

# **Mogelijkheden voor therapie-effectonderzoek in de tbs-sector: komen tot een evidence base onder zorgprogramma's**

Edwin de Beurs

Marko Barendregt

Nederlands Instituut voor Forensische Psychiatrie en Psychologie (NIFP)  
Afdeling Onderzoek en Ontwikkeling  
Postbus 13369  
3507 LJ Utrecht

## **Colofon**

Dit onderzoek is uitgevoerd in opdracht van het Wetenschappelijk Onderzoek en Documentatiecentrum van het Ministerie van Justitie. Het onderzoek werd begeleid door:

Dr. Joke Harte (voorzitter)  
Drs. Marleen Nagtegaal  
Dr. Pieter Osinga  
Dr. Nienke Tenneij

©2008, WODC, Ministerie van Justitie, auteursrecht voorbehouden

## ***Samenvatting***

De studie ‘Mogelijkheden voor therapie-effectonderzoek in de tbs-sector’ had twee doelen: (1) het inventariseren van onderzoeksmethoden en designs die in Nederland en in het buitenland worden gebruikt om de effectiviteit van behandelingen te meten (m.a.w. om therapie-effectonderzoek uit te voeren) en (2) inzicht te krijgen in hoeverre deze methoden bruikbaar zijn om mechanismen die ten grondslag liggen aan verschillen in behandelrespons te onderzoeken. De achterliggende vraag was welke methoden van therapie-effectonderzoek toepasbaar en realiseerbaar zijn in de tbs-sector. Effectiviteits- en responsiviteitsonderzoek binnen de tbs-sector zijn nodig om een evidence base aan te kunnen leggen voor zorgprogramma’s in de tbs, met behandelingen waarvoor empirische evidentie bestaat. Zorgprogramma’s kunnen dan een beschrijving omvatten van de optimale match van patiëntkenmerken en evidence-based behandelingen, m.a.w. wat werkt bij wie onder welke omstandigheden.

Dit rapport bestaat uit drie delen. In het eerste deel wordt de onderzoeksvraag nader gespecificeerd en de aanpak beschreven (hoofdstuk 1). Tevens wordt het kader geschetst waarin het therapie-effectonderzoek moet worden uitgevoerd. De tbs-sector wordt kort beschreven en met enige cijfers geïllustreerd (hoofdstuk 2). Uit deze beschrijving komt naar voren dat de tbs-populatie een heterogene populatie is wat betreft delicthistorie en wat betreft psychopathologie. Bovendien is de psychopathologie veelal ernstig en gecompliceerd, wat ondermeer blijkt uit de hoge mate van co-morbiditeit van meerdere stoornissen bij het merendeel van de tbs-gestelden.

In het tweede deel (hoofdstuk 3 en 4) wordt de achtergrond van evidence-based behandelen beschreven. Evidence-based behandelen is een ontwikkeling vanuit de somatische zorg die in de afgelopen jaren de ggz en nu ook de tbs heeft bereikt. Om tot evidence-based behandelen te komen, moeten er drie stappen gezet worden: uitvoeren van wetenschappelijk onderzoek ter verkrijging van bewijs omtrent therapie-effect, het opstellen van richtlijnen op basis van het wetenschappelijk bewijs en het implementeren van die richtlijnen op het niveau van de individuele behandelaar of patiënt. In de tbs-sector moet allereerst aandacht zijn voor het opbouwen van de evidence base middels wetenschappelijk onderzoek. Welke interventies en behandelprogramma’s zijn effectief in de tbs? Bij het beschrijven van de onderzoeksmethoden in dit rapport richten wij ons primair op de evaluatie van specifieke interventies en behandelprogramma’s in de tbs en niet zozeer op de tbs-maatregel in zijn totaliteit. Onderzoek van de tbs-maatregel in zijn totaliteit is lastig en omdat de maatregel uit een black box van onderdelen bestaat (zoals verschillende vormen van interventies en een begeleidingstraject na de behandeling) zal dit type onderzoek weinig bijdragen aan het opbouwen van een evidence base van werkzame (specifieke) interventies.

Een belangrijk aandachtspunt bij therapie-effectonderzoek in de tbs is welke uitkomstmaat gebruikt zou kunnen of moeten worden. Reductie van recidive is het primaire doel van de tbs-behandeling. Voor gecontroleerd effectonderzoek van specifieke interventies binnen de tbs lijkt recidive echter een minder geschikte uitkomstmaat, vanwege de geruime tijdspanne tussen beëindiging van de onderzochte interventie en de bepaling van eventuele recidive en het feit dat recidive een weinig specifieke en sensitieve uitkomstmaat van interventies is, waardoor (te) grote aantallen proefpersonen nodig zijn om een effect te kunnen aantonen. De mate waarin specifieke doelstellingen van de interventie zijn bereikt, zoals afname van bepaalde symptomen of de mate van reductie van recidiverisico gemeten met risicotaxatie-instrumenten, zijn meer geschikt als uitkomstmaten.

In het derde deel van dit rapport worden methoden voor therapie-effectonderzoek beschreven. In hoofdstuk 5 beschrijven we een aantal algemene methodologische eigenschappen van onderzoeksdesigns voor therapie-effectonderzoek. We staan stil bij interne en externe validiteit van effectonderzoek, wat samenhangt met het aantonen dat een interventie *werkzaam* is (Werkt behandeling Y bij problematiek X beter dan niets?) en dat een interventie *effectief* is in de klinische praktijk (Werkt behandeling Y ook onder minder gecontroleerde omstandigheden buiten de academische setting en is de interventie afdoende?). Onderzoekdesigns onderscheiden zich in welke mate zij inzicht kunnen verschaffen in de factoren die van invloed zijn op het al dan niet slagen van een therapie bij een individu, de zogenaamde responsiviteitsfactoren. Naast deze eigenschappen van een design stellen verschillende onderzoeksdesigns uiteenlopende eisen aan de wijze waarop het onderzoek in de praktijk moet worden opgezet en uitgevoerd. Een aantal van deze eisen (waaronder randomisatie, het geven van een placebobehandeling en statistische eisen) wordt geïntroduceerd. Deze eisen kunnen een specifiek onderzoeksdesign meer of minder uitvoerbaar maken binnen het tbs-kader en dienen daarom meegewogen te worden bij de beoordeling van de geschiktheid van een design voor toepassing in de tbs.

Drie soorten van effectonderzoek worden besproken in hoofdstuk 6: klinisch onderzoek met Randomised Controlled Trials (RCT's), single case designs en outcomes onderzoek. RCT's zijn de gouden standaard voor therapie-effectonderzoek en de voorkeursmethode om de werkzaamheid (d.w.z. de uitkomst ten opzichte van een niet behandelde vergelijkingsgroep) van interventies te onderzoeken. Bij een RCT in zijn simpelste vorm worden patiënten random (willekeurig) toebedeeld aan een experimentele behandelconditie die naar verwachting een werkzaam onderdeel bevat of een controleconditie zonder werkzaam bestanddeel (placebo) of de gebruikelijke behandeling (treatment as usual). Groepsgewijze vergelijking wijst dan uit of de experimentele behandeling superieur is aan geen behandeling of de controlebehandeling. Het RCT-design is een waarlijk experimenteel design en biedt de onderzoeker veel controlemogelijkheden, zoals strenge in- en exclusiecriteria voor deelnemers en toediening van behandeling onder gecontroleerde omstandigheden. De interne validiteit van dergelijk onderzoek is dan ook hoog. De uitgevoerde controle op patiënteninstroom en behandeling gaat echter vaak ten koste van de externe validiteit van het onderzoek: de prototypische patiënt en de zorgvuldig volgens protocol uitgevoerde behandeling is niet kenmerkend voor de alledaagse klinische praktijk. Single case designs zijn experimentele studies waarbij veranderingen binnen één proefpersoon worden onderzocht. Single case designs zijn ook waarlijk experimentele designs, in die zin dat het effect van wel versus geen behandeling binnen een persoon onderzocht wordt. Deze designs zijn vooral geschikt om snel een indruk te krijgen van de werkzaamheid van een interventie. Echter, de generaliseerbaarheid van de bevindingen is beperkt totdat de interventie bij een representatieve groep deelnemers is geëvalueerd. Daarnaast is toepassing van dit design slechts mogelijk voor een beperkt aantal interventies, namelijk interventies waarbij interventie en non-interventieperiodes met elkaar afgewisseld kunnen worden. RCT's en single case designs lenen zich beide minder voor onderzoek naar responsiviteitsfactoren, d.w.z. onderzoek dat zich richt op de vraag wie baat heeft bij bepaalde therapeutische interventies in de tbs en wie niet, en wat de optimale omstandigheden zijn waaronder een interventie werkzaam is. Deze designs lenen zich hier minder goed voor, omdat variatie in responsiviteitsvariabelen juist zoveel mogelijk onder controle van de onderzoeker gebracht wordt vanwege het belang van de interne validiteit van het onderzoek.

