

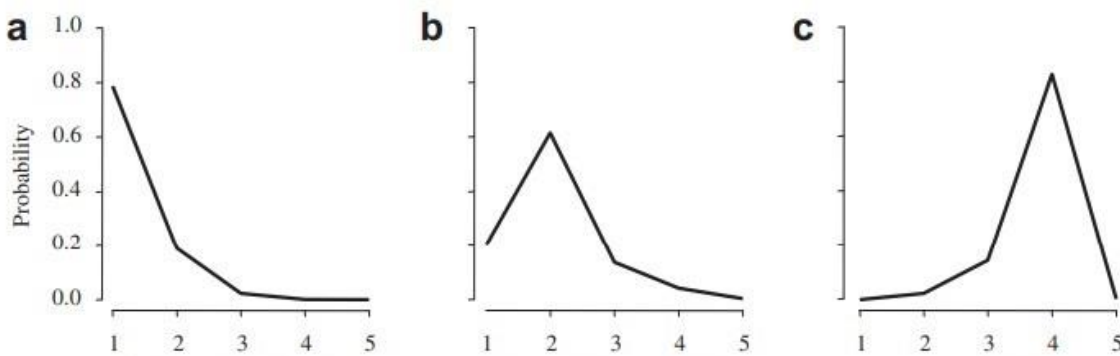


# Is de Surface Under the Cumulative Ranking-curve (SUCRA) een betrouwbare manier om de resultaten van een netwerk meta-analyse klinisch te interpreteren?

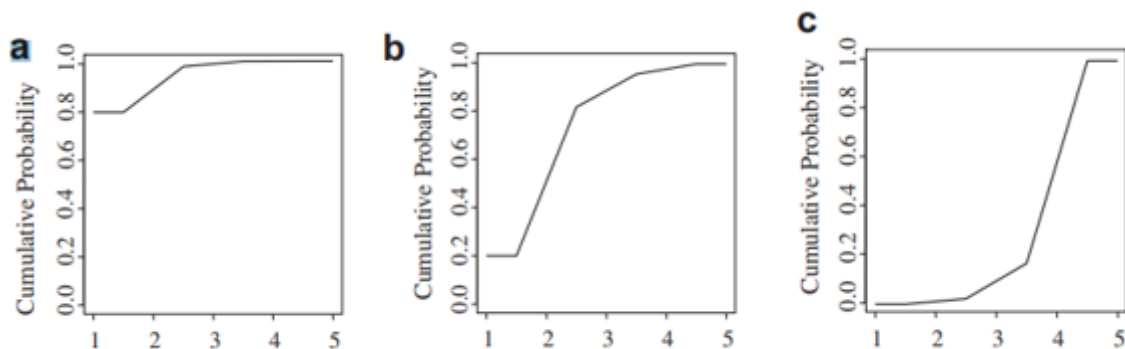
Tom Poelman, Vakgroep Volksgezondheid en Eerstelijnszorg, UGent

Een ‘klassieke’ meta-analyse combineert de resultaten van meerdere gerandomiseerde gecontroleerde studies die een interventie direct vergelijken met placebo of met een alternatieve interventie. Het doel is om een nauwkeurigere schatting van het effect te bekomen. Maar, in de klinische praktijk komen voor eenzelfde indicatie vaak meerdere alternatieve behandelingen in aanmerking die niet vergeleken zijn met placebo, noch met elkaar onderling in head-to-head-studies. Netwerk meta-analyses proberen dit hiaat op te vullen door directe vergelijkingen te combineren met indirecte vergelijkingen (1-3). Ook hier is het doel om de nauwkeurigheid van de schatting van het effect te vergroten. Het nut en de valkuilen van een netwerk meta-analyse kwamen in Minerva reeds meermaals aan bod (2,3). Zo moeten de geïncludeerde studies voldoende homogeen, vergelijkbaar en consistent zijn om betrouwbare conclusies te kunnen trekken (3).

