

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/18691> holds various files of this Leiden University dissertation.

Author: Jong, Kim de

Title: A chance for change : building an outcome monitoring feedback system for outpatient mental health care

Date: 2012-04-17

Dutch summary

S

Doelstelling

Het hoofddoel van dit proefschrift was om een feedbackmodel voor *routine outcome monitoring* (ROM) te ontwikkelen voor ambulante kortdurende psychiatrische en psychotherapeutische behandelingen in Nederland. Om dit doel te bereiken moest een aantal stappen worden gezet. De eerste stap was om de kwaliteit van de Nederlandse vertaling van de Outcome Questionnaire (OQ-45; Lambert et al., 2004) te onderzoeken (Hoofdstuk 2). Deze vragenlijst wordt gebruikt als uitkomstmaat in de onderzoeken in de rest van het proefschrift. De tweede stap was om te bepalen wat gemiddeld gezien het behandelverloop is van patiënten en welke factoren geassocieerd zijn met een verhoogd risico op negatieve behandeluitkomsten (Hoofdstuk 4). Hiervoor werden gegevens verzameld bij drie GGZ-instellingen. Er werd gebruik gemaakt van geavanceerde statistische technieken om negatieve behandeluitkomsten te voorspellen. De derde stap was het onderzoeken van de effectiviteit van het geven van feedback over het behandelverloop. Hiervoor werden twee onderzoeken gedaan. In elk onderzoek was de feedback anders vormgegeven. Eerst werd bekeken hoe groot de steekproeven in de onderzoeken moesten zijn om een effect te kunnen vinden (*power*) (Hoofdstuk 3). In het eerste feedbackonderzoek werd alleen aan behandelaars feedback gegeven. Dit werd vergeleken met geen feedback (Hoofdstuk 5). In het tweede onderzoek werd feedback aan behandelaars vergeleken met feedback aan zowel behandelaars als patiënten en geen feedback (Hoofdstuk 6). Verder werd onderzocht of eigenschappen van behandelaars invloed hebben op de mate waarin behandelaars gebruik maken van de feedback en op de effectiviteit van de feedback (Hoofdstuk 5).

Cross-culturele validatie van de OQ-45

De cross-culturele validatie van de OQ-45 liet zien dat de Amerikaanse en Nederlandse versies van de OQ-45 vergelijkbaar waren in betrouwbaarheid en validiteit, maar er werden verschillen gevonden in de factorstructuur en de normen. Voor de drie-factor structuur van de OQ-45 met de subschalen Symptomatische Distress (SD), Interpersoonlijke Relaties (IR) en Sociale Rol (SR) was geen sterk bewijs in de Amerikaanse versie. In de Nederlandse versie paste deze structuur beter, maar was nog steeds niet voldoende goed. Verdere analyses leidden tot het toevoegen van een extra factor, die aanvullend op de drie oorspronkelijke factoren gebruikt kan worden. Deze factor, Angst en Somatische Distress (ASD) genoemd, is vooral interessant voor het meten van uitkomsten bij patiënten met angst- of psychosomatische stoornissen. Een vergelijking van de Amerikaanse en Nederlandse normen liet zien dat de Nederlandse normale populatie en poliklinische populatie wat lager scoren

dan Amerikaanse populaties. Deze verschillen leidden ertoe dat de grenswaarde voor normaal functioneren bij de Nederlandse populatie (55) acht punten lager ligt dan de Amerikaanse grenswaarde. De sensitiviteit en specificiteit waren vergelijkbaar met die van de originele Amerikaanse versie. De reliable change index (een maat voor statistisch betrouwbare verandering) was gelijk voor de twee versies (14 punten). Een verschil tussen de Amerikaanse en Nederlandse normen was dat er in de Nederlandse populatie sekseverschillen gevonden werden. Mannen hadden meer problemen op de sociale rol schaal, terwijl vrouwen een hoger klachtenniveau hadden op de schaal symptomatische distress en de schaal angst- en somatische distress.

De betrouwbaarheid van de subschalen en totaalscore was voldoende tot goed, met uitzondering van de sociale rol schaal. De gevoeligheid voor verandering was uitstekend en de OQ-45 kon goed onderscheiden tussen de normale en disfunctionele populaties. De concurrente validiteit van de schalen voor symptomatische distress en angst- en somatische distress was goed, maar was minder sterk voor de schalen interpersoonlijke relaties en sociale rol. Samenvattend kan gezegd worden dat de Nederlandse versie van de OQ-45 voldoende tot goede psychometrische kenmerken heeft.

