

DE VERANDERINGS- GEVOELIGHEID VAN OQ-45 EN KKL BIJ ROM

DOOR: Liv Pijck, Mathijs Deen, Julia van den Berg, Irma Huijbrechts & Kees Korrelboom

Eerder onderzoek suggereert dat de voortgang van een behandeling beter is te meten met de OQ-SD dan met de Korte Klachtenlijst. Daardoor lijken instellingen die de KKL gebruiken in een ROM-benchmark slechter te presteren. De KKL blijkt bij nieuw onderzoek in de tweede lijn echter goed bruikbaar.

Behalve dat gegevens van Routine Outcome Monitoring (ROM) in de behandeling worden gebruikt, worden ze landelijk verzameld door Stichting Benchmark GGZ (SBG) om op geaggregeerd niveau uitspraken te doen over de kwaliteit van zorg in de ggz. Het is soms lastig om deze twee doelen van ROM met elkaar te verenigen. Zorgaanbieder PsyQ stelt sinds de start met ROM in 2007 het belang van ROM voor de behandeling boven het belang van de benchmark door SBG.

PsyQ levert ambulante tweedelijnszorg met evidence-based behandelingen in acht specialis- tische behandelprogramma's. ROM vormt hierbij een belangrijke ondersteuning. Vanwege de klantvriendelijkheid heeft PsyQ destijds gekozen voor een korte generieke vragenlijst¹ die ook voor de benchmark kan worden gebruikt, met daarnaast een klachtenspecifieke vragenlijst per behan- delprogramma.

PsyQ gebruikt als generieke lijst de *Korte Klachten Lijst* (KKL).²⁻⁴ Andere instellingen ge- bruiken bijvoorbeeld de *Outcome Questionnaire-45* (OQ-45). De keuze voor de vragenlijst kan de benchmark beïnvloeden, omdat het ene instrument veranderingen beter detecteert dan het andere instrument.

¹ **Huijbrechts IPAM, Appelo M, Korrelboom CW, et al.**

Routine Outcome Measurement binnen PsyQ: normering van de 4 K's. *Dth* 2009; 29: 243-253.

² **Appelo M, Lange A.**

Meten van klachten en veerkracht in vierentwintig vragen. *Dth* 2007; 27: 197-204.

³ **Lange A, Appelo MT.**

De Korte Klachten Lijst (KKL) Handleiding. *Houten: Bohn Stafleu van Loghum; 2007.*

⁴ **Lange A, Schrieken B, Ven JP van de, Blankers M.**

De Korte Klachten Lijst. *Dth* 2000; 20: 384-392.

DE PATIËTENGROEP

Uit een eerder onderzoek van De Beurs en anderen naar de equivalentie van de KKL en een subschaal van de OQ-45, de OQ-SD (Symptomatische Distress), bleek dat de KKL 10 tot 15% minder veranderingsgevoelig was dan de OQ-SD.⁶ Deze resultaten zijn echter gebaseerd op een steekproef van beperkte omvang ($n=101$). De data waren bovendien verzameld bij een eerstelijnspatiëntengroep, die bij de start van de behandeling gemiddeld lager scoort op de KKL dan tweedelijnspatiënten: namelijk 14,3 (SD=6,2).⁶ Afhankelijk van de onderzochte populatie worden in de tweede lijn gemiddelden van 18,2 (SD=9,1)¹ en 21,8 (SD=10,4)³ gevonden. Daardoor kan de door De Beurs en anderen onderzochte patiëntengroep niet als representatief voor de tweedelijnspatiënten beschouwd worden. De lagere beginscore gaf de groep eerstelijnspatiënten minder 'ruimte' voor verbetering dan tweedelijnspatiënten. In dit onderzoek vergeleken wij de veranderingsgevoeligheid van de KKL met de OQ-45 daarom opnieuw bij een grote steekproef van patiënten in de tweedelijns-ggz.