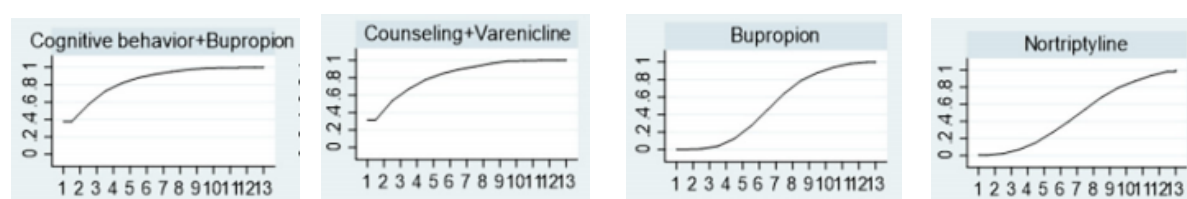
Vaak is het moeilijk om de resultaten van een netwerk meta-analyse correct te interpreteren door de overload aan effectschattingen van de verschillende directe en indirecte vergelijkingen (4). Elke puntschatting gaat bovendien gepaard met een bepaalde graad van onzekerheid. De clinicus moet echter een keuze kunnen maken en stelt zich in de eerste plaats de vraag: ‘Wat is *mogelijk* de beste behandeling voor deze indicatie?’ Op basis van de resultaten van de netwerk meta-analyse kan men voor elke behandeling grafisch voorstellen hoe waarschijnlijk (0,0 tot 1,0 in de Y-as) ze de beste, tweede beste, derde beste,... (X-as) is. In het voorbeeld hieronder zien we dat behandeling A 80% kans heeft om de beste behandeling te zijn en behandeling C 0% kans heeft om de beste behandeling te zijn (5).



Wanneer er veel behandelingen vergeleken worden, wordt het op een bepaald ogenblik moeilijk om op basis van deze grafieken te zien hoe behandelingen zich ten opzichte van elkaar verhouden (4). Daarom is het nuttig om de onderzochte behandelingen te rangschikken op basis van een cumulatieve waarschijnlijkheid. Voor elke behandeling kan men dan grafisch voorstellen hoe waarschijnlijk (0,0 tot 1,0 in de Y-as) ze de beste optie is, maar ook hoe waarschijnlijk ze bij de beste twee, beste drie,...behandelingen behoort (X-as). In het voorbeeld hieronder zien we dat behandeling A 80% kans heeft om de beste behandeling te zijn, maar dat ze daarnaast 90% kans heeft om bij de twee beste behandelingen te behoren en 98% kans heeft om bij de drie beste behandelingen te behoren (5).



Zo zien we in de duiding van de studie over rookstop bij COPD-patiënten (6,7) dat de combinatie cognitieve gedragstherapie + bupropion 40% kans en de combinatie cognitieve gedragstherapie + varenicline 30% kans heeft om de ‘beste’ behandeling te zijn, terwijl de kans voor bupropion alleen en nortriptyline alleen 0% bedraagt om ten opzichte van de andere behandelingen er als beste voor rookstop uit te komen.



Op basis van deze curves lijkt cognitieve gedragstherapie + bupropion er als beste uit te komen, maar... kan je op basis van deze grafieken ook iets zeggen over de onderlinge ranking van bupropion alleen of nortriptyline alleen?

Om dit op te lossen berekent men voor elke behandeling de **AUC** onder de cumulatieve waarschijnlijkheidscurves. Met de zogenaamde Surface under the cumulative ranking curve (SUCRA) kunnen we dan met één cijfer de positie van elke behandeling in de rangschikking voor beste behandeling bepalen. De waarde varieert van 0 tot 100%. Hoe hoger de SUCRA-waarde, hoe meer waarschijnlijk een bepaalde behandeling in de top van de rangorde zit. Hoe lager de SUCRA-waarde, hoe meer waarschijnlijk een bepaalde behandeling onderaan de rangorde bengelt.