Onderzoek gericht op het vaststellen van de effectiviteit van behandelingen in de klinische praktijk staat bekend als “outcomes” onderzoek. Dit houdt in dat er routinematig gegevens worden verzameld over de behandeling die wordt uitgevoerd en dat functioneren en psychische gezondheid van een patiënt regelmatig in kaart worden gebracht. Hiertoe worden op gezette tijden gestandaardiseerde meetinstrumenten toegepast die gevoelig zijn voor verandering. Dit resulteert uiteindelijk in zicht op de effectiviteit van interventies die in de klinische praktijk van de tbs worden uitgevoerd. Outcomes onderzoek is quasi-experimenteel en de resultaten hebben dus minder bewijskracht vergeleken met waarlijk experimentele designs, zoals RCT's en single case designs. De infrastructuur die nodig is voor outcomes onderzoek, kan overigens ook ingezet worden voor het evalueren van nieuwe experimentele behandelingen, al dan niet in een gecontroleerd onderzoeksdesign (RCT's). Outcomes onderzoek leent zich, i.t.t. RCT's en single case designs, bij uitstek voor onderzoek naar responsiviteitsvariabelen, zoals welke persoons- of delictkenmerken en welke behandelingsomstandigheden samenhangen met een gunstige uitkomst. Dit komt omdat variatie in responsiviteitsvariabelen wordt toegestaan in dit design.

De drie onderzoeksdesigns worden onderling vergeleken op methodologische aspecten én realiseerbaarheid in tbs-verband. Geconcludeerd wordt dat, hoewel RCT's de aangewezen methoden zijn voor het aantonen van de werkzaamheid van een behandeling, deze moeilijk realiseerbaar zijn binnen tbs-verband. Ondermeer de eisen van randomisatie en het eventueel onthouden van behandeling aan de controlegroep spelen hierbij een rol. Ook de omvang van een groep homogene patiënten die nodig is voor een RCT is dusdanig dat het aantal RCT's dat binnen de relatief kleine tbs-sector opgezet kan worden beperkt zal zijn. Outcomes onderzoek is minder geschikt voor het aantonen van de werkzaamheid van behandelingen dan RCT's (interne validiteit). Met dit type onderzoek is echter wel goed de effectiviteit van interventies in de praktijk aan de tonen (externe validiteit) en outcomes onderzoek is makkelijker realiseerbaar binnen het tbs-kader dan RCT's. Single case designs hebben met RCT's gemeen dat deze sterk kunnen zijn in het aantonen van werkzaamheid. Vanzelfsprekend zijn single case designs naar verhouding gemakkelijker te implementeren binnen een tbs-behandeling dan RCT's. Een beperking aan single case designs is dat ze niet bij elke interventie bruikbaar zijn. In hoeverre bevindingen uit een single case studie ook te generaliseren zijn naar andere patiënten wordt pas bekend na een ruime serie replicaties.

Onderzoeksinspanningen in de tbs zouden zich moeten richten op het genereren van een evidence base onder behandelingen. Wat werkt voor wie? Waar mogelijk moeten belemmeringen voor onderzoek worden geslecht. In het laatste hoofdstuk worden randvoorwaarden opgesteld voor therapie-effectonderzoek in de tbs en wordt een aantal aanbevelingen gedaan om een evidence base onder de tbs-behandeling tot stand te brengen. Er wordt gepleit voor het opzetten van een infrastructuur voor outcome monitoring (de praktische implementatie van outcomes onderzoek) waarbinnen tevens de uitvoering van kleinschalige RCT's en single case studies mogelijk is. Deze infrastructuur behelst monitoring van de aard van de behandeling en een set van therapie-uitkomstmaten. Onderzoek van therapie-effect vergt een forse inspanning en valide resultaten worden pas bereikt bij voldoende grote aantallen patiënten. Afgezien van single case designs, ontstijgen beide andere onderzoeksdesigns de mogelijkheden van een enkele tbs-instelling. Het is daarom zaak samenwerking op onderzoeksgebied tussen instellingen te stimuleren, bijvoorbeeld door een gemeenschappelijke set van meetinstrumenten voor outcome monitoring na te streven. Een andere manier van samenwerking is onderzoek naar de

werkzaamheid van interventies in de tbs vooral middels multi-center trials te organiseren. Onderlinge samenwerking zou als voorwaarde gesteld kunnen worden voor financiering van dit soort onderzoek.

# Inhoud

Deel 1 Onderzoeksvragen en kader .....	9
1. Onderzoeksvraag en aanpak .....	11
1.1 Probleemstelling .....	11
1.2 Onderzoeksdoelen .....	11
1.3 Opbouw van dit rapport .....	11
1.4 Verantwoording .....	12
2. Kader: de tbs in cijfers .....	13
2.1 Instellingen .....	13
2.2 Instroom en uitstroom .....	13
2.3 Subtypen tbs-gestelden .....	13
2.4 Persoonsgegevens .....	13
2.5 Delicten .....	14
2.6 Psychopathologie .....	14
2.7 Conclusie .....	14
Deel 2 Evidence-based behandelen in de medische en ggz en in de tbs .....	15
3. Evidence-based behandelen .....	17
3.1 Wat is evidence-based behandelen? .....	17
3.2 Evidence-based behandelen in de tbs .....	19
4. De reikwijdte van de evidence base .....	21
4.1 Het effect van de tbs-maatregel vs. het effect van specifieke tbs-interventies en behandelmodules .....	21
4.2 De uitkomstvariabele .....	22
4.3 Conclusie .....	25
Deel 3 Methoden van onderzoek voor een evidence base in de tbs .....	27
5. Algemene methodologische eigenschappen therapie-effectonderzoek .....	29
5.1 Methodologisch relevante aspecten .....	29
5.2 Eisen van designs .....	33
6. Specifieke onderzoeksdesigns voor therapie-effectonderzoek .....	37
6.1. Randomised Controlled Trials (RCT's) .....	37
6.2 Outcomes onderzoek .....	42
6.3 Single case designs .....	47
6.4 Samenvatting en conclusie .....	53
7. Conclusie en aanbevelingen .....	57
7.1 Outcomes onderzoek .....	57
7.2 Kleinschalige RCT's en Single Case Designs .....	57
7.3 Systematische monitoring .....	58
7.4 Samenwerking .....	59
Literatuur .....	67





## Deel 1 Onderzoeksvragen en kader



## **1. Onderzoeksvraag en aanpak**

### **1.1 Probleemstelling**

Er is onvoldoende kennis over welke behandelingen werkzaam en effectief zijn in de tbs-sector. Bovendien zijn de bestaande behandelingen veelal niet evidence-based. Er is een beleidsvoornemen van het Ministerie van Justitie om in 2008 middelen ter beschikking te stellen ter financiering van onderzoek naar de effectiviteit van behandelingen in de tbs, met als uiteindelijk doel een evidence-based zorgprogrammering. Alvorens dit effectiviteitsonderzoek ter hand te nemen, zal de vraag moeten worden beantwoord hoe dergelijk onderzoek zou kunnen worden uitgevoerd binnen de tbs-sector.

### **1.2 Onderzoekdoelen**

Het eerste doel van deze studie is het inventariseren van onderzoeksmethoden en designs die in Nederland en in het buitenland worden gebruikt om de effectiviteit van behandelingen te meten (m.a.w. om therapie-effectonderzoek uit te voeren). Een tweede doel is inzicht te krijgen in hoeverre deze methoden bruikbaar zijn om onderzoek te doen naar mechanismen die ten grondslag liggen aan verschillen in behandelrespons. De achterliggende vraag is welke methoden van therapie-effectonderzoek toepasbaar en realiseerbaar zijn in de tbs-sector.

### **1.3 Opbouw van dit rapport**

Het onderzoek is uitgevoerd middels literatuurstudie aangevuld met een aantal vraaggesprekken met onderzoekers uit de tbs-sector en de universiteiten. De vraag naar onderzoek van behandelingseffectiviteit wordt uitgesplitst in een aantal subonderwerpen die in dit rapport in verschillende delen worden behandeld. Naast het eerste deel, dat de onderzoeksvragen en het kader schept, staan we in het tweede deel van dit rapport stil bij de vraag wat evidence-based behandelen is, hoe dit vorm kan krijgen en wat het impliceert voor de tbs. Daarmee wordt de context geschetst waarin therapie-effectonderzoek zal plaatsvinden.

Het derde deel van dit rapport gaat in op de vraag welke onderzoeksdesigns er zijn, welke is opgesplitst in twee delen. Ten eerste de vraag hoe de *werkzaamheid* van behandeling (d.w.z. of een behandeling effect heeft of niet) is vast te stellen. Aangebouwde werkzaamheid vormt de basis van evidence-based behandelen. Een tweede vraag is die van de *klinische effectiviteit* van behandelingen, hetgeen betrekking heeft op het effect van een behandeling onder minder gecontroleerde omstandigheden. Dit staat in de literatuur ook wel bekend als outcomes onderzoek. Deze twee vormen van therapie-effectonderzoek worden beschreven en hun sterke en zwakke kanten worden geïnventariseerd. Methodologische vereisten die beide onderzoeksvormen met zich meebrengen komen aan bod en worden geïllustreerd met voorbeeldonderzoek uit de literatuur. De toepasbaarheid binnen het tbs-kader en de eisen waaraan voorgesteld onderzoek minimaal dient te voldoen (bijvoorbeeld ten aanzien van power, betrouwbaarheid en sensitiviteit van meetinstrumenten, alsook interne en externe validiteit van het onderzoeksdesign) worden beschreven. Ten slotte komt de praktische vraag aan de orde hoe we in de tbs therapie-effectonderzoek het beste kunnen uitvoeren. Het gaat daarbij zowel om de vraag welke onderzoeksdesigns het meest voor de hand liggen als aan welke randvoorwaarden voldaan moet worden.

