisch review. Voor een richtlijn gaat het om de mate van zekerheid dat het effect een aanbeveling ondersteunt terwijl dit bij een systematische review draait om de mate van zekerheid dat de effectgrootte het echte effect weerspiegelt. Zie hieronder voor de definities.

For guideline panels: The quality of evidence reflects the extent to which our **confidence in an estimate of the effect** is adequate to **support a particular recommendation**. **For authors of systematic reviews: The quality of evidence** reflects the extent to which we are **confident that an estimate of the effect is correct**. Bron: (1)

De zekerheid van het bewijs wordt geëvalueerd per uitkomstmaat, en dus voor alle studies samen die bijdragen aan een effectgrootte voor die uitkomstmaat. In eenzelfde systematische review kan de zekerheid van het bewijs verschillen tussen de uitkomstmaten. Zo zal bij niet-geblindeerde studies de zekerheid van het bewijs van een subjectieve uitkomstmaat zoals "pijn" lager worden beoordeeld dan een objectieve uitkomstmaat zoals "mortaliteit".

De beoordeling van de zekerheid van het bewijs wordt uitgedrukt in een van volgende 4 niveaus: hoog, matig, laag of zeer laag.

Om de zekerheid van bewijs te bepalen, begint de GRADE-methode bij het studiedesign. Bestaat de systematische review uit RCTs, dan start je met een hoog niveau van zekerheid van bewijs. Een systematische review met observationele studies start met een laag niveau van zekerheid van bewijs. Vervolgens zijn er een aantal factoren die het vertrouwen in de zekerheid van het bewijs kunnen verhogen ("upgraders") of verlagen ("downgraders). Er bestaan 5 downgraders en 3 upgraders. Iedere factor kan het niveau van zekerheid 1 of soms 2 niveaus doen dalen of stijgen (zie Tabel 4).

Tabel 4. Factoren die de zekerheid van het bewijs kunnen verlagen of verhogen

Studiedesign	Factoren	Oordeel
RCT	Problemen in studieopzet Inconsistentie Imprecisie Indirectheid Risico op publicatie bias	Ernstig (-1) of zeer ernstig (-2) Ernstig (-1) of zeer ernstig (-2) Ernstig (-1) of zeer ernstig (-2) Ernstig (-1) of zeer ernstig (-2) Ernstig (-1)
Observationeel	Groot effect Dosis-responsgradiënt Restconfounding of “plausible confounding”	Groot (+1) of zeer groot (+2) Aanwezig (+1) Waarschijnlijk aanwezig (+1)

Algemene instructies:

GRADE is geen kwantitatief systeem om de zekerheid van het bewijs te beoordelen, dat betekent dat de beoordeling niet zwart-wit is. En dit maakt het ook moeilijk. Probeer jezelf voor te stellen dat alle factoren worden beoordeeld op een continue schaal. Dus de studies voor de bepaalde uitkomst, bijvoorbeeld ‘mortaliteit’ hebben problemen in de studieopzet. Dit kan gaan van kleine problemen die slechts gering zullen doorwegen op de effectgrootte van mortaliteit tot zeer grote problemen waarbij de effectgrootte van de review heel waarschijnlijk niet accuraat zijn. Het gaat erom te bepalen wanneer is het probleem groot genoeg zodat ik minder vertrouwen heb in het eindresultaat. Dit is een subjectieve beslissing. De kracht van de GRADE-methode is om de argumenten voor dergelijke beslissingen te rapporteren waardoor het oordeel transparant wordt. Bijvoorbeeld “we verwachten dat het gebrek aan blindering in de studies de effectgrootte van mortaliteit niet belangrijk heeft beïnvloed omdat dit een objectieve uitkomstmaat is. Daarom is niet gedowngraded voor problemen in de studie-opzet.”

De GRADE-methode hanteert een conservatief beoordelingssysteem. Je moet dus zeker zijn dat de factor die problemen geeft, de resultaten van de effectschatting heeft beïnvloed. Als er een klein probleem is m.b.t. 2 factoren die elk niet zwaar genoeg doorwegen om de zekerheid van het bewijs samen met twee niveaus te verlagen, kun je ervoor kiezen om ze samen te nemen en ze met 1 niveau te verlagen. Ook hierbij is het belangrijk om de argumenten hiervoor goed te documenteren.

Factoren die de zekerheid van bewijs verlagen (“downgraders”)

Deze factoren zijn van toepassing op gerandomiseerde, niet-gerandomiseerde studies en observationeel onderzoek en kunnen leiden tot een verlaging van het niveau van zekerheid van het bewijs. In het geval van gerandomiseerde studies als basis voor een richtlijn, start je met een hoog niveau van zekerheid van bewijs dat kan worden verlaagd tot ‘matig’, ‘laag’ of ‘zeer laag’ als één, twee of drie downgraders van toepassing zijn. Lager dan ‘zeer laag’ is niet mogelijk, zelfs als er meer dan drie downgraders van toepassing zijn. Start je met niet-gerandomiseerde en observationele studies als basis voor een richtlijn, begin je met een laag niveau van zekerheid dat wordt verlaagd tot zeer laag als er een downgrader is. De zekerheid van bewijs van observationele studies zonder een controlegroep wordt beoordeeld als ‘zeer laag’.

2.1 Problemen in de studieopzet

Problemen in de studieopzet kunnen de resultaten van deze studies vertekenen. In dat geval heb je minder vertrouwen in de resultaten ervan. De GRADE-methode stelt zich dus steeds de vraag of men

door eventuele problemen in de studieopzet minder vertrouwen heeft in de resultaten van de individuele studies en dus in de effectgrootte berekend in de systematische review.