DEELNEMERS EN PROCEDURE

PsyQ verzendt sinds 2009 aan het begin, tijdens en aan het eind van de behandeling digitale ROM-vragenlijsten aan patiënten. De patiënt krijgt een e-mail met een link waarmee via een beveiligde verbinding de vragenlijsten kunnen worden ingevuld. Na afronding zijn de resultaten direct door de behandelaar te zien in het Elektronisch Patiënten Dossier.

Van 1 maart 2012 tot 14 januari 2013 werd de OQ-45 naast de KKL afgenomen bij alle nieuw ingeschreven patiënten van de behandelprogramma's Angststoornissen, Depressie, Psychotrauma en Somatiek & Psyche van de vestigingen Den Haag, Rotterdam en Spijkenisse. Deze programma's geven een goede dwarsdoorsnede van de problematiek bij Volwassenen Kortdurende zorg en maken het leeuwendeel uit van de patiënten die in de benchmark worden gemeten.

Het onderzoek is uitgevoerd op PsyQ-locaties waar de ROM redelijk tot goed was geïmplementeerd, opdat de OQ-45 relatief probleemloos kon worden ingevoegd in het bestaande proces en de extra belasting van de werkvloer tot het minimum beperkt zou blijven. Uiteraard kregen de behandelaars wel uitleg over de OQ-45, zodat zij konden profiteren van de extra informatie die deze opleverde.

De OQ-45 en de KKL zijn bij 624 individuele patiënten minstens tweemaal gelijktijdig afgenomen: eenmaal direct na de aanmelding en nog eens in de loop van de behandeling. Om er zeker van te zijn dat tussen de metingen in ieder geval enige behandeling heeft plaatsgevonden, moest de periode tussen twee metingen minimaal een maand zijn.

MEETINSTRUMENTEN

De Korte Klachten Lijst (KKL) vraagt op een vijfpuntsschaal van 0 ('geen') tot 4 ('heel erg') hoeveel last de respondent heeft van dertien symptomen: angsten, concentratieproblemen, geheugenproblemen, depressiviteit, lichamelijke klachten, prikkelbaarheid, relatieproblemen, zelfmoordgedachten en/of -pogingen, eetstoornissen, automutilatie, seksuele problemen, slaapproblemen en verslavingen. Daarnaast kan de respondent op een open item zelf aanvullend een klacht invullen. De totaalscore wordt berekend door de scores op de eerste dertien vragen op te tellen. De range is 0-52. Een hogere score duidt op meer klachten. Uit psychometrisch onderzoek blijken de interne

5

Lambert MJ, Morton JJ, Hatfield DR, et al.

Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire) (3th ed.)

Wilmington DE: American Professional Credential Service LLC; 2004.

6

Beurs E de, Barendrecht M, Flens G, et al.

Vooruitgang in de behandeling meten – Een vergelijking van vragenlijsten voor zelfrapportage.

Maandblad Geestelijke volksgezondheid 2012; 67: 259-265.

7

Beurs E de, Hollander-Gijsman ME den, Buwalda V, et al.

The Outcome Questionnaire (OQ-45): een meetinstrument voor meer dan alleen psychische klachten.

De Psycholoog 2005; 40: 53-63.

8

Jong K de, Nugter A, Polak MG, et al.
The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation.

Clinical Psychology and Psychotherapy 2007; 14: 288-301.

consistentie (tussen de 0,78 en 0,85) en de test-hertestbetrouwbaarheid (tussen de 0,69 en de 0,77) van de KKL bij verschillende groepen voldoende te zijn.³

De *Outcome Questionnaire* (OQ-45) is speciaal ontwikkeld om verandering van klachten te meten gedurende een behandeling.^{5,7-8} De vragenlijst bestaat uit 45 items die gescoord worden op een vijfpuntsschaal van ‘nooit’ (0) tot ‘bijna altijd’ (4).

Het functioneren van de patiënt wordt in kaart gebracht op drie subschalen, waarvan alleen de eerste is gebruikt voor de vergelijking met de KKL; Symptomen (Symptomatische Distress - SD) waaronder angst, depressie, gebruik van alcohol en/of drugs. Deze wordt ook in de benchmark van SBG gebruikt en komt qua meetpretentie het meest overeen met de KKL. De score wordt berekend door de scores op 25 items op te tellen; de range is 0-100. Een hogere score duidt op meer symptomen.