Zoals eerder vermeld, zijn een aantal onderzoekers in het veld benaderd en geïnterviewd ten einde goed inzicht te verkrijgen in de huidige stand van zaken met

betrekking tot therapie-effectonderzoek, beschikbaarheid van zorgprogramma's, protocollen, meetinstrumenten en onderzoeksinfrastructuur.

#### **1.4 Verantwoording**

Het onderhavige rapport kwam tot stand op grond van een onderzoek dat werd uitgevoerd in opdracht van het WODC. De opdracht luidde om in twee maanden een beknopte inventarisatie te plegen van geschikte onderzoeksdesigns voor therapie-effectonderzoek en responsiviteitsonderzoek in de tbs-sector. Gezien de beperkte tijdsduur die ons ter beschikking stond moet dit rapport niet beschouwd worden als het definitieve overzicht van wat beschikbaar is op dit gebied, doch als het resultaat van een eerste exploratie. Het beoogt geen volledig inzicht te bieden van het complexe tbs-veld, noch van de (internationale) literatuur op het omvangrijke terrein van therapie-effectonderzoek.

Naast bestudering van de onderzoeksliteratuur werd een aantal betrokkenen uit het tbs-veld geïnterviewd. De groep geïnterviewden is klein vanwege de beperkt beschikbare tijd en deze selectie is vanzelfsprekend niet bedoeld als representatief voor het gehele tbs-veld. De volgende personen werden geïnterviewd: Dr. Eddie Brand, Dr. Erik Bulten, Dr. Klaus Drieschner, Prof. Dr. Paul M.G. Emmelkamp, Prof. Dr. Henk Nijman, Drs. Wineke Smid en Dr. Vivienne de Vogel. We zeggen hen hier dank toe voor hun medewerking.

Tot slot is een woord van dank op zijn plaats aan de begeleidingscommissie, onder voorzitterschap van Dr. Joke Harte van de Vrije Universiteit Amsterdam, voor de kritische maar opbouwende kanttekeningen die de leden bij de conceptversies van deze rapportage hebben geplaatst.

## **2. Kader: de tbs in cijfers<sup>1</sup>**

### **2.1 Instellingen**

In Nederland zijn 12 klinieken door het Ministerie van Justitie aangewezen voor de verpleging van tbs-patiënten, namelijk: de Oostvaarderskliniek te Amsterdam en Utrecht, FPC Veldzicht te Balkbrug, de Dr. S. van Mesdagkliniek te Groningen, de Pompestichting te Nijmegen, de Kijvelanden te Poortugaal, FPC Oldenkotte te Rekken, de Van der Hoeven Stichting te Utrecht, FPI De Rooyse Wissel te Venray, Stichting Hoeve Boschoord te Boschoord, GGZ Eindhoven en de Kempen, Divisie forensische psychiatrie GGZ Drente te Assen en FPK AMC de Meren te Amsterdam. Naast intramuraal kunnen tbs-gestelden ook poliklinisch behandeld of nabehandeld worden bij een aantal instellingen.

### **2.2 Instroom en uitstroom**

Het aantal opleggingen van een tbs-maatregel schommelt tussen de 150 en 200 per jaar. In 2006 is de tbs-maatregel 171 keer opgelegd (<http://www.dji.nl>). De verblijfsduur van een tbs-gestelde was in 2006 gemiddeld 89 maanden.

Het aantal personen dat de tbs verlaat is relatief gering, zo rond de 115 per jaar (111 in 2006, waarvan een kwart door overlijden of door afsluiting van een in duur beperkte tbs-maatregel). Nog altijd is de instroom aanzienlijk groter dan de uitstroom en deze discrepantie verklaart de forse groei van intramuraal verblijvende tbs-gestelden van 681 tbs-gestelden in 1995 tot 1494 personen in 2006 ([www.dji.nl](http://www.dji.nl)). Op dit moment (november 2007) zijn er 1785 intramurale tbs-gestelden.

### **2.3 Subtypen tbs-gestelden**

Men kan verschillende indelingen maken van tbs-gestelden, bijvoorbeeld naar de aard van de psychopathologie of naar het type delict waarvoor de persoon is veroordeeld. Voor onderzoek naar de effectiviteit van interventies zal de indeling naar aard van de psychopathologie veelal de meest relevante zijn. In de onderzoeksliteratuur wordt bijvoorbeeld gerapporteerd over de effectiviteit van een behandeling voor psychotische stoornissen of de effectiviteit van een bepaalde therapievorm bij persoonlijkheidsstoornissen bij forensische patiënten. Onderzoek naar de behandeling van specifieke dadergroepen komt vooral bij seksueel delinquenten voor. In de tbs-sector worden zorgprogramma's samengesteld voor psychotische patiënten, plegers van zedendelicten, plegers van gewelddadige delicten en patiënten in de longstay.

### **2.4 Persoonsgegevens**

Bijna alle tbs-gestelden zijn man (95%). Het grootste deel van de tbs-gestelden is van autochtone afkomst. Echter, allochtonen zijn met 31% van de tbs-populatie tegen 18% in de bevolking in 2000 relatief oververtegenwoordigd. De grootste groep van hen heeft een Antilliaanse of Surinaamse achtergrond (15%) gevolgd door een Turkse, Marokkaanse of Mediterrane achtergrond (7%). Er zijn opmerkelijke verschillen tussen de autochtone en allochtone subpopulaties van tbs-gestelden. Zo komen psychotische stoornissen meer voor onder allochtone (43%) dan onder autochtone (16%) tbs-gestelden. In 2006 is de proportie vrouwen toegenomen tot 6% en de proportie allochtonen (gemeten naar geboorteland) stabiel gebleven op zo'n 30% (<http://www.dji.nl>).

---

<sup>1</sup> Het cijfermatige deel van dit hoofdstuk is gebaseerd op van Emmerik (2001), tenzij anders vermeld

## 2.5 Delicten

Met betrekking tot delictachtergrond wijkt de tbs-populatie aanmerkelijk af van de gedetineerdenpopulatie. Van de tbs-gestelden is 98% veroordeeld voor een geweldsdelict, eventueel in combinatie met een seksueel delict of een vermogensdelict. Van de gedetineerdenpopulatie is een derde deel veroordeeld voor geweldsdelicten, 30% voor vermogensdelicten en 15% voor overtreding van de opiumwet. De resterende 22% is veroordeeld voor een variëteit aan delicten.

Bij iets meer dan de helft van de tbs-patiënten was het slachtoffer van het delict dat zij pleegden een bekende van de dader. Het geweld waarmee het delict gepaard ging is bij tbs-gestelden overwegend zeer ernstig: bij 80% van de tbs-gestelden was sprake van letsel en bij een kwart van de gevallen zelfs een dodelijke afloop. Bij 28% van de tbs-gestelden was het delict tevens een seksueel delict, 12% met een slachtoffer jonger dan 16 jaar (40% van alle seksuele delinquenten) en 16% met een slachtoffer van 16 jaar of ouder. In 7% van de seksuele delicten was sprake van een dodelijke afloop.

## 2.6 Psychopathologie

Van de tbs-populatie heeft 25% een psychotische stoornis (17% schizofrenie) en 43% een stoornis gerelateerd aan middelengebruik. Er was sprake van intensief gebruik van middelen of verslaving ten tijde van het delict bij 65% van de populatie, vooral alcohol (41%), softdrugs (31%) en harddrugs (27%) en in mindere mate gokken (7%) en medicijnen (6%).

De overgrote meerderheid van de tbs-gestelden heeft een persoonlijkheidsstoornis (80%), waarvan de helft de diagnose persoonlijkheidsstoornis “Niet Anders Omschreven” (NAO) heeft gekregen. Bij de andere helft gaat het vooral om cluster B stoornissen (antisociaal 18%, borderline 9%, narcistisch 5% en theatraal 0,5%). Co-morbiditeit van DSM As-I en AS-II stoornissen is eerder regel dan uitzondering: bij 13% is sprake van een enkelvoudige As-I stoornis, bij 27% van een enkelvoudige persoonlijkheidsstoornis en bij 60% is sprake van een comorbide As-I en As-II stoornis.

## 2.7 Conclusie

Uit de cijfermatige beschrijving van de tbs-gestelden komt een beeld naar voren van zeer ernstige en heterogene psychiatrische problematiek. Er is sprake van een hoge co-morbiditeit en een groot percentage persoonlijkheidspathologie, dat zich niet gemakkelijk laat veranderen met therapeutische interventies.

