

Van meten naar verbeteren: een data-gedreven lerend netwerk in de ambulante ggz

M.J.M. Merkx, B. Dumas, A. Seinen, N. Gramberg, D. Schwegler, V. Karthaus, N. Kakes, K.M.F. Gorgels

- Achtergrond** De kwaliteit van zorg in de ggz staat onder druk door personeelstekorten en een toenemende zorgvraag. Om de zorgkwaliteit te verbeteren, hebben vier ambulante ggz-instellingen in 2022 een lerend netwerk opgericht voor angst-, stemmings-, en trauma- en stressorgerelateerde stoornissen waarbinnen data over behandelprocessen en -resultaten werden gedeeld en systematisch vergeleken. Dit droeg bij aan de identificatie van best practices en genereert inzichten voor het optimaliseren van werkwijzen voor deze cliëntengroep, teneinde de efficiëntie en effectiviteit van behandelingen te vergroten.
- Methode** Een observationele cohortstudie werd uitgevoerd bij vier ggz-instellingen. Data over behandel efficiëntie en behandelresultaten werden verzameld uit het elektronisch patiëntendossier en met routine outcome monitoring (ROM).
- Resultaten** Binnen alle vier de instellingen werden behandel efficiëntie (behandelduur) en -effect (symptomatisch herstel) geanalyseerd. Er was een aanzienlijke praktijkvariatie in behandel efficiëntie. Voor alle instellingen was het behandel effect statistisch significant. De gevonden effectgroottes (Cohens d) varieerden van groot tot zeer groot. Echter, de gevonden effectgroottes varieerden tussen instellingen, hetgeen wees op ruimte voor verdere verbetering.
- Conclusie** Samenwerking van instellingen – in plaats van concurrentie – heeft waardevolle inzichten opgeleverd in efficiëntere en effectievere behandelprocessen door het systematisch vergelijken van data. De verdere verbetering van de ambulante behandeling van angst-, stemmings-, en trauma- en stressorgerelateerde stoornissen zou met name gericht moeten zijn op het verhogen van de behandel efficiëntie. Methoden zoals chatcontact en het continu monitoren van behandelresultaten tijdens de behandeling (within-treatmentrespons) kunnen hierbij een sleutelrol spelen. Een lerend netwerk levert een waardevolle bijdrage aan de verbetering van de kwaliteit van ambulante ggz-zorg.

De Wetenschappelijke Raad voor het Regeringsbeleid benadrukt dat de kwaliteit van de Nederlandse gezondheidszorg hoog is, maar onder druk staat.¹ Kwaliteit van zorg is onder meer gerelateerd aan doeltreffendheid, doelmatigheid en toegankelijkheid.² De aanhoudende personele krapte in de ggz en de (verwachte) toename van zorgvraag dragen bij aan de toenemende druk op de zorgkwaliteit.

Sinds de jaren 70 is het percentage mensen, werkzaam in de gezondheidszorg, meer dan verdubbeld, van minder dan 7% naar ruim 15% van de beroepsbevolking.¹ Over 40 jaar zou één op de drie professionals in de zorg moeten werken om aan de dan verwachte vraag naar zorg te kunnen voldoen, inclusief de ggz. Dit lijkt onhaalbaar door het huidige tekort van 8600 FTE ggz-professionals, oplopend tot 22.100 FTE in 2034.^{3,4} Daarnaast neemt ook het aantal mensen met psychische aandoeningen toe, zowel wereldwijd als in Nederland.⁵ In Nederland steeg de 12-maandsprevalentie van DSM-IV-stoornissen van 17,4% (2007-2009) naar 26,1%

(2019-2022).⁶ Ook bij adolescenten is er een toename van het aantal mentale problemen, vooral angst- en stemmingsstoornissen bij meisjes, en ADHD en autismespectrumstoornissen bij jongens.⁷ Aangezien 75% van de psychische klachten zich manifesteert voor het 24ste levensjaar, zal de vraag naar ggz-hulp blijven stijgen, terwijl de wachtlijsten al zorgwekkend hoog zijn.⁸ De groeiende vraag naar ggz-hulp en het tekort aan professionals drukken de zorgkwaliteit, waardoor niet iedereen tijdig geholpen kan worden. Een mogelijke oplossing is het verhogen van de efficiëntie en de effectiviteit van behandelingen. Dit kan de doorstroom verbeteren, wachttijden verkorten en het aantal heraanmeldingen verminderen, wat leidt tot een beter toegankelijke en duurzamere ambulante ggz-zorg. In dit licht hebben vier ambulante ggz-instellingen een gezamenlijk lerend netwerk gevormd gericht op angst-, stemmings- en trauma- en stressorgerelateerde stoornissen volgens de DSM-5.⁹ Het lerend netwerk, gestart in 2022, heeft als doel de kwaliteit van zorg te verbeteren

door *best practices* te identificeren door het delen en systematisch vergelijken van behandelproces en -resultaten, op basis van data. De verkregen inzichten kunnen gebruikt worden om te leren van elkaars interne processen en de werkwijze eventueel te verbeteren.