Voorbeelden van problemen in de studieopzet zijn: geen of slechte randomisatie, niet-geblindeerde toewijzing van de deelnemers aan de interventie- of controlegroep, gebrek aan blindering (van deelnemers, personeel of uitkomstbeoordelaars), selectieve uitval van deelnemers. Er is geen vaststaande lijst van problemen in de studieopzet. GRADE adviseert alle problemen te bekijken die een effect kunnen hebben op de effectgrootte.

Hoe beoordeel je dit?

Een systematische review onderzoekt en beschrijft de kwaliteit van alle geïncludeerde studies aan de hand van een instrument zoals de “Cochrane Risk of Bias Tool”. Een GRADE-beoordeling gebeurt per uitkomstmaat. Je beoordeelt alleen opzet van de studies die bijdragen aan de effectgrootte van die uitkomstmaat. Als in een systematische review, die 20 studies bevat, slechts 4 studies een bepaalde uitkomstmaat meten, bekijk je dus alleen de studieopzet van deze 4 studies om de zekerheid van het bewijs voor deze uitkomstmaat in te schalen.

Als het meeste bewijs komt van studies met een laag risico op bias, dan kun je concluderen dat er geen belangrijk probleem is in dit gebied. Dit betekent niet dat alle studies perfect zijn uitgevoerd en beschreven, maar wel dat men er algemeen van kan uitgaan dat er een laag risico op bias is. Als sommige studies kritische problemen hebben in hun studieopzet die doorwegen in de resultaten, kun je besluiten om met 1 niveau te verlagen, bijvoorbeeld bij het ontbreken van blindering voor een subjectieve uitkomstmaat zoals levenskwaliteit.

Het is niet de bedoeling om het aantal studies met problemen in hun studieopzet te tellen. Het is beter om de relatieve bijdrage van elke probleem in de studieopzet per uitkomst in te schalen. Studies die zwaarder doorwegen in de meta-analyse, wegen het best ook zwaarder door bij het maken van dit algemene oordeel.

Tot slot, wees conservatief bij deze beoordeling. Om de kwaliteit van het bewijs te verlagen, moet je zeker zijn dat er een belangrijk risico op vertekening is over alle studies heen die bijdragen aan de effectgrootte. Je verlaagt het niveau met -1 als je van mening bent dat er een ernstig risico op vertekening is en -2 bij een zeer ernstig risico op vertekening door problemen in de studieopzet.

Voorbeeld

Een effectgrootte van een terugkerende trombose tijdens een follow-up van 12 maanden is berekend op basis van 5 RCTs met 1781 deelnemers. Twee studies zijn onduidelijk over allocation concealment, 1 studie (met ruim 800 deelnemers) heeft een hoog risico op bias door incomplete outcome data en alle studies hebben een gebrek aan blindering van patiënten en personeel. Hier is met 1 niveau gedowngraded vanwege een ernstig risico op vertekening door problemen in de studieopzet. (Bron: Kahale et al, 2018)

2.2 Inconsistentie

Voor deze factor zoeken we naar grote verschillen tussen de resultaten van de verschillende studies die bijdragen aan de effectschatting van de uitkomstmaat, de zogenaamde inconsistentie of heterogeniteit. De resultaten tussen studies kunnen verschillen omdat ze zijn uitgevoerd in verschillende patiëntengroepen of omdat de interventie niet helemaal gelijk is (verschil in dosis, duur, frequentie of wijze van uitvoering). Dit zijn verklaarbare verschillen en worden binnen de GRADE-beoordeling niet als onrustwekkend beschouwd. Bij de GRADE-beoordeling kijk je naar onverklaarbare variaties. Onverklaarbare verschillen verlagen het vertrouwen in de gepoolde effectschatting van de uitkomstmaat.

Hoe beoordeel je dit?

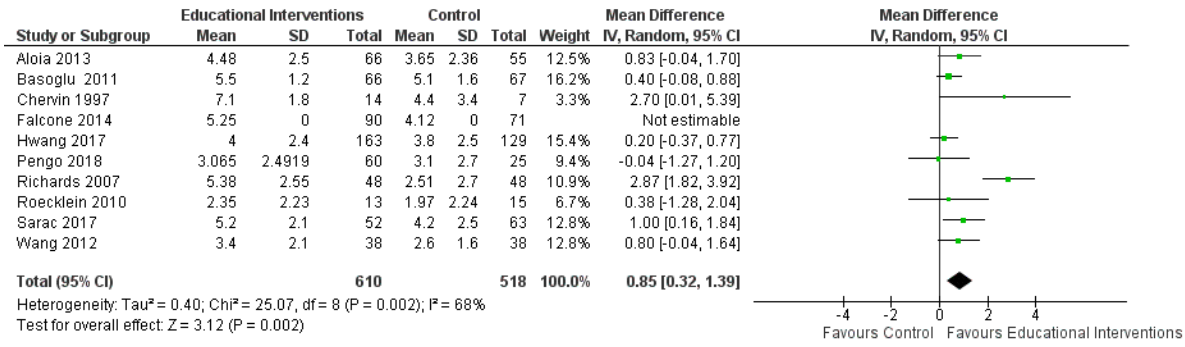
De belangrijkste manier om de resultaten van de individuele studies te bekijken in via de “forest plot”, een figuur met de resultaten van de individuele studies en de gepoolde effectgrootte.