Als we de cijfers bezien, zijn de meest voorkomende vormen van gedrag en psychopathologie die in aanmerking komen voor interventies in de tbs agressief en gewelddadig gedrag, persoonlijkheidspathologie (in totaal 80% van de tbs-gestelden, waarvan slechts de helft een goed omschreven diagnose heeft), middelenmisbruik en verslaving (65%), psychotische stoornissen (25%) en seksuele delinquentie (30%). Deze indeling sluit deels aan bij de ontwikkeling van de zorgprogramma's voor de tbs-sector (voor psychotische stoornissen, zeden- en geweldsdelicten en voor longstay tbs-gestelden) met uitzondering van de verslaving.

## **Deel 2 Evidence-based behandelen in de medische en ggz en in de tbs**





### ***3. Evidence-based behandelen***

#### **3.1 Wat is evidence-based behandelen?**

Net als elders in de geestelijke gezondheidszorg is in de tbs-sector een ontwikkeling gaande om de behandeling meer te organiseren aan de hand van zorgprogramma's voor een aantal beschreven doelgroepen van patiënten. Die zorgprogramma's bieden idealiter niet alleen informatie over welke behandelingen het predicaat "evidence-based" verdienen bij een bepaalde doelgroep, maar ook richtlijnen voor "wat werkt bij wie". Een eerste stap zal worden gezet door de Erkenningscommissie Gedragsinterventies Justitie, onder voorzitterschap van Prof. Dr. W. van Tilburg. Deze commissie richt zich nu al op de beoordeling van justitiële interventies voor jeugdigen en volwassenen en zal zich in de toekomst ook richten op interventies in de tbs-sector.

Evidence-based behandelen staat in de tbs-sector op dit moment nog in de kinderschoenen. Dit heeft ten dele te maken met het gegeven dat er vooralsnog slechts beperkte kennis is over welke behandelingen en interventies werkzaam en effectief zijn bij tbs-gestelden. Daarmee ontbreekt de evidence base voor de behandeling nog grotendeels. We schetsen hier kort de ontwikkeling van evidence-based medicine in de geneeskunde en in het bijzonder in de geestelijke gezondheidszorg.

##### *3.1.1 Evidence-based medicine*

Aan de basis van evidence-based behandelen staat empirische kennis verkregen uit wetenschappelijk onderzoek over wat werkt bij wie en onder welke omstandigheden. De tegenpool van evidence-based behandelen is expert-based behandelen, waarbij niet primair wetenschappelijke kennis, maar inzichten en ervaringen van de behandelaar zelf de maatstaf vormen voor de keuzes in de behandeling. In plaats van evidence-based behandelen wordt ook wel gesproken over "research-based" behandelen. Deze term verdient eigenlijk de voorkeur, want het geeft beter weer dat het onderzoeken van nieuwe behandelingen niet ophoudt na het verzamelen van het eerste bewijsmateriaal ten aanzien van de effectiviteit van een behandeling, maar een continu proces inhoudt van wetenschappelijk toetsen. In de onderzoeksliteratuur heeft de term evidence-based ingang gevonden en wij sluiten ons daarbij aan.

"Evidence-based medicine" komt voort uit een ontwikkeling in de medische wetenschap die al werd ingezet in de jaren '80 van de vorige eeuw. De term werd voor het eerst gebruikt door Eddy (1990). Het Centre for Evidence-Based Medicine definieert evidence-based medicine als volgt:

"Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett et al., 1996).

Er wordt bij de toepassing van evidence-based medicine onderscheid gemaakt tussen (1) Evidence-Based Guidelines, de richtlijnen voor behandeling die voor allerlei stoornissen worden ontwikkeld, gebaseerd op resultaten van wetenschappelijk effectonderzoek, en (2) Evidence-Based Individual Decision Making (Eddy, 1990). Dit laatste houdt in dat een interventie, ingreep of behandeling bij een individuele patiënt wordt verricht volgens bovenstaande richtlijnen of protocollen.

### 3.1.2 *Cochrane Collaboration en de Campbell Collaboration Crime and Justice Group*

Om te komen tot individueel evidence-based behandelen is het noodzakelijk dat er op wetenschappelijk bewijs gebaseerde richtlijnen worden opgesteld. Voor deze richtlijnen is het essentieel dat onderzoeksresultaten, die de evidence base vormen, goed worden gedocumenteerd. In de medische wetenschap wordt de kennis onder meer verzameld, geïnventariseerd en georganiseerd door een internationaal instituut, de Cochrane Collaboration (zie ook par. 3.2).<sup>2</sup> Met literatuuroverzichten en meta-analyses wordt door dit instituut de kennis over werkzame behandelingen geïnventariseerd en beschikbaar gesteld. Middels wetenschappelijke publicaties en onderwijs worden de medische beroepsbeoefenaren gevoed met deze informatie.

Overzichten van evidence-based behandelingen in de medische wereld worden volgens een bepaalde methodiek gemaakt en verzameld in de Cochrane Library. Evidence wordt verzameld door zogenaamde Collaborative Review Groups (CRG's). Zo'n groep stelt zich tot doel om, na een inventarisatie van alle beschikbare studies naar een behandeling, bij voorkeur aan de hand van een meta-analyse, een uitspraak te doen over de werkzaamheid van die interventie. Onderzochte interventies lopen uiteen van vitamine C ter preventie van verkoudheid tot antidepressieve medicatie bij anorexia nervosa. De verzamelde kennis wordt vastgelegd in Evidence-Based Guidelines.

Een vergelijkbaar initiatief voor de evaluatie van het effect van sociale en educatie interventies is de Campbell Collaboration. Een van de eerste CRG's hield zich bezig met misdaad en straf (Farrington & Petrosino, 2001). Er zijn inmiddels rapporten in de Campbell Library opgenomen over bijvoorbeeld de effectiviteit van behandeling voor middelengebruik in detentie (effect is matig, [www.campbellcollaboration.org/doc-pdf/Incarceration-BasedDrugTxSept06final.pdf](http://www.campbellcollaboration.org/doc-pdf/Incarceration-BasedDrugTxSept06final.pdf)) en het effect van heropvoedingskampen op recidive (effect is nihil, [www.campbellcollaboration.org/doc-pdf/Wilson\\_bootcamps\\_rev.pdf](http://www.campbellcollaboration.org/doc-pdf/Wilson_bootcamps_rev.pdf)). Op stapel staat een meta-analyse van de effectiviteit van cognitieve-gedragstherapie met veroordeelde adolescenten en volwassenen.

### 3.1.3 *Kritiek op evidence-based medicine*

De toenemende toepassing van evidence-based medicine ontmoet ook kritiek. Er kunnen onder andere vraagtekens geplaatst worden bij de kwaliteit van de "evidence". Zo heeft bijvoorbeeld publicatiebias een negatief effect op de beschikbaarheid van informatie over niet effectieve interventies. Met deze bias wordt bedoeld op het gegeven dat de kans dat een studie met een positieve uitkomst (een experimentele behandeling is superieur, een hypothese wordt bevestigd) in een tijdschrift wordt gepubliceerd groter is, dan wanneer de studie een negatief resultaat heeft opgeleverd. Dit leidt tot een vertekening van de werkelijkheid. Er is een kerkhof voor niet-effectief gebleken behandelingen dat vol ligt met onbekende soldaten. Om publicatiebias enigszins tegen te gaan, of in ieder geval inzicht te krijgen in de omvang van het probleem, is er een internationale studieregistratie opgezet voor Randomised Controlled Trials (RCT's), het CONSORT-statement initiatief. Alvorens een interventie studie van start gaat, dient hij aangemeld te zijn (zie [www.consort-statement.org/Statement/revisedstatement.htm](http://www.consort-statement.org/Statement/revisedstatement.htm)). De onderzoekers van aangemelde studies zijn geboden na verloop van tijd de uitkomsten van de trial te melden bij de organisatie.

---

<sup>2</sup> Archie Cochrane was een Schotse epidemioloog die al in 1972 pleitte voor een meer empirische basis onder het medisch handelen. Als eerbetoon aan deze founding father is de organisatie naar hem vernoemd.

Een ander kritiek punt op evidence-based medicine is dat wat werkt volgens een meta-analyse van een groot aantal studies, nog niet hoeft te werken in het individuele geval van een patiënt die men nu behandelt. Nog een punt van zorg is dat een meer uniforme behandeling volgens richtlijnen en protocollen wellicht de mogelijkheden om vernieuwende behandelingen uit te voeren inperkt, wat een verdere toename van kennis in de weg staat.