Zoals misschien wel te verwachten was, blijkt in de duiding van de studie over rookstop bij COPD-patiënten (6,7) cognitieve gedragstherapie + bupropion de hoogste SUCRA te hebben, zijnde 85%, gevolgd door counseling + varenicline met een SUCRA van 82%. Moeten we als clinicus nu cognitieve gedragstherapie + bupropion verkiezen boven cognitieve gedragstherapie + varenicline? Op basis van de rangorde lijkt bupropion (51%) nu ook beter te zijn dan nortriptyline (47,7%). Merkwaardig is wel dat de SUCRA voor bupropion alleen groter is dan voor counseling + bupropion (35,9%). De duider wijst er terecht op dat dit een hoogst onwaarschijnlijk resultaat is dat indruist tegen eerder onderzoek.

Daarom gaan we in de volgende paragraaf in op de betrouwbaarheid van SUCRA om de keuze voor een bepaalde behandeling te verantwoorden (4).

### 1) SUCRA houdt geen rekening met de graad van bewijs van de directe en indirecte vergelijkingen.

De SUCRA-ratings kunnen afkomstig zijn ofwel van een groot aantal studies met weinig methodologische beperkingen en dus een hoge zekerheid van bewijs, ofwel van een klein aantal studies met een hoog risico van bias (geen concealment of allocation, geen blinding, hoge en ongelijke studie-uitval), met studieresultaten die onnauwkeurig (brede betrouwbaarheidsintervallen), inconsistent (veel statistische heterogeniteit) en indirect (veel klinische heterogeniteit in studiepopulatie of uitkomstmeting) zijn, met een groot risico op publicatiebias en dus een lage tot zeer lage zekerheid van bewijs. Er werd bovendien aangetoond dat het aantal studies, de studiegrootte en de consistentie bepalend zijn voor de robuustheid van de SUCRA-ratings (8).

Wanneer we terugkeren naar onze duiding over rookstop bij COPD-patiënten (6,7) zien we dat volgens SUCRA cognitieve gedragstherapie + bupropion op de eerste plaats en counseling + varenicline op de tweede plaats staat. Wanneer we echter in het supplement de tabel met directe en indirecte vergelijkingen raadplegen zien we dat beide vergelijkingen gebaseerd zijn op een indirect resultaat (geen head-to-head-studies beschikbaar), met belangrijke inconsistentie en onnauwkeurigheid waardoor het resultaat gepaard gaat met een zeer lage zekerheid van het bewijs. Ook voor de vergelijking van bupropion met bupropion + counseling kunnen we niets besluiten als we rekening houden met deze evidentietabel. De SUCRA-rangschikking voor cognitieve gedragstherapie + bupropion, counseling + varenicline, bupropion en bupropion + counseling is dus zeer zwak onderbouwd. Bijgevolg zou het dan ook onjuist zijn om met deze rangschikking veel rekening te houden bij de keuze voor een bepaalde behandeling.

Certainty assessment							N° of patients		Effect	Certainty
N° of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Intervention-1	Intervention-2	OR with 95% CI	
<b>Bupropion vs Counseling+Bupronion</b>										
-	RCT	Not serious	Serious	Very serious	Not serious	Undetected	-	-	1.31 (0.42,4.11)	Very Low
<b>Cognitive behavior+Bupropion vs Counseling+Varenicline</b>										
-	RCT	Not serious	Serious	Very serious	Serious	Undetected	-	-	2.69 (0.49,14.80)	Very Low

**2) SUCRA houdt rekening met één uitkomstmaat, terwijl de clinicus vaak geïnteresseerd is in verschillende relevante uitkomstmaten.** Een behandeling met de meeste winst voor een bepaalde uitkomstmaat kan er voor een andere uitkomstmaat (zoals veiligheid) als slechtste behandeling uitkomen.

In een eerdere duiding van Minerva over het effect van geneesmiddelen bij paniekstoornissen hield men wel rekening met twee relevante uitkomstmaten (9,10). SUCRA rangschikte benzodiazepines, tricyclische antidepressiva en SSRI's bovenaan op vlak van remissie. Maar, omdat benzodiazepines en tricyclische antidepressiva meer ongewenste effecten hadden, wonnen SSRI's het pleit voor de 'beste' behandeling.