Bekijk vooral eerst of de betrouwbaarheidsintervallen (BI) overlappen. Is dit het geval, dan lijken de resultaten consistent. Dit kun je bevestigen door de resultaten van I^2 te bekijken, een maat die de mate van heterogeniteit weergeeft. Hiervoor geeft GRADE de volgende indeling

- < 40% lage heterogeniteit
- 30-60% matige heterogeniteit
- 50-90% substantiële heterogeniteit
- 75-100% aanzienlijke heterogeniteit

De aanwezigheid van inconsistentie heeft niets te maken met aan welke zijde van nullijn van de plot de resultaten liggen. Resultaten tussen studies kunnen inconsistent zijn, ook al laten ze allemaal positieve effecten zien. Daarnaast kun je ook de resultaten van andere statistische tests (zoals Chi^2, tau^2) gebruiken om te bepalen of de resultaten heterogeen zijn en dus inconsistent. Bedenk wel dat deze tests hun beperkingen hebben en niet als enige argument gebruikt kunnen worden om te besluiten tot inconsistentie.

Figuur 1. Voorbeeld van inconsistentie van resultaten van de systematische review over het effect van educatie interventies voor het gebruik van luchtdrukmachines bij mensen met slaap apneu. Deze forest plot betreft het effect op gebruik van de machine (uren/nacht) (Bron: Askland et al, 2020)



“2 There was minimal or no variability in direction of elect, with all (or nearly all) studies favouring the intervention arm. Magnitude of effect varied substantially (4 studies with CIs excluding null). CIs have reasonable overlap. Substantial statistical heterogeneity P = 0.002, I^2 = 66%. Therefore, inconsistency was downgraded by one level to 'serious.'”

2.3 Imprecisie

De factor “imprecisie” heeft betrekking op de mate van precisie van de effectgrootte. Als de resultaten weinig precies of nauwkeurig zijn, verlaagt dit onze zekerheid erin.

De eerste vraag die we ons moeten stellen is of de studies ons genoeg informatie geven om na te gaan of de resultaten precies zijn. Hiervoor bekijken we het totale aantal patiënten en het aantal events (gebeurtenissen) die bijdragen aan een bepaalde uitkomstmaat. Voor het totale aantal patiënten, baseren we ons op de standaard steekproefgrootte voor een enkele studie met voldoende power. Dit noemen we de “Optimal Information Size” (OIS). Als we niet minstens dat aantal patiënten in de meta-analyse hebben, dan besluiten we dat de effectschatting imprecies is. Ook als er heel weinig events zijn, bijvoorbeeld het aantal patiënten met een ernstig neveneffect, zijn gevonden in de studies die die uitkomst hebben geëvalueerd, concluderen we dat het bewijs imprecies is.

Ten tweede moeten we de breedte van het betrouwbaarheidsinterval bekijken. Als het betrouwbaarheidsinterval wijst op twee mogelijke conclusies, dan kan men ook concluderen dat de effectschatter imprecies is. Een voorbeeld hiervan is een betrouwbaarheidsinterval voor een relatief risico op neveneffecten van 0.95 tot 2.50. Dit suggereert dat het risico op neveneffecten nauwelijks verhoogd kan zijn maar dat het ook kan zijn dat dit risico met 150% is verhoogd in de behandelgroep.

<u>Hoe beoordeel je dit?</u>

Er bestaan vuistregels voor een minimale steekproefgrootte en het minimale aantallen events in een meta-analyse, maar deze moeten met voorzichtigheid worden gebruikt omdat ze mogelijk niet geschikt zijn voor de uitkomst waar het om gaat. Als een analyse minder dan 400 events heeft (voor dichotome uitkomstmaten, ofwel uitkomstmaten met 2 mogelijkheden, zoals wel of niet overlijden) of 400 personen (voor continue uitkomstmaten, ofwel uitkomstmaten met een range van mogelijkheden, zoals gewicht in kilogram) voor de interventie- en controlegroepen samen, dan kan men overwegen om het niveau van zekerheid te verlagen, tenzij de minimale OIS is bereikt.

Om de breedte van het betrouwbaarheidsinterval te kunnen beoordelen, bepaalt de richtlijnontwikkelingsgroep voor de uitkomst een drempelwaarde, die aangeeft wanneer een bepaald effect belangrijk is. Een voorbeeld is een verschil in pijnscore van 2 op een schaal van 0 tot 10. Als een betrouwbaarheidsinterval voorbij de 2 gaat en ook nul omvat, zoals een gemiddeld verschil van 1.3 (95%BI -0.5 tot 3.1), zijn de resultaten imprecies omdat de conclusie kan zijn ‘de interventie is effectief’, maar ook ‘de interventie heeft geen effect’. Voor richtlijnen is het belangrijk om vooraf met de richtlijnontwikkelingsgroep te bepalen waar de grens ligt van een belangrijk effect.

Als vuistregel om te bepalen of een effect van een dichotome uitkomstmaat belangrijk is, wordt een toename met 25% gebruikt. Dit komt overeen met een RR groter dan 1.25 of kleiner dan 0.75.

Bij ernstige imprecisie, bijvoorbeeld bij zowel weinig events als een betrouwbaarheidsinterval dat meerdere mogelijke conclusies omvat, kan men besluiten met 2 niveaus af te waarderen.