In dit artikel beschrijven we een observationele, naturalistische studie naar de behandel efficiëntie en -effecten van de vier ambulante ggz-instellingen die het lerende netwerk vormen. Data zijn verzameld uit het elektronisch patiëntendossier (EPD) en met *routine outcome monitoring* (ROM). We evalueren en bespreken de eerste inzichten en hopen dat dit onderzoek een bescheiden bijdrage kan leveren aan een toekomstbestendigere en toegankelijke ggz.

METHODE

Cliënten en setting

Voor deze observationele studie includeerden we cliënten van 18 jaar en ouder die in 2022 een nieuwe ambulante behandeling startten en beëindigden voor een angst-, stemmings- of trauma- en stressorgerelateerde stoornis (DSM-5). Omdat behandelingen die eind 2022 waren gestart veelal pas in de daaropvolgende jaren werden afgerond, kon afronding van de behandeling ook in 2023 of 2024 hebben plaatsgevonden. Cliënten werden geïnccludeerd indien de ernst van psychische klachten zowel bij aanvang als bij afronding van de behandeling was geregistreerd. Alle behandelingen waren uitsluitend gefactureerd volgens het zorgprestatie model.

De vier deelnemende instellingen behandelen de veelvoorkomende psychische aandoeningen, behalve matige of ernstige stoornissen in het gebruik van een middel. Het behandelportfolio van de deelnemende instellingen bestaat vooral uit cognitieve gedragstherapie (CGT), inclusief de zogenaamde derde generatie CGT en *eye movement desensitization and reprocessing* (EMDR), met name bij posttraumatische stressstoornis. Alle instellingen hadden in 2022 meer dan 100 behandelaren in dienst, van wie circa 30% met een BIG-registratie; de overige 70% waren grotendeels basispsychologen.

Voor de analyses gebruikten we data uit het EPD en uit ROM. Behandel efficiëntie werd geoperationaliseerd als het aantal directe behandelminuten, uitgesplitst naar intake en behandeling. Behandeleffecten baseerden we op ROM-vragenlijsten, afgenomen voorafgaande aan en bij afronding van de behandeling.

Behandelefficiëntie: directe tijd in minuten

Behandelefficiëntie werd geoperationaliseerd als de aan de zorgverzekeraar gedeclareerde directe contacttijd tussen behandelaar en cliënt (behandelkamer of chat), onderscheiden in *intake en diagnostiek* en *behandeling*.

Gemiddeld aantal minuten intake en diagnostiek

Het gemiddelde aantal minuten intake, vallend onder prestatie diagnostiek, per cliënt per instelling werd berekend door de totale directe tijd voor intake en diagnostiek te delen door het aantal cliënten die in 2022 een nieuwe behandeling waren gestart, een intake kregen en de behandeling hadden afgerond.

Gemiddeld aantal minuten behandeling

Het gemiddelde aantal minuten behandeling per cliënt per instelling werd berekend door de totale directe tijd te delen door het aantal cliënten die in 2022 nieuwe behandeling waren gestart, een intake hadden gehad en de behandeling hadden afgerond.

Vragenlijsten

Voor het in kaart brengen van psychische klachten gebruikten we twee zelfrapportagevragenlijsten, namelijk de *Symptom Questionnaire-48* en de *Brief Symptom Inventory*.^{10, 11} Eerstgenoemde vragenlijst werd door drie instellingen gebruikt, laatstgenoemde door één instelling. De SQ-48 is een zelfrapportagevragenlijst die de frequentie van psychische klachten in de afgelopen week meet, met antwoordcategorieën van 0 (nooit) tot 4 (zeer vaak). Deze heeft goede validiteit, hoge interne consistentie ($\alpha = 0,97$) en test-hertestbetrouwbaarheid (ICC = 0,93). De SQ-48 bevat zeven op klachten gerichte subschalen, en daarnaast de subschalen vitaliteit en werk. De totaalscore (0-148) wordt berekend door de scores van de klachtenschalen op te tellen; hogere scores wijzen op meer distress.