De betrouwbaarheid en validiteit van de Nederlandse vertaling van de OQ-45 zijn voldoende tot goed.⁷⁻⁸ In de vergelijkbaarheidsstudie van De Beurs en anderen werd de veranderingsgevoeligheid van de OQ-SD als goed beoordeeld.⁶

ANALYSE VAN VERANDERINGSGEVOELIGHEID

Er zijn verschillende manieren om naar veranderingsgevoeligheid te kijken. Omdat onderzoekers het niet eens zijn over wat de beste maat is,⁹ hebben wij ervoor gekozen dezelfde drie maten voor effectgrootte te gebruiken als De Beurs en collega's in 2012.⁶

Cohen's d: de effectgrootte-index van Cohen. Dit is het verschil tussen de gemiddelde score op de voormeting en de gemiddelde score op de nameting, gedeeld door de standaarddeviatie (SD) van de scores op de voormeting.

Standardised Response Mean (SRM): het gestandaardiseerde responsgemiddelde. Dit is het verschil tussen de gemiddelde score op de voormeting en de gemiddelde score op de nameting, gedeeld door de SD van de verschillen tussen voor- en nameting.

ΔT : Het verschil tussen de genormaliseerde T-scores van voor- en nameting. De benchmarkmethode die SBG gebruikt is vergelijking van de genormaliseerde T-scores. Voor de technische uitleg verwijzen we naar De Beurs.¹⁰ Voor dit artikel is van belang om te begrijpen dat ruwe scores op verschillende vragenlijsten moeilijk met elkaar kunnen worden vergeleken. Deze scores zijn afhankelijk van de gemiddelde score en de spreiding van de vragenlijst, en moeten daarom gestandaardiseerd worden. De T-score is een zodanige transformatie van de ruwe score dat ze wel vergelijkbaar worden.

Omdat verschillen alleen valide zijn bij een intervalmeetniveau, zijn de scores bovendien genormaliseerd. De genormaliseerde T-scores werden apart berekend voor de voor- en nametingen; vervolgens werd het verschil tussen de twee berekend.

BOOTSTRAPPING

Om te bepalen of de verschillen tussen de KKL en de OQ-SD op de drie maten statistisch significant zijn, moeten er betrouwbaarheidsintervallen worden berekend. In het huidige onderzoek is gekozen voor een non-parametrische bootstrap.¹¹⁻¹²

De bootstrapprocedure bestaat uit het aselect en met teruglegging trekken van een groot aantal (gewoonlijk 5000) nieuwe steekproeven uit de originele data. Elke nieuwe bootstrapsteekproef

9

Norman GR, Wyrwich KW & Patrick DL.
The mathematical relationship among different forms of responsiveness coefficients.

Quality of Life Research 2007; 16: 815-822.

10

Beurs E de.
De genormaliseerde T-score: Een 'euro' voor testuitslagen.

Maandblad Geestelijke volksgezondheid 2010; 65: 684-695.

11

Efron B, Tibshirani R.
An introduction to the bootstrap.
New York: Chapman & Hall; 1993.

12

Davison AC, Hinkley DV.
Bootstrap methods and their application.
Cambridge, UK: Cambridge University Press; 1997.