Figuur 2. Voorbeeld van imprecisie van resultaten van de systematische review over het effect van anticoagulantia voor de behandeling van veneuze trombose bij mensen met kanker. Dit voorbeeld betreft het effect op algehele mortaliteit. (Bron: Kahale et al, 2018)

Summary of findings 2. Direct oral anticoagulant secondary prophylaxis compared to vitamin K antagonist secondary prophylaxis in people with active cancer with venous thromboembolism

DOAC secondary prophylaxis compared to VKA secondary prophylaxis in people with active cancer with VTE

Population: people with cancer with VTE receiving secondary prophylaxis

Setting: outpatient

Intervention: DOAC prophylaxis

Control: VKA prophylaxis

Outcomes	Nº of participants (studies)	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with VKA secondary prophylaxis	Risk difference with DOAC secondary prophylaxis
All-cause mortality up to 12 months	1031 (4 RCTs)	⊕⊕⊝⊝ **Low**[a,b]	**RR 0.93** (0.71 to 1.21)	Study population 176 per 1000	12 fewer per 1000 (51 fewer to 37 more)

"b Downgraded one level due to serious imprecision, 95% CI was consistent with the possibility of important benefit (51 fewer per 1000) and possibility of important harm (37 more per 1000); included 174 events."

2.4 Indirectheid

Direct bewijs bestaat uit onderzoek naar de juiste interventies in de juiste populatie met uitkomstmaten die relevant zijn voor de patiënten. Met de juiste interventies en populatie bedoelen we de interventies en patiëntengroep waarover de richtlijn gaat.

Er zijn verschillende bronnen van indirect bewijs, zoals de patiëntpopulatie, de experimentele interventie, de controle-interventie en de uitkomstmaten. Een andere bron van indirect bewijs zijn analyses met indirecte vergelijkingen. Wanneer, voor een richtlijn, de vergelijking tussen twee medicijnen gevraagd wordt, kan het zijn dat er enkel studies beschikbaar zijn die beide medicijnen vergelijken met placebo, maar niet met elkaar. Via netwerkanalyse kan je beide studies in één analyse combineren maar hier is sprake van indirect bewijs omdat je nooit volledig zeker kan zijn dat beide placebo-interventies vergelijkbaar zijn.

<u>Hoe beoordeel je dit?</u>

Om directheid te beoordelen, bekijken we de klinische vraag van de richtlijn, de vraag die we wilden beantwoorden, en gaan we na of het gevonden bewijs geschikt is om deze vraag te beantwoorden. Hiervoor kan men bijvoorbeeld de inclusiecriteria van de richtlijn vergelijken met de gegevens in de tabel met de studiekenmerken van geïncludeerde studies.

Het kan zijn dat studies enkel een beperkt aspect van de klinische vraag hebben bekeken, bv. studies die speciale populaties onderzoeken zoals personen met een hoog risico op co-morbiditeiten. Bepaal dan of er redenen zijn waarom dit bewijs niet zou gelden voor andere patiënten. Zijn er individuele biologische verschillen of verschillen in setting, zodanig dat we bij andere patiënten een ander effect mogen verwachten? Andere voorbeelden van indirect bewijs zijn studies die een specifieke interventie onderzoeken, geïmplementeerd door grondig opgeleide experts, studies die een interventie vergelijken met een controle-interventie die minder doeltreffend is dan de huidige standaardzorg. Ook het meten van surrogaatuitkomsten (uitkomstmaten die niet direct klinisch relevant zijn, zoals de botmineraaldichtheid voor mensen met osteoporose) is een vorm van indirect bewijs.

Bij verregaande indirectheid, bijvoorbeeld als er alleen dierexperimentele studies zijn over de werkzaamheid van een interventie, kan je besluiten het bewijs met 2 niveaus te verlagen.

Figuur 3. Voorbeeld van indirectheid van resultaten van de systematische review over het effect van anticoagulantia voor de behandeling van veneuze trombose bij mensen met kanker. Dit voorbeeld betreft het effect op algehele mortaliteit. (Bron: Kahale et al, 2018)

Summary of findings 2. Direct oral anticoagulant secondary prophylaxis compared to vitamin K antagonist secondary prophylaxis in people with active cancer with venous thromboembolism

DOAC secondary prophylaxis compared to VKA secondary prophylaxis in people with active cancer with VTE

Population: people with cancer with VTE receiving secondary prophylaxis

Setting: outpatient

Intervention: DOAC prophylaxis

Control: VKA prophylaxis

Outcomes	№ of participants (studies)	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with VKA secondary prophylaxis	Risk difference with DOAC secondary prophylaxis
All-cause mortality up to 12 months	1031 (4 RCTs)	⊕⊕⊖⊖ **Low**[a,b]	**RR 0.93** (0.71 to 1.21)	Study population 176 per 1000	12 fewer per 1000 (51 fewer to 37 more)

"A Downgraded one level due to serious indirectness. Two studies (RECOVER I-II and RE-MEDY) included people with a diagnosis of cancer within five years before enrolment"

2.5 Publicatiebias

Hierbij stellen we ons de vraag: "hebben we alle studies gevonden over onze klinische vraag, of is het mogelijk dat we studies missen met onduidelijk of negatieve resultaten die nooit zijn gepubliceerd?"

We kunnen nooit zeker zijn of er al dan geen publicatiebias is. Het beste wat schrijvers van een systematische review kunnen doen is publicatiebias zoveel mogelijk voorkomen en hun best doen alle studies te vinden door het uitvoeren van een omvattende zoekactie, pogingen doen om niet-gepubliceerde en grijze literatuur te vinden, auteurs in het veld proberen te contacteren, etc.

Trialregisters zijn een belangrijke bron van informatie hierbij en nemen in belang toe. Steeds meer tijdschriften eisen een registratie van een studie voordat ze die willen publiceren. Registers hebben het potentieel om een belangrijk verschil te maken in het verminderen van het risico op publicatiebias. Maar op dit moment zijn er nog beperkingen aangaande de volledigheid van de gegevens die geregistreerd worden en de toepassing ervan door de tijdschriften.

Hoe beoordeel je dit?