Ook de BSI is een zelfrapportagevragenlijst die de ervaren intensiteit van psychische klachten in de afgelopen week meet, met antwoordcategorieën van 0 (helemaal geen) tot 4 (heel veel). De BSI heeft hoge validiteit, interne consistentie ($\alpha = 0,96$) en test-hertestbetrouwbaarheid (ICC = 0,90). De BSI meet de ernst van klachten op negen schalen, waaronder somatische klachten, depressie, angst en psychoticisme. De gemiddelde score van 53 items weerspiegelt de ernst van psychisch disfunctioneren; hogere scores wijzen op meer distress.

T-score bij aanvang van de behandeling

Omdat de ernst van de psychische klachten gemeten werd met twee verschillende vragenlijsten, harmoniseerden we de startscores naar T-scores (normaalverdeling, $M = 50$; $SD = 10$) om vergelijkbaarheid te waarborgen.¹²

Symptomatisch herstel

Symptomatisch herstel werd geoperationaliseerd als de effectgrootte van het verschil tussen de ernst van psychische klachten bij aanvang en bij afronding van de behandeling, berekend met Cohens d . Effectgroottes van 0,20-0,49 worden geïnterpreteerd als klein, waarden van 0,50-0,79 als middelgroot, 0,80-1,20 als groot en waarden boven 1,20 als zeer groot.¹³

Tabel 1. Cliëntkenmerken bij aanvang van een behandeling per behandelinstelling

	A	B	C	D
Gem. leeftijd (in jaren)	37,1 (14,7)	40,8 (13,7)	32,7 (11,6)	34,1 (13,1)
Geslacht (% vrouwen)	72,7	60,1	52,4	72,5
Primaire classificatie DSM-5				
Angststoornissen	998	1325	1661	1305
Stemmingsstoornissen	1315	2448	1006	1174
Trauma- en stressorgerelateerde stoornissen	1181	1312	448	1051
Volledigheid (% (n))				
Angststoornissen	54,6 (n = 545)	82,6 (n = 1096)	53,9 (n = 895)	41,8 (n = 545)
Stemmingsstoornissen	39,0 (n = 513)	80,3 (n = 1958)	50,3 (n = 506)	36,9 (n = 433)
Trauma- en stressorgerelateerde stoornissen	41,9 (n = 494)	81,0 (n = 1063)	46,0 (n = 207)	37,2 (n = 391)
T-score bij start (SD)				
Angststoornissen	47,9 (7,0)	47,1 (6,7)	46,5 (6,8)	47,8 (7,8)
Stemmingsstoornissen	49,3 (7,9)	49,0 (6,7)	47,8 (6,9)	49,9 (7,5)
Trauma- en stressorgerelateerde stoornissen	47,2 (8,0)	47,3 (7,9)	44,8 (7,8)	48,3 (8,6)

Statistische analyses

Verschillen tussen voor- en nameting van de afzonderlijke instellingen toetsten we met een t-toets voor gepaarde steekproeven. Vervolgen werd voor de effectgrootte Cohens d berekend. Elke instelling voerde zelf de analyses uit.

RESULTATEN

Beschrijvende kenmerken

Een overzicht van de beschrijvende kenmerken van patiënten is opgenomen in [tabel 1](#). De gemiddelde leeftijd varieerde per instelling: instelling A: 37,1 jaar, B: 40,8, C: 32,7 en D: 34,1 jaar. In instelling A kwamen stemmingsstoornissen het meest voor (n = 1315), gevolgd door trauma- en stressorgerelateerde (n = 1181) en angststoornissen (n = 998). In B domineerden eveneens stemmingsstoornissen (n = 2488), terwijl angst- (n = 1325) en trauma- en stressorgerelateerde stoornissen (n = 1312) in vergelijkbare mate voorkwamen. Instelling C en D lieten eenzelfde patroon zien: angststoornissen kwamen het meest voor (n = 1661 en n = 1305), gevolgd door depressieve (n = 1006 en n = 1174) en trauma- en stressorgerelateerde stoornissen (n = 448 en n = 1051). De volledigheidspersentages (voor- en nameting) varieerden binnen en tussen instellingen: bij A tussen 39% (stemmingsstoornissen) en 54% (angststoornissen), bij B boven 80%, bij C rond 50% en bij D rond 40%. De ernst van psychische klachten (T-scores) was over het algemeen vergelijkbaar tussen instellingen, met uitzondering van trauma- en stressorgerelateerde stoornissen.

