

# Introductie van Bayesiaanse analyses binnen de vaktherapie

Een eerste verkenning



Onderzoek naar effecten van vaktherapie is geworteld in de praktijk en wordt veelal uitgevoerd door vaktherapeuten en onderzoekers. Vanuit deze co-creatie zou er meer gebruik kunnen worden gemaakt van de Bayesiaanse methodiek. Deze methodiek is interessant omdat praktijkkennis als uitgangspunt kan worden gebruikt, en ook vanwege het feit dat er daarom per definitie geen grote aantallen respondenten nodig zijn. Hier zal de Bayesiaanse methodiek worden geïntroduceerd door een traditionele analyse te vergelijken met een soortgelijke Bayesiaanse analyse. De gebruikte data komen uit een onderzoek naar effecten van muziektherapie in de forensische praktijk, waarbij is gekeken of muziektherapie het uiten van emoties en copingvaardigheden kon verbeteren.

## Inleiding

Als het gaat om onderzoek binnen de vaktherapie dienen we enerzijds oog te hebben voor de context van de vaktherapeutische behandeling en de cliëntendoelgroep en anderzijds moeten we streven naar gedegen wetenschappelijke methodologie en statistiek (Van Hooren, 2018; Raad voor Volksgezondheid en Samenleving, 2017). Al vaker is voorgesteld dat bij onderzoek binnen vaktherapie, de Bayesiaanse methodiek goed aansluit, vanwege het gebruik van voorkennis maar ook vanwege het feit dat daarom per definitie geen grote aantallen respondenten nodig zijn (Waterink & Van Hooren, ingediend). De methodiek is gebaseerd op Bayesiaanse kansrekening, waarbij van te voren verwachtingen worden gecombineerd met de verzamelde data uit onderzoek. Een eerste Bayesiaanse statistische analyse gaat uit van een ruwe modellering, waarbij op basis van ervaring, kennis, overtuigingen en inzichten een schatting wordt gedaan. Samen met verzamelde informatie wordt vervolgens met behulp van de regel van Bayes een verbeterde schatting van de kansverdeling verkregen (zie voor meer uitleg Waterink & Van Hooren, ingediend).

De Bayesiaanse methodiek is nog niet overal geaccepteerd (Wagenmaker, 2015). Maar daar lijkt, in het geval van de psychologie, nu langzaam verandering in te komen. Dat blijkt onder andere uit een *systematic review* (Van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg & Depaoli, 2017), waarbij 1.579 psychologische artikelen zijn gevonden die gaan over de Bayesiaanse methodiek tussen 1990 en 2015. Daaruit kwam naar voren dat het gebruik van de Bayesiaanse methodiek in de toegepaste psychologie een constante stijging laat zien en zelfs een vlucht heeft genomen. De reden hiervoor is dat de methodiek op een flexibele manier gebruikt kan worden om vele verschillende vraagstellingen te beantwoorden (Van de Schoot et al., 2017; Wagenmakers & Gronau, 2018; Woertman, Groenewoud & Van der Wilt, 2014) en omdat er nu eindelijk toegankelijke softwarepakketten zijn die vraagstukken op een Bayesiaanse manier kunnen toetsen (Marsman & Wagenmakers, 2017).

In onderzoek waarbij traditionele statistiek wordt gebruikt, gaat het om het toetsen van hypothesen, zoals: 'Werkt een vaktherapeutische behandeling wel of niet?' Om dat te bepalen worden passende analyses gekozen met een zelf gekozen grenswaarde, aangeduid via de p-waarde. In de traditionele statistiek is de grens van die p-waarde dus arbitrair. Bij de Bayesiaanse statistiek gaat dat anders. Daar gaat het om het relatieve verschil tussen het bewijs voor de hypothesen die tegen elkaar worden afgezet. Het relatieve verschil wordt uitgedrukt in een cijfer, de Bayes-factor. De Bayes-factor geeft aan hoe goed hypothesen ten opzichte van elkaar bij de verzamelde data passen.

