

Reactie op

'Glasgow Antipsychotica Bijwerkingen Schaal voor Clozapine; validatie van de Nederlandstalige versie'

Met veel belangstelling lezen wij het artikel van Huisman e.a. (2021) over de betrouwbaarheid en validiteit van de reviseerde versie van de *Glasgow Antipsychotica Bijwerkingen Schaal voor Clozapine* (GASS-C-NLR2). De auteurs onderzochten de test-hertestbetrouwbaarheid met Spearmans correlatiecoëfficiënt. De score bleek 0,68 te zijn bij een interval van een week. Hieruit concluderen de auteurs dat de GASS-C-NLR2 een hoge test-hertestbetrouwbaarheid heeft. Bij deze conclusie willen wij een kanttekening plaatsen: de toegepaste statistische analyse voor de test-hertestbetrouwbaarheid is volgens ons minder geschikt voor een betrouwbaarheidsanalyse in deze context. Wij zullen dit toelichten.

Over het algemeen wordt betrouwbaarheid van metingen gedefinieerd als de mate waarin een meting consistent is en vrij van fouten (Portney & Watkins 2000). De test-hertestbetrouwbaarheid zoals Huisman e.a. die onderzoeken, betreft een analyse van twee metingen met een tijdsinterval van meerdere dagen tot weken betreffende dezelfde patiënt. Dit type onderzoek is belangrijk en vormt een essentieel onderdeel van psychometrische evaluatie van instrumenten. Een lage betrouwbaarheid kan namelijk verstrekende consequenties hebben. Deze vergroot de kans dat onjuiste medische beslissingen worden genomen, vermindert de zeggingskracht (power) van wetenschappelijk onderzoek, kan leiden tot een hogere placeborespons en onbetrouwbaarheid verzwakt de grootte van gevonden associaties.

Spearmans correlatiecoëfficiënt is een non-parametrische correlatiecoëfficiënt om de grootte en richting van een associatie tussen twee metingen te berekenen. Deze methode houdt echter geen rekening met de overeenkomst of het gemiddelde verschil tussen metingen, hetgeen

Tabel 1. Fictief voorbeeld test-hertestbetrouwbaarheid

Meting 1	Meting 2
1	3
2	4
3	5
1	3
2	3
3	5
1	3
2	4
3	5
1	3

AUTEURS

Steven Berendsen, arts in opleiding tot psychiater, promovendus, Arkin, UMC Amsterdam, locatie AMC.

Lieuwe de Haan, psychiater en hoogleraar Psychotische Stoornissen, afd. Psychiatrie, Vroege Psychose, UMC locatie AMC, Arkin Amsterdam.

Correspondentieadres

Steven Berendsen (s.berendsen@amsterdamumc.nl).

Geen strijdige belangen meegedeeld.

essentieel is voor een betrouwbaarheidsanalyse. Een geschiktere analyse voor ordinale data is bijvoorbeeld een intraklassecorrelatiecoëfficiënt (ICC), Krippendorffs *alpha of agreement coefficient-2*.

Om dit te illustreren geven wij een fictief voorbeeld in **tabel 1**, gebruikmakend van scoremogelijkheden van de GASS-C-NLR2. Meting 1 vertegenwoordigt de eerste afname van het meetinstrument en meting 2 het tweede moment. Net zoals in de studie van Huisman e.a. worden deze twee metingen gebruikt voor de bepaling van test-hertestbetrouwbaarheid.

Hiervoor hebben wij drie typen analyses uitgevoerd (SPSS, versie 26): Spearmans correlatiecoëfficiënt, percentage overeenkomst en de ICC (*2-way, mixed effects model, absolute overeenkomst*).

Spearmans correlatiecoëfficiënt blijkt 0,927 ($p < 0,001$). Op basis hiervan kunnen we terecht concluderen dat er een sterke correlatie is tussen de twee metingen. Bij visuele inspectie van de data valt echter op dat er geen overeenkomst is tussen de twee metingen (percentage overeenkomst = 0%) en de ICC-score is 0,290. Volgens de gangbare interpretatie duidt de ICC-score op onvoldoende betrouwbaarheid (Koo & Li 2016).

Uit dit voorbeeld blijkt dat een hoge score op Spearmans correlatiecoëfficiënt geen garantie is voor een hoge ICC-score. Metingen kunnen dus ver uit elkaar liggen, terwijl Spearmans correlatiecoëfficiënt hoog is. Dit maakt deze analyse volgens ons minder geschikt voor een test-hertestbetrouwbaarheidsanalyse.

Het belang van betrouwbaarheid is helder en psychometrisch onderzoek van klinisch relevante instrumenten zoals Huisman e.a. presenteren, is erg belangrijk. Het type

betrouwbaarheidsanalyse verdient aandacht, er zijn veel varianten in omloop, elk met voor- en nadelen. Er zijn verschillende overzichtartikelen om de keuze voor de juiste analysemethode te vergemakkelijken (Hallgren 2012).