Cliënten van instelling C scoorden hier lager (44,8) dan cliënten van de overige instellingen (47,2-48,3), wat kon duiden op een afwijkend instroomprofiel. Voor stemmingsstoornissen werd in alle instellingen de hoogste ernst gevonden (A: 49,3; B: 49,0; C: 47,8; D: 49,9), hetgeen onderstreepte dat cliënten met deze classificatie doorgaans ernstiger klachten rapporteren bij aanvang.

Behandelefficiënte: behandelminuten

Gemiddeld aantal minuten intake en diagnostiek

Hoewel de doelen van het intakeproces bij de vier instellingen vergelijkbaar waren, namelijk probleemanalyse, classificatie volgens de DSM-5, professionele duiding en gezamenlijk beslissen over behandeling, varieerde de tijd die aan intake en diagnostiek werd besteed. Bij instelling A en D lag de gemiddelde tijdsbesteding hoger (152-167 minuten respectievelijk 162-175 minuten), met een brede spreiding. In instelling B en C waren de tijden lager en consistent: 112-114 minuten (SD's: 17-20) in B en 74-75 minuten (SD's: 4-6) in C. De verschillen tussen enerzijds instelling A en D en anderzijds instelling B en C kwamen in alle classificatiegroepen in vergelijkbare mate naar voren. Dit duidde op meer gestructureerde intakeprocedures bij instelling B en C, waar in principe geen aanvullende diagnostiek plaatsvond, tegenover instelling A en D, waar dit wel mogelijk was.

Gemiddeld aantal minuten behandeling

Er bestond aanzienlijke variatie in het gemiddeld aantal behandelminuten tussen de instellingen. Instelling C besteedde gemiddeld het minste aantal minuten per cliënt (495-518 minuten), terwijl instelling D de hoogste

Tabel 2. Behandelefficiëntie per instelling (gemiddelden (SD))

	A	B	C	D
Minuten intake en diagnostiek				
Angststoornissen	163 (113)	113 (18)	75 (4)	167 (105)
Stemmingsstoornissen	167 (109)	112 (17)	75 (5)	175 (126)
Trauma- en stressorgerelateerde stoornissen	152 (71)	114 (20)	74 (6)	162 (90)
Minuten behandeling				
Angststoornissen	615 (373)	658 (377)	518 (185)	735 (513)
Stemmingsstoornissen	662 (365)	661 (369)	513 (196)	839 (609)
Trauma- en stressorgerelateerde stoornissen	642 (399)	716 (553)	495 (197)	799 (604)

Tabel 3. Behandeleffect

	Startmeting gem. (SD)	Eindmeting gem. (SD)	t	df	Cohens d
Instelling A* (SQ-48)**					
Angststoornissen	67,01 (20,03)	43,56 (21,46)	24,44	544	1,05
Stemmingsstoornissen	70,69 (22,38)	46,32 (24,99)	22,47	512	1,01
Trauma- en stressorgerelateerde stoornissen	64,78 (22,86)	42,00 (23,74)	19,83	493	0,89
Instelling B* (SQ-48)**					
Angststoornissen	64,52 (19,29)	32,16 (19,41)	50,33	1095	1,52
Stemmingsstoornissen	70,28 (18,08)	35,14 (21,43)	67,27	1957	1,52
Trauma- en stressorgerelateerde stoornissen	65,01 (22,50)	31,98 (22,27)	43,89	1062	1,35
Instelling C* (SQ-48)**					
Angststoornissen	62,16 (19,39)	36,76 (19,14)	37,58	895	1,26
Stemmingsstoornissen	66,07 (19,00)	40,67 (21,24)	27,18	505	1,21
Trauma- en stressorgerelateerde stoornissen	59,46 (21,03)	33,19 (18,57)	17,49	206	1,21
Instelling D* (BSI)***					
Angststoornissen	1,23 (0,55)	0,56 (0,42)	29,16	544	1,27
Stemmingsstoornissen	1,40 (0,59)	0,67 (0,54)	24,45	433	1,16
Trauma- en stressorgerelateerde stoornissen	1,24 (0,65)	0,56 (0,52)	23,05	390	1,17

*p < 0,001; **Symptom Questionnaire-48; ***Brief Symptom Inventory.