Om de Bayesiaanse methodiek verder uit te leggen, zal deze worden vergeleken met de traditionele methodiek, door hypothesen uit een eerder uitgevoerd onderzoek via traditionele statistiek en Bayesiaanse statistiek te toetsen. Het eerdere onderzoek betrof een verkennende studie naar muziektherapie

” De Bayes-factor geeft aan hoe goed hypothesen ten opzichte van elkaar bij de verzamelde data passen

rapie in de forensische praktijk (De Witte, 2015a), waarbij bij forensische patiënten met lichte verstandelijke beperkingen is gekeken of muziektherapie het uiten van hun emoties en hun copingvaardigheden kon verbeteren.

## De p-waarde versus de Bayes-factor

In de huidige onderzoekspraktijk wordt bij de statistische toetsen nog steeds voornamelijk de p-waarde gebruikt. De p-waarde is een waarde tussen 0 en 1. Hoe lager deze waarde, des te groter de kans dat de nulhypothese wordt verworpen en dus de alternatieve hypothese wordt aangenomen. In de verkennende studie van De Witte (2015a), is de nulhypothese dat de mensen die muziektherapie hebben gehad niet verschillen in het uiten van emoties en copingvaardigheden ten opzichte van de mensen die geen muziektherapie hebben gehad. De alternatieve hypothese is dus dat er wel verschillen zijn tussen deze groepen. Om na te gaan wanneer de alternatieve hypothese aangenomen wordt, kiest de onderzoeker een bepaalde p-waarde als grens. Meestal wordt gekozen voor een grenswaarde .05, of als we strenger zijn, voor .01. Hiermee kan worden aangeduid of een resultaat statistisch significant is en een resultaat dus daadwerkelijk noemenswaardig is. In het verkennend onderzoek was dat .05. In onderzoek houdt een nulhypothese in dat na de interventie er geen verschil is tussen de experimentele groep en de controlegroep. Deze grenswaarde betekent in dit geval dat als een p-waarde kleiner is dan .05, de kans dat een bepaald effect aan het toeval is te wijten kleiner is dan 5%. Dat wordt in de traditionele onderzoekswereld acceptabel gevonden. Als de gevonden p-waarde in het geval van de verkennende studie kleiner is dan .05, is er sprake van een statistisch significant effect. Er mag dan gesteld worden dat de muziektherapie-interventie een effect heeft gehad. Wel wordt dan geaccepteerd dat er een kans van ten hoogste 5% is dat het gevonden effect toch op toeval berust. Wanneer een p-waarde significant is, betekent dit niet dat de nulhypothese per definitie dus niet klopt.

Misschien is het effect zo klein dat het niet is waargenomen. Andersom, een grote gevonden p-waarde betekent niet dat er geconcludeerd mag worden dat de nulhypothese juist is. Is de p-waarde groter dan .05, dan is een kans dus groter dan 5%

## IN DIT ARTIKEL

- Bayesiaanse analyses en onderzoek vaktherapie.
- P-waarde versus Bayes-factor.
- Muziektherapie in de forensische praktijk.

” In onderzoek houdt een nulhypothese in dat na de interventie er geen verschil is tussen de experimentele groep en de controlegroep

dat het gevonden effect op toeval berust. In feite mag er dan alleen maar gezegd worden dat de nulhypothese verworpen moest worden. Verder kan er nooit bewezen worden dat de nulhypothese correct is.

Binnen de Bayesiaanse methodiek worden de nulhypothese en de alternatieve hypothese met elkaar vergeleken en uitgedrukt via de Bayes-factor. Dat kan op twee manieren. In het geval dat de alternatieve hypothese wordt vergeleken ten opzichte van de nulhypothese, wordt de Bayes-factor uitgedrukt in  $BF_{10}$ . Als, bijvoorbeeld,  $BF_{10} = 10$ , dan geeft dat aan dat in de data voor de alternatieve hypothese 10 keer meer steun te vinden is dan voor de nulhypothese. Als de het gaat om dezelfde data waarbij van de nulhypothese wordt vergeleken ten opzichte de alternatieve hypothese, wordt de Bayes-factor uitgedrukt in  $BF_{01}$ , maar heeft dan de waarde  $.10$ . In Tabel 1 staan vuistregels voor de interpretatie van de Bayes-factor uitgedrukt in  $BF_{10}$ .