bevat evenveel observaties als de originele data, maar sommige observaties komen niet in een bepaalde steekproef voor en andere observaties komen dan vaker voor. Dit laatste is het gevolg van de teruglegging. Op elk van deze steekproeven worden de analyses uitgevoerd waarin men geïnteresseerd is, waardoor er voor elke analyse evenveel uitkomsten als bootstrapsteekproeven zijn. Wanneer deze uitkomsten gerangschikt worden van laag naar hoog, en de 25e en 975e promille worden bepaald, dan ligt het 95% betrouwbaarheidsinterval van de uitkomst tussen deze waardes. Uit de huidige steekproef zijn 5000 bootstrapsamples getrokken. In elk van deze bootstrapsamples werden de drie maten voor veranderingsgevoeligheid berekend en werd vervolgens vastgesteld wat het verschil was tussen de KKL en de OQ-SD (verschil = KKL - [OQ-SD]) voor elke maat. Hierdoor zijn voor elke maat 5000 verschilwaardes gegenereerd. Voor elk van deze verschilwaardes is vervolgens het betrouwbaarheidsinterval bepaald. Als dit betrouwbaarheidsinterval de waarde 0 (nul) omvat, dan is het verschil in veranderingsgevoeligheid tussen de OQ-SD en de KKL niet statistisch significant.

RESULTATEN: KKL EVEN GEVOELIG ALS OQ-SD

Uiteindelijk bleken door 580 patiënten zowel een begin- als een vervolgmeting van de KKL en de OQ-SD gelijktijdig te zijn ingevuld, met minimaal 1 maand tussenruimte. Van de patiëntengroep was 38% man ($n=220$). De gemiddelde leeftijd was 41,8 jaar ($SD=120$). Gemiddeld zat er 4,3 maanden tussen de twee metingen ($min=1, max=10, SD=2,1$). In tabel 1 staat de verdeling over de behandelprogramma's.

Omdat de totaalscore van de KKL schever verdeeld is dan de totaalscore van de OQ-SD, zijn de correlaties tussen beide vragenlijsten op de begin- en eindmeting en de verschilcores berekend door middel van Spearman's rho. De correlatie tussen de beginmetingen is 0,84. Tussen de eindmetingen is de correlatie 0,87 en tussen de verschilcores 0,71.

De range van scores op de vragenlijsten was voor deze steekproef 0-48 op de KKL en 2-94 op de OQ-SD. De gemiddelde scores en verschilcores zijn weergegeven in tabel 2.

(TABEL 1)

VERDELING PATIËNTEN OVER DE BEHANDELPROGRAMMA'S					
BEHANDELPROGRAMMA	N	%	TIJDSVERLOOP TUSSEN TWEE METINGEN		
			M	SD	RANGE
Angststoornissen	100	17,2	3,8	2,01	1-9
Depressie	337	58,1	4,42	2,08	1-10
Psychotrauma	85	14,7	4,36	1,9	1-9
Somatiek & Psyche	58	10,0	4,23	2,06	1-9

(TABEL 2)

VERGELIJKING VAN SCORES OP VOOR- EN NAMETINGEN EN EFFECTGROOTTES								
		VOORMETING		NAMETING		COHEN'S d	SRM	ΔT
		M	SD	M	SD			
KKL	ruwe score	17,7	9,3	13,5	9,3	0,45	0,57	-
	t-score	50,1	9,8	45,3	10,2	-	-	4,7
OQ-SD	ruwe score	48,8	17,3	41,2	18,8	0,44	0,53	-
	t-score	50,0	9,9	45,8	10,5	-	-	4,2

(TABEL 3)

RESULTATEN NON-PARAMETRISCHE BOOTSTRAP			
VERSCHIL KKL-OQ-SD	95% BETROUWBAARHEIDSINTERVAL		
	2,5%		97,5%
COHEN'S d	0,01	-0,043	0,060
SRM	0,04	-0,028	0,100
ΔT	0,5	0,038	0,977

De resultaten van de non-parametrische bootstrap staan in tabel 3. Het 95% betrouwbaarheidsinterval ligt tussen het 25e en het 975e promille van de drie maten. Er is geen verschil gevonden in veranderingsgevoeligheid tussen de OQ-SD en de KKL op de Cohen's d en de SRM. Wel is er een verschil op de ΔT: de KKL scoort hoger dan de OQ, wat inhoudt dat er volgens de ΔT-methode meer verandering wordt gemeten met de KKL dan met de OQ-SD.