Power in drie-level multilevel modellen met behandelaareffecten

Multilevel analyse, een vorm van regressieanalyse met geneste data waarbij patiënten bijvoorbeeld genest zijn binnen behandelaars, is de laatste jaren steeds populairder geworden als methode om longitudinale data te analyseren. Dit komt mede doordat multilevel analyse zo flexibel om kan gaan met ontbrekende metingen. Tot op heden is er maar beperkt aandacht besteed aan poweranalyse in deze modellen. Hoofdstuk 3 laat de effecten zien van de intraclass correlatie coëfficiënt (ICC), het niveau van randomiseren, de steekproefgrootte, covariaten en drop-out op de *power* om een effect te vinden. Data uit routine outcome monitoring werden gebruikt als basis voor simulatieonderzoek. Er werd uitgegaan van een drie-level model, met behandelaars op het hoogste niveau (level 3), patiënten binnen behandelaars op het middelste niveau (level 2) en metingen binnen patiënten op het laagste niveau (level 1). De resultaten lieten zien dat het effectiever voor de power was om patiënten willekeurig aan condities toe te wijzen (*randomisatie*) dan behandelaars. Het aantal patiënten vergroten was de beste manier om de power te verhogen wanneer de randomisatie op patiëntniveau plaatsvond. Bij randomisatie op behandelaarniveau was het toevoegen van meer behandelaars effectiever. Het vergroten van het aantal metingen per patiënt had weinig effect op de power in beide onderzoeksdesigns, net als het toevoegen van een covariaat. Drop-out had eveneens geen groot effect op de power, hoewel het de power wel enigszins verlaagde, vooral wanneer de drop-out in het begin van het onderzoek geconcentreerd was.

Risicofactoren voor negatieve behandeluitkomsten

Omdat het voorkomen van negatieve uitkomsten een van de belangrijkste doelen van feedback aan behandelaars is, is het nuttig om te weten welke factoren geassocieerd worden met negatieve behandeluitkomsten. In hoofdstuk 4 was het doel om het risico op negatieve uitkomsten aan het einde van de behandeling te voorspellen. Uitkomsten werden op twee manieren bekeken: de snelheid van veranderen met multilevel analyse en het functioneren van de patiënt aan het einde van de behandeling met Classification And Regression Trees (CART). CART is een vorm van regressieanalyse waarbij automatisch gezocht wordt naar interacties tussen de voorspellers en resulteert in een voorspellingsmodel met een boomstructuur.

Er werden gegevens over het behandelverloop verzameld bij drie GGZ-instellingen. Daarbij werd geen feedback gegeven aan behandelaars of patiënten. In totaal scoorde 51% van de patiënten in de steekproef ($n = 1540$) aan het einde van de behandeling vergelijkbaar met de normale populatie en kan derhalve als verbeterd worden beschouwd. In de CART analyse vonden we dat patiënten met relatief lage scores op de OQ-45 voor de start van de behandeling en patiënten met een hoger opleidingsniveau betere kansen hadden op positieve behandeluitkomsten. Een uitgebreider CART model liet zien hoe verwachtingen van patiënten over de uitkomst van de behandeling, scores op de sociale rol schaal, scores op de Global Assessment of Functioning (GAF) en de werkaliantie in de beginfase van de behandeling (Taak schaal) op verschillende manieren met elkaar interacteerden. In de multilevel analyse waren de werkaliantie in de beginfase van de behandeling (Taak of Doel subschaal), comorbiditeit op As I, het hebben van een persoonlijkheidsstoornis als hoofddiagnose en de GAF score voorspellend voor de snelheid van verandering bij patiënten.

Er waren verschillen tussen de CART modellen en de multilevel modellen in hun gevoeligheid om negatieve uitkomsten te detecteren. Het eerste CART model (3 takken) had een hoge sensitiviteit, maar een lage specificiteit, terwijl het multilevel model een hoge specificiteit had maar een lage sensitiviteit. Het uitgebreidere CART model (10 takken) had de beste balans tussen sensitiviteit en specificiteit.