### Toepassing van de Bayesiaanse methodiek

Om te laten zien hoe de Bayesiaanse methodiek kan worden toegepast, wordt de traditionele statistiek van het hiervoor besproken verkennende onderzoek vergeleken met de situatie wanneer de onderzoeksvragen beantwoord worden via de Bayesiaanse statistiek. Hiervoor is gebruikgemaakt van data uit een onderzoek dat werd uitgevoerd binnen de forensische psychiatrische afdeling van STEVIG te Oostrum. Binnen dit onderzoek werd er gebruikgemaakt van een selecte (doelgerichte) steekproef waarbij alleen patiënten met lichte verstandelijke beperkingen in aanmerking kwamen. De onderzoeksvraag was of muziektherapie ten opzichte van *care as usual* bij deze doelgroep leidt tot meer verbetering in het uiten van emoties en copingvaardigheden. De nulhypothese is dat beide groepen gelijk zijn aan elkaar. De alternatieve hypothese dat beide groepen niet aan elkaar gelijk zijn.

Er zijn tien patiënten geselecteerd uit een totale populatie van 28 patiënten. Door randomisatie is er een controle- en een experimentele groep gevormd. Alle patiënten is vooraf om toestemming gevraagd tot medewerking aan het onderzoek. Eén van de patiënten die door de randomisatie geselecteerd was voor de experimentele groep, gaf aan niet deel te willen ne-

## TABEL 1

Interpretatie van de Bayes-factor ( $BF_{10}$ ) (naar Wetzels & Wagenmakers, 2012)

Bayes factor ( $BF_{10}$ )	Interpretatie
> 100	Beslissend bewijs voor de alternatieve hypothese
30 - 100	Zeer sterk bewijs voor de alternatieve hypothese
10 - 30	Sterk bewijs voor de alternatieve hypothese
3 - 10	Substantieel bewijs voor de alternatieve hypothese
1 - 3	Anekdotisch bewijs voor de alternatieve hypothese
1	Geen bewijs
.33 - 1.00	Anekdotisch bewijs voor de nulhypothese
.10 - .33	Substantieel bewijs voor de nulhypothese
.03 - .10	Sterk bewijs voor de nulhypothese
.10 - .001	Zeer sterk bewijs voor de nulhypothese
.001 >	Beslissend bewijs voor de nulhypothese

## TABEL 2

Gemiddelde scores van de Observer Alexithymia Scale

OAS	Experimentele groep (n=3)		Controlegroep (n=6)	
	Voormeting	Nameting	Voormeting	Nameting
Afzijdig	20.83 (3.21)	19.83 (1.89)	19.33 (6.08)	19.25 (5.81)
Gebrek aan inzicht	15.83 (1.04)	14.50 (1.80)	16.83 (2.09)	17.00 (1.58)
Somatisatie	5.33 (3.88)	6.00 (1.32)	4.83 (3.27)	4.83 (3.06)
Gebrek aan humor	7.33 (2.75)	5.33 (2.25)	7.42 (2.62)	7.83 (2.07)
Rigide	5.83 (1.53)	6.00 (1.80)	7.17 (2.82)	7.75 (3.74)

men aan de muziektherapie-interventie. Deze patiënt is toegevoegd aan de controlegroep. De experimentele groep telde daarom vier patiënten, de controlegroep telde zes patiënten. De demografische kenmerken van beide groepen zijn met elkaar vergeleken door middel van de Chi-Square Test. Daaruit bleek dat beide groepen als gelijkwaardig kunnen worden beschouwd betreffende sekse ( $t(1) = .08, p = .778$ ), het gebruik van dempende medicatie ( $t(1) = .08, p = .778$ ), deelname aan een cognitieve training gericht op agressiebeheersing ( $t(1) = .00, p = 1.000$ ), leeftijd ( $t(8) = 7.92, p = .442$ ), de hoogte van het IQ ( $t(7) = 7.92, p = .340$ ) en wat betreft totale opnameduur in maanden ( $t(5) = 6.88, p = .230$ ).