Om dit te beoordelen bekijk je allereerst de zoekstrategie van de systematische review. Is deze omvattend genoeg of is er reden om aan te nemen dat relevante studies zijn gemist?

In een meta-analyse kunnen we een visuele inspectie en tests voor asymmetrie van "funnel plots" doen. Funnel plots zijn figuren waarbij de effectgroottes van een studieresultaat zijn uitgezet tegen de steekproefgrootte. Studieresultaten in de funnel plot liggen verspreid rond de gepoolde effectschatting. Als er geen teken is van publicatiebias, dan is het resultaat een symmetrische figuur in de vorm van een omgekeerde driehoek. Als de figuur asymmetrisch is, geeft dit een vermoeden op publicatiebias. Funnel plots zijn alleen waardevol als er minstens 10 studies in zijn opgenomen.

Systematische review kunnen ook statische tests uitvoeren voor publicatiebias (bijvoorbeeld "trim-and-fill"-tests), maar deze zijn foutgevoelig en de resultaten moeten daarom voorzichtig worden geïnterpreteerd. Het is heel moeilijk om zeker te zijn dat er geen publicatiebias is en het is onmogelijk om een afkappunt aan te geven voor het verlagen van het niveau. GRADE suggereert om hiervoor met maximaal 1 niveau te verlagen.

Bepaalde factoren kunnen publicatiebias doen vermoeden, bijvoorbeeld: als de set studies vooral bestaat uit kleinschalige studies met positieve resultaten, en wanneer sponsoring door de industrie een rol speelt. Maar ook wanneer veel studies de belangrijkste uitkomst niet schijnen te rapporteren. Wanneer studies gesponsord zijn door de industrie, maar met resultaten die niet in het voordeel pleiten van het product, zou dit een argument kunnen zijn tegen een risico op publicatiebias.

Figuur 4. Voorbeeld van mogelijke publicatie bias van resultaten van de systematische review over het effect van educatie interventies voor het gebruik van luchtdrukmachines bij mensen met slaap apneu. Dit voorbeeld betreft het effect op levenskwaliteit () (Bron: Askland et al, 2020)

2.7 Quality of life: Comparison of Change from Baseline Values	The mean Quality of life: Comparison of Change from Baseline Values was 0	SMD 0.22 higher (0.01 lower to 0.45 higher)	-	294 (3 RCTs)	⊕⊝⊝⊝ VERY LOW [3 9 10 11]

"11 Our review included a comprehensive search for published reports conducted. All (or nearly all) studies, including all small studies, for this comparison found a benefit for the intervention. Thus, due to suspicion for publication bias, this outcome was downgraded by one level."

Factoren die de zekerheid van bewijs verhogen ("upgraders")

Deze factoren zijn enkel van toepassing op observationeel onderzoek. Resultaten van observationeel onderzoek worden standaard gelabeld met een niveau van lage zekerheid. De reden is omdat er in dit type studies altijd factoren spelen die zorgen dat de interventie en controlegroepen niet helemaal gelijk zijn. Deze factoren kunnen de uitkomstmaten beïnvloeden. Dit effect is niet te corrigeren, zelfs niet met aangepaste analyses. We noemen dit restconfounding.

In bepaalde omstandigheden kan het niveau van zekerheid van bewijs van observationele studies worden verhoogd. Dat kan alleen als ze niet eerder werd verlaagd, m.a.w. als er geen sprake is van problemen met de studieopzet, inconsistentie, imprecisie, indirectheid en er geen risico is op publicatiebias.

Groot effect

De zekerheid van het bewijs van observationele studies kan worden verhoogd als het effect van een interventie groot is. De achterliggende gedachte is: hoe groter het effect van een interventie, hoe sterker het bewijs voor een associatie.

De effectgrootte wordt meestal uitgedrukt in een relatief risico (RR). Een groot effect komt overeen met een RR > 2 of een RR < 0,5, en dit bewijs kan men met 1 niveau verhogen. Bij een zeer groot effect met een RR > 5 of een RR < 0,2 kan men met twee niveaus verhogen. De analyses moeten wel correct zijn uitgevoerd, d.w.z. dat men heeft gecorrigeerd voor kenmerken die het gevonden effect zouden kunnen verstoren (confounders).

Dosis-responsgradiënt

De dosis-responsgradiënt geeft een relatie aan tussen de blootstelling en het effect; hoe meer blootstelling hoe groter het effect of andersom. De aanwezigheid van een dosis-responsgradiënt verhoogt ons vertrouwen in het effect en kan leiden tot het verhogen van de zekerheid van bewijs.

Als een systematische review een dosis responsgradiënt heeft onderzocht, staat dit in de methode beschreven. Men kan bijvoorbeeld de effectgrootte apart berekenen voor verschillende doseringen van de blootstelling, of een verband berekenen tussen de blootstelling en effectgrootte.

Restconfounding

Hierboven beschreven we al dat restconfounding een belangrijke reden is om zekerheid van bewijs van observationele studies in principe als ‘laag’ te waarderen.

In sommige studies kunnen de restconfounders een tegengesteld effect laten vermoeden dan het effect aangetoond in de studie. Bijvoorbeeld, wanneer in een studie er een vermoeden is dat enkel ziekere patiënten een behandeling kregen, maar zij toch beter werden dan de gezondere patiënten die de behandeling niet kregen, dan is het plausibel dat het werkelijke effect van een interventie nog groter zal zijn dan het effect gevonden in de studies. Dit verhoogt het vertrouwen in de gevonden resultaten en kan reden zijn om de kwaliteit van het bewijs te verhogen.