Effect van feedback

Het effect van feedback op de uitkomst van de behandeling werd onderzocht in twee gecontroleerde gerandomiseerde onderzoeken. In het eerste onderzoek (Hoofdstuk 5) was het doel om de effectiviteit van 'eenvoudige' feedback (zonder waarschuwingssignalen of verwachte respons curves) te onderzoeken ten opzichte van geen feedback. In de literatuur zijn de grootste effecten van feedback gevonden bij het gebruik van meer complexe feedback modellen waarbij op basis van een statistisch

model een behandelverloop werd voorspeld voor de patiënt, waar het werkelijke behandelverloop tegen afgezet werd en waarbij de behandelaar een signaal kreeg als de patiënt teveel afweek van het verwachte beloop. De meeste outcome monitoring feedback systemen hebben deze functionaliteit echter niet en de effectiviteit van deze meer 'eenvoudige' systemen is onvoldoende onderzocht, terwijl deze wel op grote schaal gebruikt worden. Patiënten ($n = 413$) werden willekeurig toegewezen aan een controlegroep zonder feedback of de feedback conditie. Er werd verwacht dat de feedback vooral effectief zou zijn voor patiënten die niet goed vooruit gingen in de behandeling, de zogenaamde 'not on track' (NOT) patiënten. Tegen de verwachtingen in was de feedback echter niet effectiever dan geen feedback en er was geen interactie tussen feedback en NOT zijn voor patiënten. Echter, wanneer behandelaars aangaven dat ze de feedback actief gebruikten in de behandeling, had het geven van feedback aan de behandelaar een significant positief effect voor NOT patiënten.

In het tweede onderzoek (Hoofdstuk 6) was het doel om het effect te laten zien van feedback aan zowel patiënt als behandelaar. Patiënten werden willekeurig toegewezen aan drie condities: geen feedback, feedback aan alleen de behandelaar en feedback aan behandelaar en patiënt. Feedback werd gegeven zonder voorspellingsmodel met verwacht behandelverloop, maar behandelaars kregen wel feedback berichten die aangaven of de patiënt verslechterd was of onvoldoende verandering had doorgemaakt. Er werden aparte analyses gedaan voor korte behandelingen (korter dan 35 weken) en langere behandelingen (35 weken of langer). Het effect van feedback was het sterkste bij NOT patiënten in korte behandelingen, wanneer feedback werd gegeven aan zowel de behandelaar als de patiënt. Feedback aan alleen de behandelaar was ook effectief bij NOT patiënten in korte behandelingen, maar in langere behandelingen was alleen feedback aan behandelaar en patiënt effectief. De feedback had vooral effect op de snelheid waarmee patiënten vooruitgingen in de behandeling, maar had geen significant effect op het niveau van functioneren van patiënten aan het einde van de behandeling.

Behandelaar effecten

Kenmerken van de behandelaars en de manier waarop zij feedback gebruiken spelen mogelijk een belangrijke rol in de effectiviteit van feedback. Immers, als behandelaars de feedback niet constructief gebruiken, is het onwaarschijnlijk dat de feedback de uitkomsten verbetert. Er werden diverse kenmerken van de behandelaar die potentieel van invloed zijn op de effectiviteit van de feedback onderzocht (Hoofdstuk 5). Feedback wordt door behandelaars eerder geaccepteerd wanneer het van een betrouwbare bron komt, die persoonlijke relevantie heeft voor de ontvanger van de feedback. Dit concept wordt *perceived validity* genoemd. Een andere factor die van

belang is bij de acceptatie van feedback is de feedback voorkeur. Mensen met een interne feedback voorkeur vertrouwen meer op hun eigen oordeel dan op feedback uit een externe bron, terwijl mensen met een externe feedback voorkeur zelf actief feedback zoeken, ook bijvoorbeeld bij collega's. Vertrouwen in de eigen competenties (*self-efficacy*) is een andere factor die van invloed is op het feedbackproces. Het verwijst naar het vertrouwen dat iemand heeft om een bepaalde taak succesvol uit te voeren. De motivatie om de feedback te gebruiken in de behandeling is ook een belangrijke factor, omdat bekend is dat behandelaars niet altijd gemotiveerd zijn om feedback te gebruiken en deze dan waarschijnlijk niet effectief is.