De experimentele groep kreeg de muziektherapie-interventie tien weken aangeboden (één sessie van 45 minuten per week). De muziektherapie-interventie die in deze studie is gebruikt is 'Muziek als Krachtpatser'. Deze interventie is beschreven in het consensus based-behandelproduct ervaringsgerichte muziektherapie bij forensische patiënten (De Witte, 2014b). Week één tot en met vijf werd er gewerkt met percussie. Werkvormen waren gericht op *finetuning* en het kunnen aanbrengen van gradatie in spanningsopbouw op percussie-instrumenten. In week zes tot en met tien werden melodie-instrumenten toegevoegd. De therapeut intervenieerde aan de hand van muzikale parameters door te spiegelen, over te nemen, te vergroten of juist te verkleinen. De controlegroep kreeg *care as usual*. Dit betekende dat zij enkel het eigen individueel behandelprogramma volgden.

De groepen werden door middel van een voor- en nameting na tien weken met elkaar vergeleken. Het effect op gedragsniveau werd bij beide groepen gemeten met twee meetinstrumenten. Uiten van emoties werd gemeten door de in het Ne-

” Gaat het er bij traditionele statistiek om of er wel of niet een effect is; bij Bayesiaanse statistiek gaat het juist om een relatieve formulering

derlands vertaalde *Observer Alexithymia Scale* (OAS; Haviland, Warren & Riggs, 2000). De OAS heeft vijf subschalen; afzijdig, gebrek aan inzicht, somatisatie, gebrek aan humor, en rigiditeit. De schaal bevat 33 items. Totale scores variëren tussen de 0 en 99. Copingvaardigheden zijn gemeten met de Utrechtse Coping Lijst (UCL; Schreurs, Van de Willige, Tellegen & Brosschot, 1993). De UCL bestaat uit 47 items die gescoord worden via een vierpuntschaal (zelden of nooit, soms, vaak, zeer vaak) verdeeld over zeven subschalen: actief aanpakken, palliatieve reactie, vermijden, sociale steun zoeken, passief reactiepatroon, expressie van emoties en geruststellende gedachten. Zowel de hoofdbehandelaars (GZ-psychologen) als de persoonlijk begeleiders voerden deze metingen uit. De metingen van deze twee observanten werden telkens gemiddeld. Binnen de experimentele groep was in week zeven sprake van een uitval door acute beëindiging van de behandeling. Dat betekende dat de analyses van de UCL en OAS zijn uitgevoerd met een totale populatie van negen patiënten.

De statistische analyses zijn verricht met de software IBM SPSS Statistics 20. Als toets is de *Independent Samples T-test* gebruikt om de nulhypothese tegen de alternatieve hypothese te toetsen, met een significantieniveau van 5%, oftewel een p-waarde met een grens van .05. De test is telkens uitgevoerd met de gemiddelde verschijscores tussen de voor- en nametingen.



'Muziek als Krachtpatser'. De eerste vijf weken wordt er gewerkt met percussie.

” De onderzoeksvraag was of muziektherapie ten opzichte van care as usual leidt tot meer verbetering in het uiten van emoties en copingvaardigheden

De Bayesiaanse analyses zijn uitgevoerd met het open source software pakket JASP (JASP Team, 2019). Ter vergelijking is dezelfde data getoetst met een *Bayesiaanse Independent Samples T-Test*. Omdat de kwaliteit van de voorkennis belangrijk is als de  $n$  klein is (Lee & Song, 2004, Van de Schoot et al., 2014; Van de Schoot, Schalken & Olf, 2017), is een gemiddelde *effect size* van .723 (Cohen's  $D$ ) overgenomen uit een meta-analyse (De Witte et al., in press). Deze laatstgenoemde studie laat een multi-level meta-analyse zien van 47 kwantitatieve studies waarin de effecten van muziektherapie op het verminderen van stress/momentane angst zijn onderzocht.