Globaal oordeel van zekerheid van het bewijs

Op basis van de beoordeling van voorgaande factoren per uitkomstmaat wordt een globaal niveau van zekerheid van het bewijs bepaald. De globale zekerheid van bewijs is het resultaat van een gecombineerde beoordeling van het bewijs voor alle uitkomsten die kritisch zijn voor het beantwoorden van een klinische vraag. Deze stap is daarom alleen relevant voor

richtlijnontwikkelaars, bij het formuleren van een aanbeveling, en niet voor auteurs van een systematische review.

Wanneer we het overzicht bekijken van het GRADE-profiel is het waarschijnlijk dat het niveau van zekerheid van het bewijs verschilt tussen de uitkomstmaten. Om een globaal oordeel van zekerheid van bewijs te bepalen, neem je allereerst die uitkomstmaten die kritiek zijn voor het nemen van een beslissing. Het laagste niveau van zekerheid voor elk van de kritieke uitkomstmaten bepaalt het globale oordeel van zekerheid.

3. Komen van bewijs tot een aanbeveling

Een klinische aanbeveling is ofwel voor of tegen een bepaalde behandeling en kan sterk of zwak zijn. Er zijn bijgevolg vier mogelijkheden bij het formuleren van de aanbeveling. Elk van deze mogelijkheden heeft een eigen implicatie voor patiënten, zorgverleners en beleidsmakers en dit helpt richtlijnontwikkelingsgroepen om de formulering van een aanbeveling zoveel mogelijk uniform maken.

Tabel 5. Mogelijke sterktes van aanbeveling volgens de GRADE-methode

Sterk voor	Zwak voor	Zwak tegen	Sterk tegen
Beoordeling op een continue schaal			

De sterkte van een aanbeveling reflecteert de mate waarin een richtlijnontwikkelingsgroep zeker is dat de effecten van een interventie (voordelen) zwaarder wegen dan de risico's (nadelen) of andersom, voor de patiënten voor wie de aanbeveling is bedoeld. Om deze weging transparant te maken wordt het Evidence to Decision (EtD) framework aangeraden (zie hieronder).

Een sterke aanbeveling houdt in dat de richtlijnontwikkelingsgroep van mening is dat deze interventie voor de meeste patiënten de beste keuze is omdat er meer voordelen zijn dan nadelen. Dus de meeste of alle patiënten zouden gebaat zijn bij deze behandeling. Bij een sterke aanbeveling tegen een interventie betekent dit dat de richtlijnontwikkelingsgroep van mening is dat de nadelen zwaarder wegen dan de voordelen en dat de meeste, of alle patiënten, gebaat zijn deze interventie niet te krijgen.

Een zwakke aanbeveling houdt in dat de richtlijnontwikkelingsgroep van mening is dat de balans tussen voor- en nadelen eerder in evenwicht is en dat het merendeel van de patiënten, maar niet allemaal, gebaat zijn bij de interventie. Bij een zwakke aanbeveling tegen een interventie wegen de nadelen zwaarder dan de voordelen en is de richtlijnontwikkelingsgroep van mening dat de meeste, of alle patiënten gebaat zijn deze interventie niet te krijgen.

	Sterke aanbeveling voor	Zwakke aanbeveling voor
Patiënten	De meeste patiënten willen de interventie, een kleine groep wil dit niet.	De meerderheid van de patiënten wil de interventie, maat veel willen deze niet.
Clinici	De meeste personen zouden de interventie moeten krijgen. Het opvolgen van de interventie kan als kwaliteitsindicator worden gebruikt. Formele beslissingshulpen zijn waarschijnlijk niet nodig	Verschillende keuzes zijn passend voor verschillende patiënten. Je moet een patiënt helpen om te komen tot een keuze die bij hem/haar past. Beslishulpen kunnen patiënten hierbij helpen
Beleidsmakers	De aanbeveling kan worden aangepast naar het beleid in de meeste situaties voor gebruik als performance indicatoren	Beleid maken heeft substantiële debatten en betrekken van verschillende stakeholders. Het beleid varieert meer waarschijnlijk tussen regio's. Performance indicatoren zouden zich moeten richten op het feit dat voldoende discussie is geweest over de behandelopties.

Tabel 6. Implicaties van sterkte van aanbeveling voor een interventie (bron: 1)

EtD framework

Het GRADE Evidence to Decision (EtD) framewerk beschrijft hoe de voordelen en de mogelijke nadelen tegen elkaar worden afgewogen, rekening houdend met de zekerheid die we hebben in de effectgrootte van zowel de voor- als nadelen. Het neemt ook nog andere factoren in overweging:

- De balans tussen effecten en risico's ofwel voor- en nadelen
    - Hoe groter het verschil in balans tussen de voor- en nadelen, hoe waarschijnlijker een sterke aanbeveling wordt. Bijvoorbeeld:
        - Zijn de voordelen veel groter dan de nadelen (sterk voor)
        - Zijn de nadelen groot en het effect klein (sterk tegen)
    - Hoe kleiner het netto voordeel en een lagere zekerheid voor dat voordeel, hoe waarschijnlijker een zwakke aanbeveling wordt.
        - Zijn de voordelen iets groter dan de nadelen (zwak voor)
        - Zijn de nadelen iets groter dan de voordelen (zwak tegen)
- Het vertrouwen in de effectgrootte
    - Hoe hoger de zekerheid van bewijs, hoe waarschijnlijker een sterke aanbeveling
- Vertrouwen in waarden en voorkeuren van patiënten en de variatie hierin
    - Hoe meer variatie in de waarden en voorkeuren van patiënten, of bij onzekerheid hierover, hoe waarschijnlijker het is om een zwakke aanbeveling te formuleren.