De resultaten van ons onderzoek lieten zien dat kenmerken van de behandelaar de effectiviteit van feedback modereren. Behandelaars met een hoge interne feedback voorkeur hadden patiënten die minder snel vooruit gingen in de behandeling dan patiënten van behandelaars met een lage interne feedback voorkeur. Behandelaars die voor aanvang van het onderzoek gemotiveerder waren om de feedback in de behandeling te gaan gebruiken, hadden patiënten die sneller vooruit gingen. Beide effecten traden op ongeacht of de behandelaar feedback ontving of niet en lijken te suggereren dat behandelaars met een open houding ten aanzien van feedback snellere vooruitgang boeken bij hun patiënten. Opvallend genoeg ging de verandering langzamer bij patiënten waarvan de behandelaar een hoge motivatie had om de feedback te gebruiken wanneer deze de feedback ook daadwerkelijk ontving. Verder gingen patiënten in de feedback conditie waarvan de behandelaar meer vertrouwen had in de eigen competenties sneller vooruit in de behandeling dan patiënten van behandelaars die minder vertrouwen hadden of dan patiënten in de controle conditie. Er werd geen effect gevonden van een externe feedback voorkeur en *perceived validity*. Behandelaars waren meer geneigd om de feedback daadwerkelijk te gebruiken in de behandeling als ze voorafgaand aan het onderzoek hoger scoorden op de motivatievragenlijst en vrouw waren.

Consequenties voor de praktijk

Routine outcome monitoring en het geven van feedback aan behandelaars is momenteel sterk in opkomst in Nederland, mede onder druk van de zorgverzekeraars. Dit proefschrift laat zien dat feedback een ingewikkeld proces is en dat vele factoren van invloed zijn op de effectiviteit ervan. De resultaten laten zien dat feedback geven over het behandelverloop effectief kan zijn, maar niet onder alle omstandigheden. De resultaten in hoofdstuk 6 suggereren dat feedback aan zowel patiënten als behandelaars het meest effectief is.

Feedback is bij voorkeur direct beschikbaar en vaker meten biedt meer mogelijkheden om negatieve uitkomsten te voorkomen. Het kan echter kostbaar

zijn om bij alle patiënten frequent te meten. De voorspellingsmodellen in Hoofdstuk 4 kunnen een leidraad zijn om te bepalen welke patiënten een verhoogd risico hebben op negatieve uitkomsten en deze kunnen dan intensiever gevolgd worden in de behandeling. Bij andere patiënten kan dan minder vaak gemeten worden. Een alternatief is om bij alle patiënten gedurende de eerste sessies elke sessie te meten en daarna te bepalen hoe vaak verder meten noodzakelijk is.

Onderzoek laat zien dat feedback effectiever is wanneer het specifieker is. In het kader daarvan is het nuttig om de uitkomsten uit te breiden met *clinical support tools* (zie Shimokawa, Lambert & Smart, 2010), die aangeven welke problemen er zijn en welke interventies mogelijk behulpzaam zijn, wanneer patiënten *not on track* zijn. *Clinical support tools* bevatten procesmaten, zoals de werkrelatie met de behandelaar en de motivatie van de patiënt en veel behandelaars geven de voorkeur aan dit type informatie boven uitkomstmetingen. Een ander aandachtspunt is dat niet alle behandelaars de feedback gebruiken en dat deze derhalve dan ook niet effectief kan zijn. Het krijgen van negatieve feedback kan onaangenaam zijn en erin resulteren dat de behandelaar de oorzaak voor de negatieve feedback, of het gebrek aan vooruitgang van de patiënt, buiten zichzelf legt, wat de effectiviteit van de feedback zou kunnen verminderen. Er zijn ook andere factoren die ervoor kunnen zorgen dat een behandelaar de feedback niet gebruikt, zoals geen tijd hebben om de feedback te bekijken of moeite hebben om de feedback te interpreteren.

Wat op individueel niveau geldt, is ook te vertalen naar een geaggregeerd niveau, bijvoorbeeld een organisatie of team. Anders dan bij feedback aan behandelaars gaat feedback op teamniveau vaak over afgesloten behandelingen en heeft het meer tot doel om te leren van de totale uitkomsten in een bepaalde periode of om beleid op te baseren. Het is te verwachten dat ook bij teams feedback het meest effectief is bij '*not on track*' teams, oftewel teams die minder goede resultaten halen dan de standaard, maar het is ingewikkeld te bepalen wat de standaard zou moeten zijn. Benchmarken zou daar een oplossing voor kunnen bieden.