waarden liet zien (735-839 minuten). Dit betekende dat het gemiddeld aantal behandelminuten per cliënt in instelling D bijna 60% hoger lag dan het gemiddeld aantal behandelminuten per cliënt in instelling C. Instelling A en B lagen dicht bij elkaar: voor angst- en stemmingsstoornissen waren de verschillen minimaal (respectievelijk 615 versus 658 minuten en 662 versus 661 minuten). Alleen bij trauma- en stressorgerelateerde stoornissen was er een

duidelijk verschil, waarbij het gemiddeld aantal behandelminuten per cliënt bij instelling B 74 minuten hoger lag dan bij A, wat overeenkomt met bijna twee extra sessies van 45 minuten. Binnen alle instellingen waren de standaarddeviaties hoog (185-609 minuten), wat wees op grote spreiding in behandelduur en het voorkomen van zowel korte als langdurige behandeltrajecten. **Tabel 2** geeft een overzicht van de behandelefficiëntie per instelling.

Behandeleffect: symptomatisch herstel

In alle instellingen nam de ernst van psychische klachten significant af ($p < 0,001$), met effectgroottes variërend van groot tot zeer groot (zie [tabel 3](#)). In instelling A lagen Cohens d-waarden rond 1,0, in B rond 1,4-1,5 (de sterkste effecten), en in instelling C en D rond 1,2. Daarmee toonden alle instellingen klinisch relevante verbeteringen, met de grootste effecten in instelling B. Samengevat lieten alle instellingen significante en substantiële afnames in klachten zien, met variaties in effectgrootte tussen de instellingen.

DISCUSSIE

In dit artikel presenteren wij de bevindingen van een naturalistische studie naar de behandelefficiëntie en -effecten van vier ambulante ggz-instellingen voor de classificatiegroepen angst-, stemmings- en trauma- en stressorgerelateerde stoornissen. Het doel van deze studie was om onderling inzicht te verkrijgen in behandelprocessen en -effecten, door deze met elkaar te vergelijken. Deze aanpak draagt bij aan het identificeren van best practices en biedt handvatten voor het optimaliseren van behandelstrategieën, met als streven zowel de efficiëntie als de effectiviteit van de vier deelnemende instellingen te vergroten. We observeerden klinisch relevante verbeteringen, maar ook substantiële variatie in behandelefficiëntie, waar in onze optiek met name ruimte ligt voor verbetering.

Variatie in duur intake en diagnostiek

De grote variatie in het aantal minuten voor intake en diagnostiek wordt verklaard door verschillen in het intakeproces per instelling. Ten eerste bieden instelling A en D op indicatie de mogelijkheid om tijdens het intakeproces aanvullende diagnostiek, met name gericht op persoonlijkheidsstoornissen of ADHD, te doen. Ten tweede, hoewel alle instellingen een geüniformeerd intakeproces hebben met vergelijkbare doelen, verschilt de opbouw per instelling wezenlijk. Bij zowel instelling A, B als C bestaat het intakeproces uit twee gesprekken. De gemiddelde tijdsduur is in instelling C het kortst, in instelling B duidelijk langer, en in instelling A het langst, bijna tweemaal zo lang als in instelling C. Het intakeproces van instelling D bestaat uit vier onderdelen, maar is qua tijdsduur vergelijkbaar met dat van instelling A. Deze langere intakeprocedure kan mogelijk samenhangen met verschillen in de populatie die naar deze instellingen wordt verwezen. Onderzoek naar de inhoud van de verwijfsbrieven kan bijdragen aan een beter begrip van deze verschillen tussen instellingen.

Variatie in behandelduur

Ook is er variatie in behandelminuten, gerelateerd aan behandeling, tussen de instellingen. Instelling C gebruikt in vergelijking met de andere instellingen, gemiddeld genomen, minder behandelminuten.

Mogelijk dat dit verschil verklaard kan worden doordat instelling C frequenter contact heeft met een cliënt dan de andere instellingen. Instelling C maakt gebruik van frequent chatcontact tussen een cliënt en een behandelaar als ondersteuning van het veranderproces.

Wetenschappelijk onderzoek bevestigt dat een hogere contactfrequentie bijdraagt aan een efficiëntere behandeling, naast versnelde symptoomreductie,^{14,15} wat mogelijk ook de hoge gevonden effectgrootte verklaart. Een andere mogelijke verklaring voor de gevonden variatie in behandelminuten is dat bij instellingen met een kortere behandelduur mogelijk ook cliënten zijn geïnccludeerd van wie het behandeltraject tegen advies van de behandelaar is afgesloten, maar waarbij wel een eindmeting beschikbaar was. Dit kan duiden op uitval of verschillen in behandelintensiteit. Aangezien geen informatie beschikbaar is over de reden van afsluiting van het behandeltraject, kan dit niet nader worden onderzocht.