### *3.1.4 Evidence-based Medicine in de geestelijke gezondheidszorg*

De ontwikkeling van evidence-based medicine en de groeiende toepassing in de somatische zorg wordt op enige afstand gevolgd door een vergelijkbare ontwikkeling in de ggz (Drake et al., 2001). Zo zijn er de laatste jaren richtlijnen tot ontwikkeling gekomen voor de behandeling van psychopathologie, eerst voor stemmingsstoornissen (Landelijke Stuurgroep Multidisciplinaire Richtlijnen in de GGZ, 2003a), later gevolgd door richtlijnen voor angststoornissen (Landelijke Stuurgroep Multidisciplinaire Richtlijnen in de GGZ, 2003b) en richtlijnen voor de behandeling van Schizofrenie (Landelijke Stuurgroep Multidisciplinaire Richtlijnen in de GGZ, 2007). Richtlijnen voor persoonlijkheidsstoornissen worden op dit moment opgesteld en worden naar verwachting op korte termijn gepubliceerd (begin 2008).

Implementatie van de richtlijnen vindt plaats in de ggz in de vorm van ontwikkeling van zorgprogramma's. Deze zorgprogramma's zijn verzamelingen van evidence-based behandelingen, of zouden dat althans moeten zijn. In werkelijkheid heeft men bij het opstellen van zorgprogramma's in de ggz in een aantal gevallen echter het behandelaanbod in de huidige praktijk beschreven en samengevat in plaats van de gewenste praktijk van behandelingen op basis van evidence-based guidelines weer te geven (Fenema et al., 2006). Een probleem is dat, net als in het somatische veld, de (kwaliteit van de) evidence voor behandelingen in de ggz ter discussie staat.

## **3.2 Evidence-based behandelen in de tbs**

De inhoud van zorgprogramma's voor de tbs-sector zou uiteindelijk zoveel mogelijk gebaseerd dienen te zijn op de evidence base zoals die is neergelegd in de Cochrane en Campbell Libraries. De totstandkoming van een evidence-based werkwijze in de tbs vereist het volgende stappenplan:

- goed gedocumenteerd wetenschappelijk onderzoek naar effectiviteit van behandelingen, m.a.w. het opbouwen van een evidence base;
- opstellen richtlijnen en zorgprogramma's voor behandelen op basis van de evidence base;
- implementeren van richtlijnen en zorgprogramma's.

Het invoeren van behandelingen volgens zorgprogramma's voor de tbs-sector naar model van de ggz en de somatische zorg vereist allereerst de beschikbaarheid van richtlijnen, protocollen en procedures waarin evidence-based behandelingen worden beschreven: een kennisbestand over wat werkt bij wie onder welke omstandigheden in de tbs. Dit kennisbestand – de evidence base – is op dit moment niet beschikbaar en zal opgebouwd moeten worden.

Deels kan daarvoor gebruik worden gemaakt van de reeds aanwezige kennis uit de Cochrane en Campbell Libraries. Formeel is dan weliswaar niet aangetoond dat in de ggz gebruikelijke behandelingen effectief zijn in de tbs-setting, maar de behandeling van geïsoleerde syndromen zoals angst- of stemmingsstoornissen zal in de tbs weinig anders zijn dan in de reguliere ggz. Tot op zekere hoogte geldt dit ook voor meer complexe stoornissen zoals de persoonlijkheidsstoornissen. Voor sommige persoonlijkheidsstoornissen, zoals de borderline persoonlijkheidsstoornis, zijn effectieve geprotocoliseerde behandelingen ontwikkeld, zoals Schema Focused

Therapy en Dialectische Gedragstherapie (Giessen-Bloo et al., 2006). De toepasbaarheid van deze behandelingen wordt onderzocht in de klinische praktijk van de ggz met een implementatie evaluatie (Nadort, 2006).

Deze wijze van opbouwen van de evidence base voor de tbs door “leentjebuurt” te spelen in de ggz zal echter om een aantal redenen niet voldoende zijn. In de eerste plaats is de doelstelling van behandeling in ggz en in de tbs verschillend: in de ggz gaat het om het terugdringen van de ziektelast en genezing, in de tbs is de doelstelling reductie van het risico op ongewenst gedrag en recidive. In de tweede plaats is zowel de setting als de populatie van de tbs dusdanig uniek dat niet zondermeer aangenomen kan worden dat wat effectief is in de ggz ook effectief is in de tbs. Om die reden zal veel onderzoek uit de ggz gerepliceerd moeten worden in een tbs-context. Zo wordt bijvoorbeeld de effectiviteit van Schema Focused Therapy in de tbs-sector momenteel onderzocht door een aantal tbs-klinieken (zie paragraaf 6.1.2). Deze studie zal uitwijzen of deze behandeling ook voor de tbs-sector het predicaat “evidence-based” verdient. Voorts is ook in de ggz de ontwikkeling van een evidence base nog volop “work in progress”.

Het opbouwen van de evidence base door middel van wetenschappelijk onderzoek is een kwestie van lange adem. Het vullen van het kennisbestand middels internationale studies biedt deels uitkomst. Dit geldt met name voor stoornissen waar ook internationaal in forensische context behandelingen voor zijn ontwikkeld.

Na het opbouwen van een evidence base kunnen protocollen en richtlijnen worden ontwikkeld (stap 2). De gebruikelijke wijze om tot richtlijnen te komen is door middel van een commissie met experts uit het veld die streeft naar een consensusuitspraak. Vervolgens dienen de richtlijnen op de werkvloer geïmplementeerd te worden (stap 3). Overigens dient te worden opgemerkt dat het feitelijke behandelaanbod in de klinische praktijk altijd ruimer zal zijn en meer zal omvatten dan alleen behandelingen met het predicaat “evidence-based”, ook bij volledige implementatie van werken volgens zorgprogramma’s. Dat is ook een goede zaak, omdat er mogelijkheden moeten blijven om ervaring op te doen met nieuwe, experimentele, potentieel werkzame behandelingen.

## **4. De reikwijdte van de evidence base**

Voorafgaande aan het beschrijven van onderzoeksmethoden voor therapie-effectonderzoek in de tbs-sector behoeven twee aspecten een nadere toelichting. Bij therapie-effectonderzoek kan men denken aan onderzoek naar de effectiviteit van de gehele tbs of aan onderzoek naar de effectiviteit van specifieke interventies, zoals die tijdens de tbs plaatsvinden. In 4.1 wordt ingegaan op dit verschil in relatie tot het opbouwen van een evidence base. Het tweede aspect is de uitkomstmaat waaraan de effectiviteit van de tbs-behandeling wordt afgemeten. Aangezien het primaire doel van de tbs is om recidive tegen te gaan, ligt het voor de hand recidive als uitkomstmaat te nemen. Of recidive altijd de meest geschikte maat is voor therapie-effectonderzoek zal worden besproken in 4.2.

### **4.1 Het effect van de tbs-maatregel vs. het effect van specifieke tbs-interventies en behandelmodules**

De tbs-maatregel is meer dan alleen het aanbieden van psychotherapeutische interventies, maar behelst een geheel aan afdelingsactiviteiten, instellingsklimaat en een stapsgewijs gecontroleerd re-integratietraject. Wanneer het gaat om de effectiviteit van de tbs kan het dus zowel gaan om 1) de effectiviteit van dit geheel aan activiteiten binnen de tbs-maatregel, m.a.w. de tbs-maatregel in zijn totaliteit, als om 2) de effectiviteit van de specifieke interventies die in een tbs-kader aan tbs-gestelden worden aangeboden.

#### *4.1.1 De effectiviteit van de tbs-maatregel*

De vraag of de tbs-behandeling in zijn totaliteit effectief is, zou kunnen worden beantwoord aan de hand van onderzoek waarbij tbs-gestelden willekeurig zouden worden toegewezen aan een tbs-behandeling versus een andere interventie (bijvoorbeeld een andere behandeling of gevangenisstraf). Een dergelijke willekeurige toewijzing is echter vanuit maatschappelijke en juridische normen volstrekt onaanvaardbaar. Voor onderzoek naar de effectiviteit van de tbs in zijn totaliteit moet daarom altijd gezocht worden naar second-best onderzoeksmethoden die wel aanvaardbaar zijn.

Een alternatief dat soms gekozen wordt, is het vergelijken van de recidivecijfers van ex-gedetineerden en ex-tbs-gestelden. Recidivecijfers laten dan zien dat ex-tbs-patiënten minder vaak recidiveren dan ex-gedetineerden (bijvoorbeeld Wartna, Tollenaar & Essers, 2005; Wartna, El Habachi & Essers, 2006). Het nadeel van deze onderzoeksbenadering is dat men hieruit nog niet mag afleiden dat de tbs een effectievere maatregel is dan gevangenisstraf. Namelijk beide populaties verschillen op voorhand en deze verschillen kunnen leiden tot een verschil in recidivecijfers dat los kan staan van een eventueel effect van de opgelegde maatregel of straf.