In Tabel 2 en 3 zijn de gemiddelden en standaarddeviaties per groep te vinden. Uitgaande van de traditionele Independent Samples T-Test, kwam naar voren dat, wat betreft de OAS-verschiltotaalscore, er een verschil is gevonden tussen de twee groepen ( $t(7) = 2.94, p = .022$ ). De nulhypothese, dat er geen verschil is tussen de experimentele groep ( $M = 5.77, SD = 4.58$ ) en controlegroep ( $M = -.77, SD = 2.33$ ), moest dus worden verworpen. Onderliggend aan het gevonden hoofdeffect zijn er verschillen gevonden tussen de twee groepen bij de subschalen Somatisatie ( $t(7) = 5.68, p = .001$ ) en Gebrek aan humor ( $t(7) = 3.37, p = .012$ ). Er waren geen verschillen tussen de twee groepen voor Afzijdig ( $t(7) = .86, p = .416$ ), Gebrek aan inzicht ( $t(7) = 1.06, p = .324$ ) en Rigide ( $t(7) = .40, p = .703$ ). Wanneer de onderzoeksvraag wordt beantwoord via de Bayesiaanse Independent Samples T-Test, dan heeft voor de OAS-totaalscore de Bayes-factor ( $BF_{10}$ ) een waarde van 3.12. Strikt genomen betekent dit dat er een substantieel bewijs is voor de alternatieve hypothese, dus dat beide groepen niet aan elkaar





gelijk zijn. Zeer sterk bewijs voor de alternatieve hypothese is gevonden voor Somatisatie ( $BF_{10} = 32.10$ ). Voor Gebrek aan humor is er substantieel bewijs gevonden ( $BF_{10} = 4.55$ ) voor de alternatieve hypothese. Voor Afzijdig ( $BF_{10} = .64$ ), voor Gebrek aan inzicht ( $BF_{10} = .71$ ) en Rigide ( $BF_{10} = .54$ ) is er anekdotisch bewijs gevonden voor de nulhypothese.

Wat betreft de schalen uit de UCL, zijn er via de traditionele statistiek geen verschillen gevonden tussen de twee groepen, wat betreft Actief aanpakken ( $t(7) = -.38, p = .712$ ), Palliatieve reactie ( $t(7) = -.52, p = .620$ ), Vermijden ( $t(7) = 2.22, p = .062$ ), Sociale steun zoeken ( $t(7) = .25, p = .813$ ), Passief reactiepatroon ( $t(6,8) = .97, p = .363$ ), Expressie van emoties ( $t(7) = -.35, p = .736$ ) en Geruststellende gedachte ( $t(7) = .81, p = .447$ ). Voor alle subschalen geldt dat de nulhypothese niet verworpen moest worden. Via de Bayesiaanse Independent Samples T-Test is er anekdotisch bewijs is gevonden voor de nulhypothese bij de subschalen Actief aanpakken ( $BF_{10} = .54$ ), Palliatieve reactie

” Het verkrijgen van goede kwantitatieve gegevens blijft voor beide methodieken een voorwaarde voor de evaluatie van een interventie

( $BF_{10} = .56$ ), Sociale steun zoeken ( $BF_{10} = .53$ ), Passief reactiepatroon ( $BF_{10} = .62$ ), Expressie van emoties ( $BF_{10} = .54$ ) en Geruststellende gedachte ( $BF_{10} = .63$ ). Anekdotisch bewijs voor de alternatieve hypothese is gevonden voor Vermijden ( $BF_{10} = 1.65$ ).

### Vergelijking van de traditionele statistiek met de Bayesiaanse statistiek

Gaat het er bij traditionele statistiek om of er wel of niet een effect is; bij Bayesiaanse statistiek gaat het juist om een relatieve formulering (Waterink & Van Hooren, ingediend). Ten aanzien van de vergelijking was te zien dat de uitkomsten van beide analyse-methoden redelijk overeenkomen. Hoewel er sprake was van een kleine  $n$ , is in het verkennende onderzoek geconcludeerd dat er sprake was van een significant verschil tussen de experimentele groep en de controlegroep als het gaat om het uiten van emoties. Meer specifiek, de mensen die muziektherapie hadden gevolgd lieten een verbetering zien ten opzicht van de mensen die *care as usual* ontvingen. Dit gold specifiek voor somatisatie en humor. In de subschaal 'copingvaardigheden' verschilden de twee groepen niet. De vergelijkbare Bayesiaanse analyses gaven een soortgelijk beeld. Hier is de conclusie dat er drie keer meer bewijs is voor de hypothese dat de twee groepen in het uiten van emoties niet aan elkaar gelijk zijn, dan dat ze wel aan elkaar gelijk zijn. Meer specifiek, dat er 32 keer meer bewijs is gevonden dat de twee groepen verschilden als het ging om somatisatie en vier keer meer bewijs als het ging om humor.