LITERATUUR

- Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychol* 2012; 8: 23-34.
- Huisman R, Okhuijsen-Pfeifer C, Mulder EYH, Jongkind A, Cohen D, Bogers J, e.a. Glasgow Antipsychotica Bijwerkingen Schaal voor Clozapine; validatie van de Nederlandstalige versie. *Tijdschr Psychiatr* 2021; 63: 270-5.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155-63.
- Portney LG, Watkins MP. *Foundations of clinical research. Applications to practice*. Prentice Hall Health, Upper Saddle River, New Jersey 2000.

Antwoord aan

Berendsen en De Haan

Wij danken Berendsen en De Haan voor hun belangstelling voor ons artikel en hun ingezonden brief en uitleg. Als reactie geven we een uitleg van het proces dat we doorlopen hebben in het onderzoek en presenteren we de data op basis van aanvullende analyses, zoals voorgesteld door Berendsen en De Haan.

Allereerst, het proces tijdens het onderzoek. We hebben inderdaad Spearmans correlatiecoëfficiënt berekend om de associatie tussen de twee meetmomenten van de GASS-C-NL-R2 te berekenen. We hebben overwogen een ICC te berekenen. Dit hebben we echter niet gedaan omdat wij niet op de hoogte waren van de door Berendsen en De Haan gepresenteerde literatuur. In deze artikelen blijkt dat de ICC ook berekend kan worden om de test-hertestbetrouwbaarheid te berekenen; wij kenden de ICC alleen als maat voor interbeoordelaarsbetrouwbaarheid. Dat laatste is in het huidige onderzoek niet van toepassing omdat de GASS-C-NL-R2 een zelfinvulvragenlijst is, waardoor er geen vergelijking gemaakt kan worden tussen beoordelaars.

Na het lezen van de genoemde literatuur komen wij tot de conclusie dat het berekenen van de ICC inderdaad een waardevolle toevoeging is om de test-hertestbetrouwbaarheid van ons onderzoek statistisch vast te stellen. Wij zijn echter niet van mening dat niet-gebruiken ervan onze eerder gerapporteerde resultaten invalideert. Op basis van de richtlijnen van Koo en Li (2016) en de uitleg van Hallgren (2012) hebben we de ICC per item op de twee meetmomenten berekend met model 'two-way mixed' en type 'absolute agreement', zoals briefschrijvers ook doen in hun voorbeeld. De ICC's variëren van 0,511 tot 0,913 en worden weergegeven per item in tabel 1. De gemiddelde ICC over alle items is 0,743.

Daarbij is het belangrijk te bedenken dat de bijwerkingen

AUTEURS

Cynthia Okhuijsen-Pfeifer, psycholoog en promovenda, afd. Psychiatrie, Hersencentrum UMC Utrecht, Universiteit Utrecht.

Jurjen J. Luykx, psychiater-onderzoeker, afd. Psychiatrie en afd. Translationele neurowetenschappen, Hersencentrum UMC Utrecht, Universiteit Utrecht, en opleider bij GGNet, Warnsveld.

Correspondentieadres

Jurjen Luykx (j.luykx@umcutrecht.nl).

gemeten werden met minimaal één week tussen de twee meetmomenten. De mate van bijwerkingen kan verschillen over de tijd en sommige bijwerkingen komen onregelmatig voor, wat deze gerapporteerde ICC's kan doen dalen terwijl dergelijke variatie in rapportage over de tijd eigenlijk niets zegt over de test-hertestbetrouwbaarheid. Daarnaast is de beoordeling van bijwerkingen subjectief en kan worden beïnvloed door stemming en er zou bijvoorbeeld tussen de meetmomenten een verschil in dosering kunnen zijn, wat de beoordeling van bijwerkingen tussen twee meetmomenten verder kan beïnvloeden.

In de praktijk is de GASS-C-NL-R2 bedoeld als momentopname om bijwerkingen te beoordelen. Onze originele bevindingen en de huidige toevoegingen beschouwende, blijven we daarom van mening dat we voldoende aantonen, middels verschillende statistische modellen op verschillende meetmomenten, dat de GASS-C-NL-R2 een waardevol instrument is om bijwerkingen van clozapine te monitoren.

Tabel 1. Intraklassecorrelatiecoëfficiënten (ICC's) per item tussen GASS-C-NL-R2-metmoment 1 en 2, geanalyseerd met model 'two-way mixed' en type 'absolute agreement'

GASS-C-NL-R2-items
1. 0,775
2. 0,669
3. 0,839
4. 0,808
5. 0,800
6. 0,852
7. 0,727
8. 0,767
9a. 0,913
9b. 0,771
10. 0,870
11. 0,582
12. 0,651
13. 0,683
14. 0,866
15. 0,548
16. 0,511
Gemiddelde: 0,743