Effectgroottes

De effectgroottes waren in alle vier de deelnemende instellingen groot tot zeer groot en klinisch relevant voor de drie classificatiegroepen. Hoewel de gevonden waarden, uitgedrukt in Cohens d, enige variatie lieten zien tussen instellingen en classificaties, bevestigden ze in alle gevallen substantiële verbeteringen.

De gevonden effectgroottes sluiten aan bij resultaten uit andere naturalistische studies, die eveneens grote tot zeer grote effectgroottes rapporteren.¹⁶⁻¹⁸ Een kanttekening is echter dat in dergelijk onderzoek de primaire uitkomstmaat doorgaans gebaseerd is op een klacht-specifieke vragenlijst, terwijl we in onze studie een klachtgeneriek instrument gebruikten. Effectgroottes, gemeten met klachtspecifieke instrumenten, vallen doorgaans hoger uit, aangezien deze doorgaans gevoeliger zijn voor verandering binnen een specifiek domein. Dit impliceert dat de gerapporteerde effectgroottes mogelijk een onderschatting zijn van het daadwerkelijke behandelresultaat, gerelateerd aan de primaire classificatie. Het lerend netwerk heeft de ambitie om in de toekomst klachtspecifieke vragenlijsten systematisch en standaard af te nemen bij zowel de start als de afronding van de behandeling.

Ook al zijn de gevonden effectgroottes vergelijkbaar voor de verschillende instellingen, het is opvallend dat de behandelresultaten van instelling B bij alle drie de classificatiegroepen gemiddeld genomen hoger zijn dan die van de drie andere instellingen. Een mogelijke verklaring is dat instelling B bij elke individuele behandeling de behandelvoortgang monitort met een klachtgenerieke zelfrapportagevragenlijst (SQ-48), een werkwijze vergelijkbaar met *feedback-informed treatment*.¹⁹ In de instelling wordt dit *behandelmanagement* genoemd. De monitoring vindt plaats eens per vijf sessies. Deze aanpak zorgt voor een beter behandelresultaat, vooral bij cliënten die aanvankelijk minder profiteren van de behandeling.^{20,21} Bijkomend positief effect is dat deze

werkwijze ook invloed heeft op de behandel efficiëntie. Deze benadering resulteert in kortere behandeltrajecten met goed behandel effect, met name bij cliënten die snel profiteren van een behandeling.²² Het is niet nodig om trajecten van cliënten met goed resultaat na een paar behandel sessies te verlengen en deze kunnen dus snel afgerond worden.

Sterke punten en beperkingen

Een belangrijk sterk punt van deze studie is het gebruik van gegevens uit de dagelijkse klinische praktijk. Dit biedt een realistisch beeld van de behaalde behandel effecten binnen vier ambulante ggz-instellingen. Daarnaast draagt de relatief grote onderzoeksgroep bij aan de betrouwbaarheid en generaliseerbaarheid van de resultaten. De deelname van vier verschillende instellingen vergroot de externe validiteit van de bevindingen, mede doordat de behandelresultaten grotendeels vergelijkbaar bleken tussen de instellingen. Toch moet bij de interpretatie van de resultaten rekening worden gehouden met mogelijke verschillen tussen instellingen die van invloed kunnen zijn op de behandeluitkomsten, maar niet in kaart gebracht zijn.

Een belangrijke beperking van deze studie is dat in drie van de vier deelnemende instellingen de classificatie volgens de DSM-5 niet is vastgesteld met een gevalideerd (semi)gestructureerd interview, wat de betrouwbaarheid van diagnostiek beperkt. Dit bemoeilijkt de vergelijkbaarheid tussen instellingen, en hierdoor kunnen verschillen in uitkomsten mede samenhangen met de betrouwbaarheid van de gevonden classificaties in plaats van met daadwerkelijke behandelverschillen. Ten tweede varieerde de volledigheid van voor- en nametingen sterk tussen de instellingen. Dit kan hebben geleid tot een vertekening van de behandelresultaten, al blijft de aard en richting van deze mogelijke bias onduidelijk door het ontbreken van een non-responsanalyse. Het systematisch verzamelen van ROM-data blijft een uitdaging. Integratie in het behandelproces, zoals bij instelling B, waar vragenlijsten geautomatiseerd worden verstuurd en samen met de cliënt worden besproken, kan de kwaliteit en duurzaamheid van de dataverzameling versterken.