Ten aanzien van copingvaardigheden is er voor de meeste copingvaardigheden anekdotisch bewijs gevonden voor de nulhypothese, behalve voor de mate van vermijding, waarbij bleek dat er twee keer meer bewijs is voor de hypothese dat de mensen die muziektherapie hebben gevolgd verschillen van de mensen die *care as usual* hebben gehad. Dit heeft vooral te maken met het gegeven dat we bij de Bayesiaanse statistiek voorkennis kunnen meenemen. In dit voorbeeld hebben we de voorkennis vanuit een eerder verrichte meta analyse kunnen gebruiken (De Witte, et al., in press). Het ligt voor de hand dat het verkennende onderzoek van De Witte (2015a) een keer herhaald wordt, met het anekdotisch bewijs als voorkennis. Want hoe zorgvuldiger er telkens met voorkennis wordt omgegaan, hoe zinvoller de toekomstige voorspellingen, en hoe betekenisvoller de statistische conclusies worden (Wagenmakers & Gronau, 2018).

Het verkrijgen van kwantitatieve gegevens van goede kwaliteit blijft voor zowel de Bayesiaanse methodiek als de traditionele methodiek een noodzakelijke voorwaarde voor de evaluatie van een interventie. We willen daarom enkele methodologische

## TABEL 3

### Gemiddelde scores van de Utrechtse Coping Lijst

UCL	Experimentele groep (n=3)		Controlegroep (n=6)	
	Voormeting	Nameting	Voormeting	Nameting
Actief aanpakken	9.83 (1.04)	11.17 (1.89)	9.00 (1.64)	11.00 (1.84)
Palliatieve reactie	20.33 (.76)	20.83 (1.26)	19.08 (3.60)	19.08 (2.91)
Vermijden	19.67 (3.55)	18.50 (2.65)	14.67 (2.18)	15.67 (2.42)
Sociale steun zoeken	7.83 (2.36)	9.17 (3.33)	13.83 (4.20)	13.58 (3.81)
Passief reactiepatroon	18.50 (1.32)	16.67 (2.55)	18.17 (2.99)	17.25 (2.25)
Expressie van emoties	7.50 (1.80)	8.00 (1.80)	8.75 (1.29)	8.75 (1.44)
Geruststellende gedachten	9.50 (1.80)	9.67 (1.15)	8.50 (1.87)	8.83 (2.32)

## OVER DE AUTEURS

**Dr. Wim Waterink** is hoofddocent Algemene psychologie bij de Open Universiteit, waar hij momenteel als programmaleider verantwoordelijk is voor de bacheloropleiding Psychologie, met als zwaartepunt de biologische en evolutionaire grondslagen van de psychologie en de psychogerontologie.  
E: wim.waterink@ou.nl

**Martina de Witte** is docent muziektherapie en coördinator onderzoek bij de opleiding voor vaktherapie van de Hogeschool van Arnhem en Nijmegen. Als onderzoeker is zij verbonden aan behandelinstelling STEVIG ([www.stevig.nl](http://www.stevig.nl)), het lectoraat KenVaK ([www.kenvak.nl](http://www.kenvak.nl)), en aan de Universiteit van Amsterdam waar zij een promotieonderzoek uitvoert naar de effecten en inzetbaarheid van muziektherapie voor het verminderen van stress.