Een ander alternatief zou kunnen zijn binnen de tbs en detentie op zoek te gaan naar vergelijkbare groepen (bijvoorbeeld door te matchen op onder meer het delict waarvoor de betrokkene is veroordeeld, psychopathologie, leeftijd en IQ) en de recidivecijfers van beide (gematchte) groepen te vergelijken. Het is echter de vraag of dergelijk onderzoek een bijdrage levert aan het opbouwen van een evidence base van werkzame behandelingen. Bij dergelijk onderzoek naar de tbs in zijn totaliteit wordt immers een “black box” geëvalueerd, die bestaat uit een complex aan behandelingen en andere aspecten inherent aan de tbs. Reductie van recidive kan, naast de specifieke interventies die gegeven worden, immers óók het gevolg zijn van een juiste “selectie aan de uitgang” bij de tbs: alleen tbs-gestelden met goede vooruitzichten worden

weer in vrijheid gesteld. Ook verschilt het na-traject van de tbs van een reguliere gevangenisstraf. Bij een tbs-maatregel hoort een intensief re-integratietraject met begeleiding en verschillende gradaties van verloven. Vrijlating is verbonden aan een reeks voorwaarden waaraan de patiënt dient te voldoen.

Onderzoek naar de effectiviteit van de tbs in z'n totaliteit in vergelijking met gevangenisstraf kent dus beperkingen. Wanneer men de groepen (tbs-gestelden en gedetineerden) niet matcht kan niet uitgesloten worden dat een eventueel verschil in recidive het gevolg is van vooraf bestaande verschillen tussen tbs-gestelden en reguliere gedetineerden. Wanneer men de groepen wel (zo goed mogelijk) zou matchen dan is bij een eventueel verschil in recidive onduidelijk welk aspect van de maatregel (bijvoorbeeld de succesvolle behandeling van de patiënt, een goed re-integratietraject of een juiste selectie van patiënten die in aanmerking komen voor beëindiging van de tbs-maatregel) hieraan heeft bijgedragen. Onderzoek naar de effectiviteit van de tbs in zijn totaliteit draagt derhalve niet bij aan de evidence base van specifieke interventies en biedt daarmee géén handvatten voor het opstellen van richtlijnen met betrekking tot behandelingen die in tbs-kader worden geboden.

#### *4.1.2 De effectiviteit van specifieke behandelingen*

Het onderzoeken van de effectiviteit van specifieke interventies die in de tbs worden toegepast is in vergelijking met onderzoek naar de tbs in zijn totaliteit naar verhouding goed mogelijk en nuttig voor het komen tot een evidence base. Hierbij hoeft men niet te rade te gaan bij een externe populatie, maar vergelijkt men tbs-gestelden die een specifieke interventie ontvangen met tbs-gestelden die deze (nog) niet hebben ontvangen. Het vinden van een vergelijkbare controlegroep is dan relatief eenvoudig.

## **4.2 De uitkomstvariabele**

### *4.2.1 Recidive*

Recidive betekent letterlijk herhaling, maar de term wordt voornamelijk gebruikt in de zin van herhaling van strafbare feiten. Recidivecijfers geven daarbij het wederom plegen van een delict aan na een eerdere veroordeling. Het moment van herhaling doet er in principe niet veel toe: het recidive delict kan direct na vrijlating plaatsvinden, maar ook jaren later. Doorgaans hanteert men een termijn van enkele jaren voor het bepalen van het recidivecijfer.

Recidive lijkt in eerste instantie de meest voor de hand liggende uitkomstmaat om de effectiviteit van tbs-behandelingen aan af te meten. Immers, tbs-behandelingen zijn uiteindelijk gericht op het terugdringen van recidive. Er kleven echter een aantal haken en ogen aan recidive als uitkomstmaat. Het eerste is vooral praktisch van aard: het duurt geruime tijd voordat recidive vastgesteld kan worden. Het moment van recidivemeting, zelfs wanneer deze kort na het ontslag uit de kliniek plaatsvindt, ligt vaak jaren verwijderd van de beëindiging van de onderzochte interventie.

Daarnaast is het te verwachten effect van één specifieke interventie op recidive beperkt. Dit komt doordat de interventie slechts een van de vele zaken is die van invloed zal zijn op recidive. Naast de onderzochte interventie worden bijvoorbeeld andere behandelmodules en interventies (bijvoorbeeld het therapeutisch klimaat) aangeboden binnen de tbs-maatregel en vinden gebeurtenissen plaats in het leven van de tbs-gestelde (zoals het vinden of verliezen van een partner) die van invloed kunnen zijn op het al dan niet recidiveren. Het effect van een specifieke interventie op later delictgedrag binnen dit complex aan factoren zal daardoor relatief

beperkt zijn. Het praktische probleem met een dusdanig beperkt effect is dat er een extra inspanning vereist is, in de zin dat het onderzoek zal moeten worden uitgevoerd met grote groepen, om het effect wetenschappelijk aan te kunnen tonen. Voor het aantonen van een recidivereductie van 10% moet men ongeveer 250 proefpersonen in zowel de experimentele conditie als controle conditie includeren (Wartna, 2005). Een andere manier om bij een beperkt effect toch een significant verschil aan te tonen is het gebruik van een fijngevoelig instrument als uitkomstmaat. Echter, recidive is juist erg grofmazig: een ex-patiënt recidiveert wel of niet.

Hoewel recidivecijfers dus niet zo geschikt lijken om de effectiviteit van specifieke interventies te onderzoeken, kunnen zij wel gebruikt worden als algemene uitkomstmaat voor evaluatie van het geheel van het tbs-stelsel (het geheel van onder andere interventies, therapeutisch klimaat, selectie bij de beëindiging van de maatregel en nazorg). In dat geval geven recidivecijfers een absolute indicatie van de mate waarin de primaire doelstelling van de tbs-behandeling wordt bereikt. Recidivegegevens over de tijd bieden een historische controle en kunnen toekomstige verbetering (of verslechtering) van het effect van de tbs aan het licht brengen. Ook zouden recidivecijfers als ankerpunt kunnen dienen om toekomstige prestaties tegen af te zetten. Hierbij moet wel aangetekend worden dat bij dergelijke vragen niet uitgesloten kan worden dat andere factoren dan veranderingen in het tbs-stelsel van invloed zijn op de recidivecijfers, bijvoorbeeld verslechtering van het researchwerk van de politie waardoor de pakkans afneemt.

Samenvattend, voor therapie-effectonderzoek naar specifieke interventies in de tbs-sector lijkt recidive een minder geschikte uitkomstmaat, vooral vanwege de lange termijn die nodig is om de uitkomst vast te stellen en de ongevoeligheid voor het meten van beperkte effecten, waardoor een groot aantal patiënten geïncludeerd zou moeten worden om dat effect aan te kunnen tonen.

#### 4.2.2 *Recidiverisico*

In plaats van door middel van recidivegegevens kan het effect van een interventie ook vastgesteld worden aan de hand van een *benadering* van recidive (een proxy). Een mogelijk alternatief is de inschatting van de kans op recidive (risicotaxatie). Risicotaxatie kan een goede proxy zijn voor recidive als er sprake is van een sterke samenhang tussen beide variabelen. Echter, de samenhang tussen risicotaxatie-instrumenten en daadwerkelijke recidive is niet perfect. Een match tussen voorspelling en daadwerkelijke recidive in 70 tot 80% van de gevallen is doorgaans een mooi resultaat en is het hoogst haalbare bij de huidige stand van kennis en beschikbare meetinstrumenten (Snowdon et al, 2007).

Wanneer men risicotaxatie-instrumenten wil hanteren als uitkomstmaat, dan zal men moeten kiezen voor risicotaxatie-instrumenten waarin (een flink aantal) dynamische factoren zijn opgenomen. Veel risicotaxatie-instrumenten gaan (deels) uit van historische en onveranderbare indicatoren. Een justitieel verleden is bijvoorbeeld een goede voorspeller voor toekomstig recidive, maar een tbs-behandeling kan niet een criminele historie uitwissen. Deze historische indicatoren zijn dus niet geschikt voor het meten van verandering als gevolg van een interventie. De wel veranderbare indicatoren, ook wel dynamische risico indicatoren genoemd, zijn in een veelgebruikt risicotaxatie-instrument als de HCR-20 (Philipse, de Ruiters, Hildebrand, Bouman & Webster, 2000) in de minderheid. Een stap in de goede richting is de ontwikkeling van de HKT-30 (Ministerie van Justitie, 2003) en de FP40 lijsten (Brand & Van Emmerik, 2006) en FOTRES (Urbaniok, 2006) welke meer aandacht hebben voor dynamische risicofactoren.

#### 4.2.3 *Andere uitkomstmaten*

Naast risicotaxatie als proxy voor recidive kunnen ook andere uitkomstmaten gehanteerd worden om de effectiviteit van interventies te onderzoeken. Daarbij gaat het vooral om uitkomstmaten die zowel in tijd als in oorzakelijke samenhang dicht bij de interventie zelf staan. Een specifieke module als agressiehanteringstherapie, bijvoorbeeld, heeft tot doel de patiënt te leren conflictsituaties beter te hanteren. Een geschikte uitkomstmaat om het effect van een dergelijke interventie te onderzoeken zou dan de aanwezigheid van de geleerde sociale vaardigheden of het aantal agressie-incidenten op de afdeling meten. Het grote voordeel van dergelijke meer directe uitkomstmaten is dat deze fijngevoelig zijn om subtiele verandering te meten en bovendien het directe doel van de therapie evalueren. Nadeel is echter dat de relatie met het latere delictgedrag vaak slechts berust op een veronderstelling.