Om deze factoren transparant te beoordelen, zouden richtlijnontwikkelingsgroepen een framework met vragen moeten volgen (zie figuur). Per item komt de richtlijnontwikkelingsgroep tot een beslissing, beschrijft waarop die beslissing is gebaseerd en beschrijft in welke mate dit item meespeelt bij het formuleren van de aanbeveling. Dit maakt de stap van wetenschappelijk bewijs naar een aanbeveling transparant.

Hoe beoordeel je dit?

De vragen in het EtD framework zijn:

- Wat is de globale zekerheid van het bewijs?
- Is er een belangrijke onzekerheid over hoe patiënten de belangrijkste uitkomstmaten waarderen?
- Zijn de te verwachten effecten groot?
- Zijn de te verwachten neveneffecten klein?
- Zijn de voordelen groter dan de neveneffecten?
- Zijn de extra kosten lager dan de te verwachten opbrengsten?

Tabel 7. Evidence to decision framework.

Criteria	**Beoordeling**	**Gebaseerd op (bewijs)**	**Overige overwegingen**
Voor- en nadelen			
Wat is de globale zekerheid van het bewijs?	0 Hoog 0 Matig 0 Laag 0 Zeer laag		
Is er belangrijke onzekerheid over hoe mensen de belangrijkste uitkomsten beoordelen? Deze onzekerheid of variatie is	0 Belangrijk 0 Mogelijk belangrijk 0 Mogelijk niet belangrijk 0 Niet belangrijk 0 Onbekend		
Zijn de te verwachten effecten groot?	0 Nee 0 Waarschijnlijk niet 0 Onzeker 0 Waarschijnlijk wel 0 Ja 0 Het varieert		
Zijn de te verwachten neveneffecten klein?	0 Nee 0 Waarschijnlijk niet 0 Onzeker 0 Waarschijnlijk wel 0 Ja 0 Het varieert		
Zijn de gewenste effecten groter dan de ongewenste effecten?	0 Nee 0 Waarschijnlijk niet 0 Onzeker 0 Waarschijnlijk wel 0 Ja 0 Het varieert		

Kosten			
Zijn de benodigde kosten beperkt?	0 Nee 0 Waarschijnlijk niet 0 Onzeker 0 Waarschijnlijk wel 0 Ja 0 Het varieert		
Zijn de extra kosten klein tegenover de netto voordelen?	0 Nee 0 Waarschijnlijk niet 0 Onzeker 0 Waarschijnlijk wel 0 Ja 0 Het varieert		
Gelijkheid, aanvaardbaarheid en haalbaarheid			
Wat is het effect op (on)gelijkheid in de zorg? De zorgongelijkheid	0 Neemt toe 0 Neemt waarschijnlijk toe 0 Is onzeker 0 Neemt waarschijnlijk af 0 Neemt af 0 Het varieert		
Is de behandeling aanvaardbaar voor de belangrijkste stakeholders?	0 Nee 0 Waarschijnlijk niet 0 Onzeker 0 Waarschijnlijk wel 0 Ja 0 Het varieert		
Is de behandeling haalbaar om te implementeren?	0 Nee 0 Waarschijnlijk niet 0 Onzeker 0 Waarschijnlijk wel 0 Ja 0 Het varieert		

Checklist om te bepalen of GRADE is toegepast (bron: 1)

Criteria:

1. De definitie van kwaliteit van het wetenschappelijk bewijs (zekerheid in de geschatte effecten) is consistent met dat van de GRADE Werkgroep.
2. Er wordt expliciet aandacht besteed aan de 8 factoren (zie tabel 4) die de zekerheid van het bewijs kunnen verlagen of verhogen.
3. De kwaliteit van bewijs is beoordeeld voor elke belangrijke uitkomstmaat en weergegeven als 'hoog, matig, laag of zeer laag', gebaseerd op de 8 factoren (zie punt 2), waarbij de interpretatie consistent is met dat van de GRADE Werkgroep. Alternatieve bewoordingen voor deze factoren zijn mogelijk.
4. De argumenten over de factoren (zie punt 2) als basis voor de beoordeling van de zekerheid van bewijs of sterkte van aanbeveling worden transparant gerapporteerd in evidence tabellen of narratieve samenvattingen van het wetenschappelijk bewijs. In het ideale geval worden volledige GRADE-profielen gebruikt, gebaseerd op systematische reviews. Minimaal is het duidelijk beschrijven van de methoden om wetenschappelijk bewijs te identificeren en kritisch te beoordelen, inclusief de redenen om de zekerheid van bewijs te verlagen of te verhogen.
5. Er wordt expliciet aandacht besteed aan de vier criteria voor het bepalen van de sterkte van aanbeveling (balans van gewenste en ongewenste effecten, zekerheid van bewijs, waarden en voorkeuren van diegene waarom het gaat en gebruik van middelen), en hoe deze informatie is verkregen.
6. De terminologie voor sterkte van aanbeveling wordt beschreven in 2 categorieën (sterk of zwak) en de definities/interpretatie is consistent met dat van de GRADE Werkgroep. Alternatieve bewoordingen zijn mogelijk (bv conditionele aanbeveling als alternatief voor zwakke aanbeveling) maar de interpretatie en implicaties moeten worden behouden.
7. In het ideale geval worden besluiten over de sterkte van de aanbeveling transparant gerapporteerd.