**Prof. dr. Susan van Hooren** is bijzonder hoogleraar Vaktherapie bij de Open Universiteit en Lector Kennisontwikkeling Vaktherapieën (KenVaK). Ze leidt de Coöperatie KenVaK ([www.kenvak.nl](http://www.kenvak.nl)), het landelijke kennisnetwerk van vaktherapie waar zes hogescholen aan participeren: Hogeschool van Arnhem en Nijmegen, Hogeschool Utrecht, Hogeschool Leiden, Codarts Hogeschool voor de kunsten, NHL-Stenden Hogeschool en Zuyd Hogeschool. Susan is adviseur en docent bij de Master Vaktherapie.  
E: susan.vanhooren@zuyd.nl

kanttekeningen plaatsen bij het uitgevoerde onderzoek. Zo is er sprake van een zeer kleine steekproef in het verkennende onderzoek. Het is dus sterk aan te raden om het oorspronkelijke onderzoek met een grotere groep uit te voeren. Verder werd in de verkennende studie gewerkt met verschillen tussen de voor- en de nameting. Om meer power te behouden zou het meer voor de hand liggen om een geavanceerde statistische techniek te verrichten, waarbij alle ruwe scores in een model gelijktijdig worden meegenomen. Er werd in de standaardstatistiek geen analyse toegepast volgens de *intention to treat principle*. Zo werd de persoon die eerder was toegewezen aan de experimentele groep maar geen muziektherapie wilde, uiteindelijk ingedeeld in de controlegroep. Volgens de *intention to treat principle* zou deze persoon in de experimentele groep moeten blijven om uit te kunnen sluiten dat er een onderliggend nadelig effect zit aan de keus om geen muziektherapie te willen. Deze methodologische kanttekeningen doen voor nu niet af aan het doel in het voorliggend artikel, namelijk, om twee methodieken met elkaar te vergelijken. In deze vergelijking is volledig aangesloten op de eerder gemaakte keuzes.

### Conclusie

De Bayesiaanse methodiek sluit ons inziens beter aan bij de context van de vaktherapeutische praktijk dan de traditionele methodiek. De voornaamste reden hiervoor is dat steeds meer uitkomsten van vaktherapeutisch onderzoek beschikbaar komen, wat gebruikt kan worden als voorkennis in Bayesiaanse statistische analyses bij onderzoek bij met name kleine patiëntgroepen. ■

## SAMENVATTING

In het huidige artikel is traditionele statistiek vergeleken met Bayesiaanse statistiek. Voor dit vergelijkend onderzoek is gebruikgemaakt van data uit een verkennend onderzoek naar effecten van muziektherapie in de forensische praktijk, waarbij bij drie forensische patiënten met lichte verstandelijke beperkingen is gekeken of muziektherapie het uiten van hun emoties en hun copingvaardigheden kon verbeteren. In de vergelijking werden de resultaten van de uitgevoerde *In-dependent Samples T-test* van het verkennend onderzoek vergeleken met die van de Bayesiaanse variant, waarbij voorkennis werd benut vanuit een eerder verrichte meta-analyse. Het vergelijkende onderzoek liet qua uitkomsten voor beide typen analyses een soortgelijk beeld zien. In het verkennende onderzoek is, als het gaat om het uiten van emoties, geconcludeerd dat er sprake was van een significant verschil tussen de drie patiënten uit experimentele groep en de zes patiënten uit de controlegroep.

In termen van de Bayesiaanse methodiek is de conclusie dat er drie keer meer bewijs is voor de hypothese dat de twee groepen ten aanzien van het uiten van emoties niet aan elkaar gelijk zijn, dan dat ze wel aan elkaar gelijk zijn. Wat betreft de copingvaardigheden was in het verkennende onderzoek de conclusie dat voor alle subschalen de nulhypothese niet verworpen moest worden. In Bayesiaanse termen is er voor de meeste copingvaardigheden anekdotisch bewijs gevonden voor de nulhypothese, behalve voor de mate van vermijding, waarbij bleek dat er twee keer meer bewijs is voor de hypothese dat de mensen die muziektherapie hebben gevolgd verschillen van de mensen die *care as usual* hebben gehad. De Bayesiaanse methodiek sluit ons inziens daarom beter aan bij de context van de vaktherapeutische praktijk, omdat steeds meer uitkomsten van onderzoek beschikbaar komen wat gebruikt kan worden als voorkennis in Bayesiaanse analyses bij onderzoek bij kleine patiëntgroepen.