Ten derde waren er geen follow-upgegevens beschikbaar, waardoor we niet konden vaststellen of de behaalde resultaten op langere termijn behouden blijven. De deelnemende instellingen hebben daarom het voornemen uitgesproken om in de toekomst structureel follow-updata te gaan verzamelen.

Een vierde beperking betreft de naturalistische onderzoekopzet zonder controlegroep. Hierdoor kunnen we niet met zekerheid vaststellen dat de waargenomen verbeteringen uitsluitend het gevolg zijn van de behandeling; andere factoren, zoals spontaan herstel of regressie naar het gemiddelde, kunnen ook van invloed zijn geweest.

Ten slotte zijn alleen cliënten met zowel een voor- als nameting geïnccludeerd, wat kan leiden tot selectiebias,

vooral bij instellingen met een lage datavolledigheid. Gezien de gevonden effectgroottes lijkt de invloed hiervan op de uitkomsten echter beperkt.

Leerervaringen: samen geleerd

De belangrijkste leerervaring was dat samenwerking tussen verschillende instellingen in het begin uitdagend was. Het kostte tijd om elkaar als partners in plaats van concurrenten te zien en onderling vertrouwen op te bouwen – een voorwaarde voor succesvolle kennisdeling binnen het netwerk. Daarnaast bleek het verkrijgen van betrouwbare en consistente data lastig, onder meer door verschillen in definities en meetmethoden. Dit onderstreept het belang van heldere afspraken en systematische afstemming om datagedreven verbetering mogelijk te maken.

Tijdens de analyse stonden we herhaaldelijk stil bij verschillen tussen de twee gebruikte vragenlijsten: de SQ-48 en de BSI. Beide meten psychische klachten, maar er zijn verschillen in antwoordcategorieën (frequentie vs. intensiteit) en subschalen. De BSI bevat subschalen (zoals paranoïde gedachten en psychoticisme) die minder relevant zijn voor de door ons onderzochte cliëntpopulatie vanwege de gehanteerde exclusiecriteria, waardoor de gevonden effectgrootte bij instelling D mogelijk een onderschatting is van de daadwerkelijke effectgroottes. Inmiddels is instelling D overgestapt op de SQ-48, wat in de toekomst de vergelijkbaarheid van resultaten zal vergroten.

Er zijn grote verschillen in volledigheidspercentages tussen de instellingen. Instelling B scoort het hoogst, waarschijnlijk doordat het meten en bespreken van behandelresultaten vast onderdeel is van het behandelproces en het uitsturen van vragenlijsten ondersteund wordt door een geautomatiseerd verzendproces. Naar aanleiding van deze verschillen heeft instelling C het verhogen van de volledigheid tot speerpunt gemaakt. Door gerichte aansturing vanuit het management is het percentage inmiddels gestegen tot boven de 60%. Instelling D is gestart met een project om behandelmanagement, het meten en bespreken van behandelresultaten tijdens de behandeling, structureel in te bedden in de behandelpraktijk. Dit zal niet alleen een bijdrage leveren aan een verhoging van het volledigheidspercentage, maar nog belangrijker is dat dit ook zorgt voor een verhoging van de behandel efficiëntie.

Opvallend is de gevonden effectgrootte bij instelling C hoog is, zeker gezien het lagere aantal gemiddelde behandelminuten. Zoals we al beschreven, is de meest waarschijnlijke verklaring het frequente behandelcontact met chat tussen een behandelaar en een cliënt gedurende het behandelproces in deze instelling ter ondersteuning van het veranderproces. Instelling B is gestart met een pilotproject dat het chatten met een cliënt tussen twee behandel sessies mogelijk maakt. Doelstelling is om dit snel te implementeren, in de hoop 25-30% behandel tijd te besparen. Instelling A is er inmiddels op kleine schaal mee gestart.

De dataverzameling en -analyse zijn door elke instelling uitgevoerd. Dit had nogal wat voeten in de aarde. Sinds begin dit jaar neemt Alliantie Kwaliteit GGZ de analyse van de data over. Dit helpt om het beschikbaar krijgen van de benodigde data te automatiseren, waardoor er sneller resultaten beschikbaar zijn. Ook gaat dit helpen in het kader van een gedeeld begrippen- en definitiekader om kennisuitwisseling en datagedreven verbetering effectief te ondersteunen.