**3) SUCRA houdt geen rekening meer met de grootte van de verschillen in effect tussen twee behandelingen.** Dat maakt het voor de clinicus moeilijk om in te schatten of het verschil wel klinisch relevant is.

Zo zien we in ons voorbeeld over rookstop bij COPD-patiënten (6,7) dat het verschil tussen cognitieve gedragstherapie + bupropion versus counseling + varenicline in een indirecte vergelijking statistisch niet significant is, waardoor de klinische relevantie zelfs niet te bepalen is.

**4) SUCRA blijft het resultaat van een netwerk meta-analyse waarbij verschillen tussen behandelingen louter op toeval kunnen berusten.**

SUCRA houdt dus geen rekening met de breedte van de betrouwbaarheidsintervallen van de directe en indirecte vergelijkingen (8,11). Merken we hierbij ook op dat de winst in nauwkeurigheid door het poolen van indirecte vergelijkingen lager is dan van directe vergelijkingen. Vier gerandomiseerde onderzoeken met indirecte vergelijkingen geven slechts evenveel winst in nauwkeurigheid als één RCT met een directe vergelijking (1). De nauwkeurigheid neemt verder af wanneer er meer schakels in het netwerk worden opgenomen (1). Het is bijgevolg belangrijk dat het te onderzoeken geneesmiddel in ieder geval één keer direct is verbonden met het netwerk (1).

## Besluit

Netwerk meta-analyses zijn nuttig om verschillende behandelingen met elkaar te vergelijken. Surface under the cumulative ranking curve (SUCRA) werd ontwikkeld om de clinicus te helpen bij de interpretatie van het resultaat van een netwerk meta-analyse. Wanneer het aantal directe vergelijkingen echter beperkt is en wanneer de graad van bewijs van de directe en indirecte vergelijkingen laag tot zeer laag is, mogen we als clinicus geen rekening houden met het resultaat van de SUCRA.

## Referenties

1. Stolk LM. Netwerk meta-analyse van geneesmiddelen. *GEBU* 2020;54:89-95. DOI: 10.35351/gebu.nl.2020.8.15|13-08-2020
2. Chevalier P. Netwerk meta-analyses: directe en indirecte vergelijkingen. *Minerva* 2009;8(9):136.
3. Chevalier P. Indirecte vergelijkingen (vervolg). *Minerva* 2011;10(6):77.
4. Mbuagbaw L, Rochweg B, Jaeschke R, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. *Syst Rev* 2017;6:79. DOI: 10.1186/s13643-017-0473-z
5. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011;64:163-71. DOI: 10.1016/j.jclinepi.2010.03.016
6. Boudrez. Effectiviteit van rookstopinterventies bij COPD-patiënten. *Minerva* 2023;22(4):71-4.
7. Wei X., Guo K., Shang X, et al. Effects of different interventions on smoking cessation in chronic obstructive pulmonary disease patients: a systematic review and network meta-analysis. *Int J Nurs Stud* 2022;136:104362. DOI: 10.1016/j.ijnurstu.2022.104362
8. Daly CH, Neupane B, Beyene J, et al. Empirical evaluation of SUCRA-based treatment ranks in network meta-analysis: quantifying robustness using Cohen's kappa. *BMJ Open* 2019;9:e024625. DOI: 10.1136/bmjopen-2018-024625
9. Catthoor K. Het effect van verschillende geneesmiddelen bij paniekstoornis. *Minerva Duiding* 21/03/2023.
10. Chawla N, Anothaisintawee T, Charoenrungrueangchai K, et al. Drug treatment for panic disorder with or without agoraphobia: systematic review and network meta-analysis of randomised controlled trials. *BMJ* 2022;376:e066084. DOI: 10.1136/bmj-2021-066084
11. Rucker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015;15:58. DOI: 10.1186/s12874-015-0060-8