## Literatuur

- Asendorpf, J.B., Schoot, R. van de, Denissen, J.J. & Hutteman, R. (2014). Reducing bias due to systematic attrition in longitudinal studies: *The benefits of multiple imputation*. *International Journal of Behavioral Development*, 38(5), 453-460. doi: 10.1177/0165025414542713
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. doi:10.3389/fpsyg.2014.00781
- Haviland, M.G., Warren, W.L. & Riggs, M.L. (2000). An observer scale to measure alexithymia. *Psychosomatics*, 41(5), 85-92. doi:10.1176/appi.psy.41.5.385
- Hooren, S.A.H. van (2018). *Vaktherapie: Doen wat werkt* [Oratie]. Open Universiteit, Heerlen.
- JASP Team (2019). JASP (Version 0.11.1) [Computer software].
- Lee, S.Y. & Song, X.Y. (2004). Evaluation of the Bayesian and Maximum Likelihood Approaches in Analyzing Structural Equation Models with Small Sample Sizes. *Multivariate Behavioral Research*, 39(4), 653-86. doi:10.1207/s15327906mbr3904\_4
- Marsman, M. & Wagenmakers, E.J. (2017) Bayesian benefits with JASP. *European Journal of Developmental Psychology*, 14(5), 545-555. doi:10.1080/17405629.2016.1259614
- Quintana, D.S. & Williams, D.R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry* 18, 178. doi:10.1186/s12888-018-1761-4
- Raad voor Volksgezondheid en Samenleving (2017). *Zonder context geen bewijs. Over de illusie van evidence-based practice in de zorg*. Den Haag: RVS.
- Schoot, R. van de, Kaplan, D., Denissen, J., Asendorpf, J.B., Neyer, F.J. & Aken, M.A.G. van (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842-860. doi:10.1111/cdev.12169
- Schoot, R. van de, Schalken, N. & Olff, M. (2017). Systematic search of Bayesian statistics in the field of psychotraumatology. *European Journal of Psychotraumatology*, 8 (sup1). doi:10.1080/20008198.2017.1375339
- Schoot, R. van de, Winter, S.D., Ryan, O., Zonder-van-Zwijenburg, M. & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2), 217-239. doi:10.1037/met0000100
- Schreurs, P.J.G., Willige, G. van de, Brosschot, J.F., Tellegen, B. & Graus, G.M.H. (1993). *Handleiding Utrechtse Coping Lijst UCL (herziene versie)*. Lisse: Swets & Zeitlinger.
- Wagenmaker, E.J. (2015). Statistiek van het gezonde verstand. *Skepter*, 2, 23-25.
- Wagenmakers, E.J. & Gronau, Q.F. (2018). De Bayesiaanse leerproces. *STAOR*, 19(4), 8-13.
- Woertman, W.H., Groenewoud, H.M.M. & Wilt, G.J. van der (2014). Bayesiaanse statistiek: wat, hoe en waarom? *Tijdschrift voor geneeskunde*, 158, 1-6.
- Waterink, W. & Hooren, S.A.H. van (Ingediend bij Tijdschrift voor vaktherapie). *Het benutten van praktijkervaring en -kennis binnen vaktherapeutisch onderzoek: de introductie van de Bayesiaanse methodiek*.
- Wetzels, R. & Wagenmakers, E.J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, 19(6), 1057-1064. doi:10.3758/s13423-012-0295-x
- Witte, M.J. de (2014a). Muziektherapie en emotieregulatie: een pilotstudie bij forensische patiënten met een licht verstandelijke beperking Muziek als krachtpatser. *Tijdschrift voor vaktherapie*, 3, 13-22.
- Witte, M. de (2014b). Muziek als Krachtpatser: Ervangersgerichte muziektherapie bij muziktherapie bij forensische LVB patiënten, waarbij emotieregulatie problematiek een risicofactor is voor delinquent gedrag. Behandelpatser STEVIG.
- Witte, M. de, Silva Pinho, A. da, Stams, G.-J., Moonen, X., Bos, A.E.R. & Hooren, S.A.H. van (2020). Music therapy for stress reduction: A systematic review and meta-analysis. *Health psychology review*. doi: 10.1080/17437199.2020.1846580