#### 4.2.4 *Overwegingen bij meetinstrumenten.*

Vanzelfsprekend moeten meetinstrumenten betrouwbaar, accuraat en valide zijn. Deze psychometrische aspecten moeten in wetenschappelijk onderzoek aangetoond zijn. Er wordt dan op dit gebied tegenwoordig wel gesproken van “evidence-based assessment” (Hunsley & Mash, 2007). Bij uitkomstmaten kan er onderscheid gemaakt worden tussen generieke meetinstrumenten (zoals schalen voor risicotaxatie, patiëntsatisfactie en kwaliteit van leven) versus stoornis- of behandelingsspecifieke meetinstrumenten (agressie incidenten of psychotische episoden). Met generieke meetinstrumenten is de uitkomst van behandeling voor alle patiënten, afdelingen, locaties en instellingen onderling vergelijkbaar. Generieke meetinstrumenten zijn echter over het algemeen minder gevoelig voor verandering en minder geschikt om het effect van specifieke interventies te onderzoeken. Voor de evaluatie van een medicijn voor het tegengaan van psychotische symptomen dienen nu eenmaal andere meetinstrumenten gebruikt te worden dan voor de evaluatie van een interventie die gericht is op het verbeteren van sociale vaardigheden. In het algemeen geldt dat hoe preciezer en specifiek het meetinstrument het behandeldoel meet, des te sensitiever het is om een eventuele verandering te detecteren.

Er zijn drie typen instrumenten: zelfrapportage, beoordelingsschalen en event monitoring. Het gebruik van zelfrapportage in de tbs kent beperkingen. We kunnen er niet op voorhand vanuit gaan dat respondenten in een forensische context vragenlijsten naar waarheid zullen beantwoorden. Deze respondenten hebben er alle belang bij zichzelf anders (lees beter) voor te doen dan de werkelijkheid is in de context van een gewenste beëindiging van de tbs. De zelfrapportage meetinstrumenten die in de geestelijke gezondheidszorg veel gebruikt worden, zoals de SCL-90, zijn voor het forensische veld dus maar beperkt toepasbaar, al kan daar een uitweg voor zijn door te trachten vast te stellen hoe waarheidsgetrouw de respondent zelfrapportage lijsten invult (vgl. Cima, 2005).

Bij beoordelingsschalen geldt dat de interbeoordelaarsbetrouwbaarheid tussen deskundigen van algemene beoordelingen over het algemeen laag is. De betrouwbaarheid wordt wel beter bij gebruik van beoordelingsschalen met een gedetailleerde beschrijving van ankerpunten voor de beoordeling. Een belangrijke keuze bij gebruik van beoordelingsschalen is wie de beoordelingsschaal invult: de behandelaar of een onafhankelijk beoordelaar. De behandelaar kent de patiënt en zijn situatie het best en beschikt daarmee over meer informatie om tot een valide oordeel te komen dan een onafhankelijk beoordelaar. De behandelaar heeft echter ook een potentiële bias naar overschatting van het effect van eigen inspanningen (“de slager die zijn eigen vlees keurt”). Idealiter wordt gebruik gemaakt van beide vormen van



beoordelingen. De validiteit van de beoordelingen kan dan worden ondersteund door de convergentie van de beoordelingen.

Een derde methode om therapie-effect te meten is continue monitoring van gebeurtenissen naast zelfrapportage en beoordelingsschalen, bijvoorbeeld agressieve incidenten op de afdeling. Op dit gebied is door Prof. Dr. H. Nijman in Nederland belangrijk werk verricht door het ontwikkelen van de Staff Observation Aggression Scale-revised (Nijman, Allertz & Á Campo, 1995). Bij de interpretatie van dergelijke gegevens moet de onderzoeker wel een open oog houden voor de (restrictieve) omstandigheden waar de onderzochte patiënt in verkeerde ten tijde van de monitoring. Agressieve incidenten tijdens verlof hebben een ander gewicht dan dergelijke incidenten binnen de instelling en hetzelfde geldt voor het uitblijven van agressief gedrag binnen en buiten de instelling.

Ten slotte is het belangrijk om de resultaten van behandeling niet alleen uit te drukken in variabelen waarop de interventie is gericht, zoals klachten, symptomen of het gedrag, maar ook andere variabelen bij de evaluatie te betrekken, zoals het percentage deelnemers dat de behandeling voortijdig afbreekt. Zogenaamde drop-outs zijn informatief over de toepasbaarheid van de interventie: zo kan blijken dat een behandeling zeer effectief is, maar onverdraaglijk voor de meeste patiënten.

#### 4.3 Conclusie

In de tbs worden psychische stoornissen behandeld teneinde recidive te beperken. Een terechte vraag is: slaagt de tbs hierin? Zoals we hebben betoogd in dit hoofdstuk, zijn de mogelijkheden beperkt om deze vraag te beantwoorden met de uitkomst van effectstudies in de tbs. Onderzoek naar de tbs-maatregel in zijn totaliteit is namelijk lastig, en de mogelijkheden om de invloed van interventies (rechtstreeks) in verband te brengen met recidive zijn beperkt.

Wat kan er wel verwacht worden van effectstudies in het tbs-kader? Globaal genomen hetzelfde als binnen de ggz: het meten van effectiviteit van specifieke behandelinterventies en therapieën door middel van directe uitkomstmaten, zoals omschreven behandeldoelen (bijvoorbeeld een vermindering van dynamische risicofactoren, reductie van psychopathologie of het aanleren van sociale vaardigheden).

Hoewel beperkt, is aantonen van effectiviteit van interventies op directe uitkomstmaten de eerste noodzakelijke stap op weg naar een meer evidence-based benadering in de tbs. Dit wil overigens niet zeggen dat er op basis van effectiviteitsstudies geen zinvolle uitspraken gedaan kunnen worden over de belangrijke vraag naar de relatie tussen behandeling in de tbs en recidive. Wanneer de behandeling bijdraagt aan het bereiken van bepaalde behandeldoelen en bekend is dat deze behandeldoelen direct of indirect gerelateerd zijn aan recidive, dan kan men veronderstellen dat een geslaagde behandeling leidt tot recidivevermindering. Deze relatie met de recidive berust echter wel op een aantal aannames en het directe wetenschappelijk bewijs voor de recidive verminderende werking van de behandeling is dan nog niet geleverd.



## **Deel 3 Methoden van onderzoek voor een evidence base in de tbs**



## ***5. Algemene methodologische eigenschappen therapie-effectonderzoek***

In het volgende hoofdstuk zullen een drietal onderzoeksdesigns besproken worden. Om een onderzoeksdesign ten behoeve van therapie-effectonderzoek in de tbs op haar waarde te kunnen schatten, dienen er met twee aspecten rekening gehouden te worden. Enerzijds gaat het om de methodologisch sterke en zwakke punten van het onderzoeksdesign zelf. Wij zullen ons hierbij concentreren op de vraag in welke mate het design in staat is de onderzoeksvraag te beantwoorden (validiteit) en op de vraag in welke mate het design in staat is meerdere causale factoren in ogenschouw te nemen (responsiviteit). Anderzijds gaat het om de praktische punten van de toepasbaarheid van het design binnen de specifieke tbs-setting. Ieder design heeft haar eigen pakket aan eisen die niet altijd even eenvoudig te realiseren zijn in de tbs-setting. Voordat we overgaan tot de beschrijving van de specifieke designs bespreken we in dit hoofdstuk deze aspecten. Zoals in hoofdstuk 4 beargumenteerd, richten we ons in beide hoofdstukken op methoden ten behoeve van het opstellen van evidence-based behandelprogramma's bestaande uit specifieke interventies en dus niet op onderzoek naar de effectiviteit van de tbs-behandeling is zijn totaliteit.

### **5.1 Methodologisch relevante aspecten**

#### *5.1.1 Interne validiteit en werkzaamheid*

In wetenschappelijk onderzoek wordt onderscheid gemaakt tussen interne en externe validiteit van het onderzoeksdesign (Rossi & Freeman, 1996). Met interne validiteit wordt bedoeld de validiteit van het onderzoek ten gevolge van de methodologische gestrengheid waarmee het onderzoek is opgezet en uitgevoerd. Het gaat hierbij in eerste instantie om de vraag of het design een ondubbelzinnig antwoord mogelijk maakt op de onderzoeksvraag. Een onderzoeksopzet met een hoge interne validiteit staat toe dat geobserveerde verschillen terecht worden toegewezen aan de experimentele condities zonder dat er alternatieve verklaringen mogelijk zijn. Causale relaties zijn dan eenduidig af te leiden uit de resultaten.