DISCUSSIE: KKL BRUIKBAAR IN TWEDE LIJN

Het doel van deze studie was om in een relevante steekproef van patiënten uit de tweedelijns-ggz te onderzoeken of de KKL inderdaad minder veranderingsgevoelig is dan de OQ-SD, zoals is gebleken uit het onderzoek van De Beurs en anderen.⁶ Dit blijkt niet het geval: volgens twee verschillende indices is de KKL even veranderingsgevoelig, en volgens een derde index zelfs veranderingsgevoeliger dan de OQ-SD. Het voldoet dus aan een belangrijke voorwaarde voor gebruik als benchmarkinstrument.

De verschillen tussen deze twee onderzoeken zijn waarschijnlijk te verklaren door de variatie in de steekproef. Het huidige onderzoek omvatte een groter aantal patiënten, die bovendien in

de tweede lijn in plaats van in de eerste lijn werden behandeld. Eerder werd het belang van het normaliseren van de scores beschreven. Doorgaans zal normaliseren weinig invloed hebben op de scores van de OQ-SD, omdat deze bij benadering normaal verdeeld zijn. Normaliseren van de ruwe scores van de KKL heeft meer invloed, omdat deze rechtsscheef verdeeld zijn. Dit kan invloed hebben op de resultaten, al is onduidelijk in welke richting.

Opvallend is dat de effectgroottes in dit onderzoek voor zowel de KKL als de OQ-SD in absolute zin bescheiden zijn. In de reguliere interne rapportages van PsyQ worden hogere effectgroottes gemeld. De meest waarschijnlijk verklaring hiervoor is dat de tweede meting niet altijd aan het eind van de behandeling plaatsvond, maar vaak een tussentijdse meting was. De gemiddelde tijdsduur tussen eerste en tweede meting was immers slechts 4,3 maanden. De meeste behandelingen in de tweede lijn duren langer.

Doordat de resultaten vergeleken worden met die van De Beurs en anderen uit 2012⁶ ontstaat de indruk dat de KKL wellicht veranderingsgevoeliger is in de tweede lijn dan in de eerste lijn. Momenteel maakt de ggz in Nederland een grote verandering door met de invoering van de generalistische basis-ggz en gespecialiseerde ggz. De verwachting is dat patiënten met relatief lichte problematiek nu behandeld zullen worden in de generalistische basis-ggz. De gespecialiseerde ggz zal daardoor gemiddeld genomen een patiëntenpopulatie met een ernstiger klachtenniveau krijgen dan de huidige tweede lijn kent. De KKL zal in de gespecialiseerde ggz mogelijk nog iets veranderingsgevoeliger zijn dan uit het huidige onderzoek naar voren komt. Of dit werkelijk zo is zal uit vervolgonderzoek moeten blijken.

SAMENVATTING

Bij Routine Outcome Monitoring heeft de keuze voor een vragenlijst invloed op de benchmark, omdat het ene instrument veranderingen beter detecteert dan het andere. In eerder onderzoek bleek de KKL minder veranderingsgevoelig dan de OQ-SD. De resultaten waren echter gebaseerd op een bescheiden steekproef uit een eerstelijnspatiëntengroep. Uit nieuw onderzoek met een grote steekproef van tweedelijnspatiënten blijkt de KKL vergelijkbaar veranderingsgevoelig op twee verschillende maten, en op een derde zelfs veranderingsgevoeliger dan de OQ-SD.

L. Pijck, MSc., is onderzoeker en ROM functionaris bij PsyQ.

l.pijck@psyq.nl

M.L. Deen, MSc., statisticus Parnassia Academie en onderzoeker instituut Psychologie, Universiteit Leiden.

Dr. J.F. van den Berg, senior onderzoeker Parnassia en docent Klinische Psychologie, Universiteit Leiden.

Dr. I.P.A.M. Huijbrechts, senior onderzoeker PsyQ Nederland.

Dr. K. Korrelboom is hoofd wetenschappelijk onderzoek en zorginnovatie, PsyQ Haaglanden, en werkzaam bij het departement medische en klinische psychologie, Tilburg University.