CONCLUSIE

Door instellingen als partners te laten samenwerken (in plaats van als concurrenten), en behandelprocessen en -resultaten systematisch samen te vergelijken, kunnen effectievere en efficiëntere werkwijzen worden geïdentificeerd. Dit geldt in het bijzonder voor het vergroten van de behandel efficiëntie. In onze optiek kan een datagedreven lerend netwerk hierin een waardevolle rol spelen en draagt het bij aan het bestendigen, en mogelijk zelfs verbeteren, van de kwaliteit van de ggz-zorg.

LITERATUUR

- 1 Wetenschappelijke Raad voor het Regeringsbeleid. Kiezen voor houdbare zorg. mensen, middelen en maatschappelijk draagvlak. Den Haag: WRR; 2021.
- 2 Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington: National Academies Press; 2001.
- 3 ABF Research. Prognosemodel Zorg En Welzijn; 2023.
- 4 Ministerie van VWS. Den Haag: Stories Dashboard – Geestelijke Gezondheidszorg (GGZ) – Nederland; 2024.
- 5 Santomauro DF, Mantilla Herrera AM, Shadid J, e.a. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021; 398: 10312.
- 6 Ten Have M, Tuithof M, van Dorsselaer S, e.a. Prevalence and trends of common mental disorders from 2007-2009 to 2019-2022: results from the Netherlands Mental Health Survey and Incidence Studies (NEMESIS), including comparison of prevalence rates before vs. during the COVID-19 pandemic. World Psychiatry 2023; 22: 275-85.
- 7 Cybulski L, Ashcroft DM, Carr MJ, e.a. Temporal trends in annual incidence rates for psychiatric disorders and self-harm among children and adolescents in the UK, 2003-2018. BMC Psychiatry 2021; 21: 229.
- 8 Kessler RC, Amminger GP, Aguilar-Gaxiola S, e.a. Age of onset of mental disorders: A review of recent literature. Curr Opin Psychiatry 2007; 20: 359-64.
- 9 American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Arlington: American Psychiatric Association; 2013.
- 10 Carlier I, Schulte-Van Maaren Y, Wardenaar K, e.a. Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders. Psychiatry Res 2012; 200: 904-10.
- 11 Derogatis LR. BSI Brief Symptom Inventory. Administration, Scoring, and. Procedures Manual (4de ed.) Minneapolis: National Computer Systems; 1993.
- 12 de Beurs E, Boehnke JR, Fried EI. Common measures or common metrics? A plea to harmonize measurement results. Clin Psychol Psychother 2022; 29:1755-67.
- 13 Sawilowsky SS. New effect size rules of thumb. J Mod Appl Stat Methods 2009; 8: 26.
- 14 Cuijpers P, Huibers M, Ebert DD, e.a. How much psychotherapy is needed to treat depression? A meta-regression analysis. J Affect Disord 2013; 149: 1-13.
- 15 Foa EB, McLean CP, Zang Y, e.a. Effect of prolonged exposure therapy delivered over 2 weeks vs 8 weeks vs present-centered therapy on PTSD symptom severity in military personnel: A randomized clinical trial. JAMA 2018; 319: 354-64.
- 16 Hans E, Hiller W. A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. Clin Psychol Rev 2013; 33: 954-64.
- 17 Hans E, Hiller W. Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: a meta-analysis of nonrandomized effectiveness studies. J Consult Clin Psychol 2013; 81: 75-88.

De overige literatuurverwijzingen zijn online te raadplegen.

AUTEURS

Maarten Merkkx, GZ-psycholoog/psychotherapeut, directeur Inhoudelijk Beleid en Onderzoek, Mental Care Group. Honorary Fellow, Faculteit Maatschappij en Gedragswetenschappen, Universiteit van Amsterdam.

Bertien Dumas, arts n.p. en algemeen directeur iPractice.

Aldert Seinen, klinisch psycholoog en bestuurder Forta Volwassenen, Forta.

Nine Gramberg, GZ-psycholoog en directeur zorg iPractice.

Dennis Schwegler, adviseur zorg, Mentaal Beter, Mental Care Group.

Vincent Karthaus, data-analist, Forta.

Niels Kakes, data-analist, HSK, Mental Care Group.

Koen Gorgels, onderzoeker, Coöperatie VGZ.

Correspondentie

Dr. Maarten Merkkx (m.merkx@hsk.nl).

Geen strijdige belangen gemeld.

Het artikel werd voor publicatie geaccepteerd op 20-10-2025.

Citeren

Tijdschr Psychiatr. 2026;68(1):19-25