Literatuurlijst

Bronnen gebruikt voor deze handleiding

1. GRADE Handboek (https://gdt.gradepro.org/app/handbook/handbook.html)
2. Leopoldino AO, Machado GC, Ferreira PH, Pinheiro MB, Day R, McLachlan AJ, Hunter DJ, Ferreira ML. Paracetamol versus placebo for knee and hip osteoarthritis. Cochrane Database of Systematic Reviews 2019, Issue 2. Art. No.: CD013273. DOI: 10.1002/14651858.CD013273.
3. Kahale LA, Hakoum MB, Tsolakian IG, Alturki F, Matar CF, Terrenato I, Sperati F, Barba M, Yosuico VED, Schünemann H, Akl EA. Anticoagulation for the long-term treatment of venous thromboembolism in people with cancer. Cochrane Database of Systematic Reviews 2018, Issue 6. Art. No.: CD006650. DOI: 10.1002/14651858.CD006650.pub5.
4. Askland K, Wright L, Wozniak DR, Emmanuel T, Caston J, Smith I. Educational, supportive and behavioural interventions to improve usage of continuous positive airway pressure machines in adults with obstructive sleep apnoea. Cochrane Database of Systematic Reviews 2020, Issue 4. Art. No.: CD007736. DOI: 10.1002/14651858.CD007736.pub3.

GRADE-methoden voor andere onderzoeksdesigns/vragen

- Effect van interventies narratief samengenomen
    - Murad MH et al. Rating the certainty in evidence in the absence of a single estimate of effect. Evid Based Med June 2017; volume 22, number 3.
- Netwerk meta-analyse
    - Salanti G, et al. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014 Jul 3;9(7):e99682.
- diagnostische test accuracy studies
    - Schünemann H et al. GRADE guidelines 21. part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. JCE
    - Schünemann H et al. GRADE guidelines 22. part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. JCE Feb 2020.
- Prognostische studies
    - Iorio A et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ 2015;350:h870
- Patiëntvoorkeuren en waarden
    - Zhang Y et al. Assessing the certainty of evidence in the importance of outcomes or values and preferences—Risk of bias and indirectness. JCE Feb 2018.
    - Zhang Y et al. Assessing the certainty of evidence in the importance of outcomes or values and preferences—inconsistency, imprecision, and other domains. JCE May 2018.
- Economische studies
    - Brunetti M et al Considering resource use and rating the quality of economic evidence. JCE Aug 2012
- Kwalitatieve studies (GRADE-CERqual)

- Lewin S et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med 12, no. 10 (2015) 001895.
- Website: GRADE-CERqual: https://www.cerqual.org/

Voor geüpdatete informatie, raadpleeg de website:
https://www.gradeworkinggroup.org/

Meer informatie over GRADE

GRADE online learning module on imprecision (http://cebgrade.mcmaster.ca/Imprecision/index.html),

GRADE guidelines JCE serie (2011- : serie van artikelen over GRADE voor auteurs van richtlijnen en systematische reviews: https://www.jclinepi.com/content/jce-GRADE-Series)

1. Introduction—GRADE evidence profiles and summary of findings tables (Guyatt G et al JCE Jan 2011)

2. Framing the question and deciding on important outcomes (Gordon H. Guyatt GH et al. JCE Jan 2011)

3. Rating the quality of evidence (Howard Balshem H et al. JCE Jan 2011)

4. Rating the quality of evidence—study limitations (risk of bias) (Guyatt GH et al., JCE Jan 2011)

5. Rating the quality of evidence—publication bias (Guyatt GH et al., JCE Aug 2011)

6. Rating the quality of evidence—imprecision (Guyatt GH et al., JCE Aug 2011)

7. Rating the quality of evidence—inconsistency (Guyatt GH et al., JCE Aug 2011)

8. Rating the quality of evidence—indirectness (Guyatt GH et al., JCE Aug 2011)

9. Rating up the quality of evidence (Guyatt GH et al., JCE Aug 2011)

10. Considering resource use and rating the quality of economic evidence (Brunetti M et al., JCE Aug 2012)

11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes (Guyatt G et al., April 2012)

12. Preparing Summary of Findings tables—binary outcomes (Guyatt GH et al., JCE May 2012)

13. Preparing Summary of Findings tables and evidence profiles—continuous outcomes (Guyatt GH et al., JCE Oct 2012)

14. Going from evidence to recommendations: the significance and presentation of recommendations (Andrews JC et al., JCE April 2013)

15. Going from evidence to recommendation—determinants of a recommendation's direction and strength (Andrews JC et al., JCE April 2013)

16. GRADE evidence to decision frameworks for tests in clinical practice and public health (Schünemann HJ et al., JCE Feb 2016)

17. Assessing the risk of bias associated with missing participant outcome data in a body of evidence (Guyatt GH et al., JCE May 2017)

18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence (Schünemann HJ et al., JCE Feb 2018)

19. Assessing the certainty of evidence in the importance of outcomes or values and preferences—Risk of bias and indirectness (Zhang Y et al., JCE Feb 2018)

20. Assessing the certainty of evidence in the importance of outcomes or values and preferences—inconsistency, imprecision, and other domains (Zhang Y et al., JCE May 2018)

21. Part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy. Schünemann HJ et al., JCE Feb 2020.

21. Part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables. Schünemann HJ et al., JCE Feb 2020.

22. The GRADE approach for tests and strategies—from test accuracy to patient-important outcomes and recommendations (Schünemann HJ et al., JCE Feb 2019)

26. Informative statements to communicate the findings of systematic reviews of interventions (Santesso N et al., JCE Nov 2019)

27. How to calculate absolute effects for time-to-event outcomes in summary of findings tables and Evidence Profiles (Skoetz N et al., JCE Nov 2019)

GRADE BMJ serie (over EtD framework)

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction (Alonso-Coello P et al. BMJ 2016;353:i2016)

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines (Alonso_Coelle P et al., BMJ 2016;353:i2089)