Interne validiteit hangt nauw samen met de gevonden werkzaamheid van de onderzochte behandeling. Designs met hoge interne validiteit zijn vooral bedoeld om de werkzaamheid van behandelingen of de superioriteit van nieuwe behandelingen versus de gebruikelijke behandeling te onderzoeken onder streng gecontroleerde omstandigheden. In de Angelsaksische literatuur duidt men dit type onderzoek aan met de term “efficacy” onderzoek (onderzoek naar de werkzaamheid van een interventie). De interne validiteit heeft zo een directe relatie met de waarde van de resultaten van het onderzoek voor de evidence base. Er zijn verschillende maatregelen mogelijk om de interne validiteit van therapie-effectonderzoek te vergroten, zoals een zorgvuldige en gedetailleerde omschrijving van de onafhankelijke variabele of experimentele manipulatie (protocollen en draaiboeken voor de behandeling) en controle op de toepassing hiervan, het gebruiken van betrouwbare en sensitieve meetinstrumenten voor het meten van de afhankelijke variabelen en welomschreven en goed gediagnosticeerde deelnemers aan de trial.

#### *5.1.2 Externe validiteit en effectiviteit*

Met externe validiteit van het onderzoek wordt bedoeld de zeggingskracht van de resultaten voor de werkelijkheid (buiten het laboratorium). Bij therapie-effectonderzoek gaat het om de vraag in hoeverre de resultaten van een wetenschappelijk onderzoek te generaliseren zijn naar de klinische praktijk van

alledag. De generaliseerbaarheid hangt onder meer af van de omstandigheden waaronder de behandeling werd gegeven tijdens de trial en of deze voldoende lijken op de omstandigheden die heersen in de klinische werkelijkheid. Andere factoren voor de generaliseerbaarheid zijn de inclusie- en exclusiecriteria die men heeft gehanteerd ten aanzien van deelnemers aan de trial (zowel patiënten als behandelaars).

Onderzoeksdesigns die zich richten op de externe validiteit zijn niet zozeer gericht op de werkzaamheid van de behandeling, maar meer op de vraag of de behandeling effect sorteert in de klinische werkelijkheid. Men spreekt dan van zogenaamd “effectiveness” onderzoek (Howard et al., 1996). De vraag die bij dit type onderzoek wordt gesteld is: wat is de uitkomst van behandelingen, die in onderzoek onder gecontroleerde omstandigheden hun werkzaamheid hebben bewezen, *in de klinische praktijk*. Vragen zijn gericht op effectiviteit in termen van reductie van klachten, aantal herstelde patiënten en toegenomen kwaliteit van leven en doelmatigheid van de behandeling in termen van kosteneffectiviteit.

Interne en externe validiteit van een onderzoek staan soms op gespannen voet met elkaar: optimalisatie van de ene vorm van validiteit gaat dan ten koste van de andere. Dat geldt bijvoorbeeld bij de selectie van deelnemers aan een trial. Om de werkzaamheid van een nieuwe behandeling te evalueren, is de selectie van een zo homogeen mogelijke deelnemersgroep van prototypische patiënten van belang (bijvoorbeeld patiënten met alleen een psychose of patiënten met alleen een borderline persoonlijkheidsstoornis, m.a.w. zonder co-morbide stoornissen). De externe validiteit van de bevindingen is dan echter beperkt, want in de klinische werkelijkheid treffen we prototypische patiënten met een enkelvoudige stoornis weinig aan (vgl. Westen, Novotny & Thompson-Brenner, 2004). De meeste patiënten hebben gelijktijdig meerdere stoornissen en zijn bijvoorbeeld zowel depressief als angstig of zowel verslaafd als psychotisch. Een ander voorbeeld waarbij interne en externe validiteit op gespannen voet staan, is de zorg die men in een trial besteedt aan de juiste uitvoering van de behandeling. Therapeuten worden uitgebreid getraind en gesuperviseerd en de behandeling wordt gemonitord op getrouwe uitvoering van het draaiboek of behandelprotocol. Dat komt de interne validiteit van de studie ten goede (de interventie wordt uitgevoerd zoals hij bedoeld is), maar in de klinische praktijk van alledag ontbreekt deze aandacht voor “treatment integrity” en zal vaak een wat verdunde versie van de behandeling toegepast worden.

### 5.1.3 Responsiviteit (prognostische variabelen)

Naast vragen of een behandeling werkt en effectief is op groepsniveau, is het vervolgens relevant te weten voor wie en onder welke omstandigheden. Bij toepassing van een bewezen effectieve behandeling, is er doorgaans een aanzienlijke groep patiënten die geen baat heeft bij de behandeling. Internationale studies onder seksueel delinquenten, bijvoorbeeld, laten zien dat behandeling recidive reduceert (Veen & De Ruiter, 2005). Desondanks recidiveert ongeveer 1 op de 10 delinquenten na behandeling (Hanson, 2002). Het is belangrijk inzicht te krijgen in de factoren die bepalen of een behandeling al dan niet aanslaat, alsook in de factoren die van invloed zijn op het al dan niet volhouden van de behandeling, en op basis daarvan programma's te ontwikkelen die drop-outs zoveel mogelijk voorkomen. Dergelijke factoren worden binnen de medische wetenschap “prognostische variabelen”

genoemd. Binnen de forensische psychiatrie staan deze ook bekend onder de naam “responsiviteitsfactoren”.<sup>3</sup>

Responsiviteitsfactoren kunnen onderverdeeld worden in enerzijds interne factoren en anderzijds externe factoren (Looman, Dickie & Abracen, 2005; Kennedy, 2000). Onder interne factoren verstaat men factoren van het individu of de patiënt zelf die van invloed kunnen zijn op diens vermogen om van een behandeling te profiteren. Hierbij kan gedacht worden aan leeftijd, culturele factoren, geslacht, behandelmotivatie en behandelbereidheid, ontkenning van het delict, intellectuele capaciteiten, psychische stoornis, attitudes als vijandigheid of weerstand. Externe responsiviteitsfactoren betreffen eigenschappen van het behandelteam en -setting, zoals vorm van de behandeling (groepstherapie vs. individueel), open vs. gesloten groepsformat, therapeutische communities, eigenschappen van de therapeut (o.a. empathie, warmte, belonend, directief/niet-confrontatief en interpersoonlijke vaardigheden). Zie ook Kader 1 “*Voorbeelden Responsiviteitsonderzoek*”.

*5.1.3.1 Motivatie als interne responsiviteitsfactor.* Een belangrijke responsiviteitsfactor die bij effectiviteitstudies in ogenschouw genomen moet worden, is behandelmotivatie (Drieschner, Lammers & Van der Staak, 2004). De variabele behandelmotivatie is in de eerste plaats relevant als risicofactor (patiënten met een lage motivatie voor behandeling hebben een verhoogd risico op recidive) en als zodanig ook opgenomen in risicotaxatie-instrumenten als de HCR-20 (Philipse et al., 2000) en HKT-30 (Ministerie van Justitie, 2003). Uit Nederlands onderzoek blijkt behandelmotivatie een van de weinige dynamische items uit de HCR-20 en HKT-30 die samenhangt met recidive (Hildebrand, Hesper, Spreen & Nijman, 2005). Bij de relatie tussen motivatie en recidive zijn overigens ook kanttekeningen te plaatsen. Zo blijkt uit een recente meta-analyse onder bijna 30.000 seksueel delinquenten geen verband tussen motivatie en recidive (Hanson & Morton-Bourgon, 2005).

Behalve als directe risicofactor voor recidive is behandelmotivatie ook indirect van invloed vanwege het effect op de uitkomst van behandeling. Gezien het gedwongen karakter van de tbs is motivatie bij de tbs-gestelde om te veranderen, anders dan bij patiënten in de reguliere ggz die (vaak) zelf hulp zoeken, niet vanzelfsprekend. Daarbij komt dat ongemotiveerde patiënten zich moeilijk laten behandelen. Er zijn belangrijke verschillen tussen delinquenten die een gedwongen behandeling opgelegd krijgen en zij die vrijwillig een behandeling ondergaan. Vergelijkend onderzoek onder daders van huiselijk geweld die door de rechter een behandeling opgelegd kregen versus daders die vrijwillig behandeling zochten, heeft laten zien dat beide groepen nauwelijks verschillen op ondermeer achtergrondkenmerken, persoonlijke variabelen en levensstijl (Bowen & Gilchrist, 2004). De daders met een opgelegde behandeling zijn echter beduidend minder gemotiveerd en leggen minder vaak de oorzaak van hun gedrag binnen zichzelf (een interne locus of control). Bovendien tonen zij lagere niveaus van woede in relatie tot oneerlijkheid of frustraties en interrupties. Dergelijke verschillen maken dat

---

<sup>3</sup> De term responsiviteit stamt uit het risk-needs-responsivity model van Andrews & Bonta (1990, 1994; zie Ward, Melsler, & Yates, 2007), waarin wordt gesteld dat er drie principes ten grondslag liggen aan de effectiviteit van een forensische behandeling. Ten eerste het risicoprincipe: de behandeling bij mensen met een hoog risico dient intensiever en meer gestructureerd te zijn dan bij mensen met een laag risico. Ten tweede het behoeftenprincipe: de behandeling moet gericht zijn op vervulling van behoeften waarvan bekend is dat deze recidive vergroten (criminogenic needs). Aan deze voorwaarden wordt in het tbs-kader al snel voldaan aangezien er een selectie plaatsvindt op hoog risico patiënten en de behandeling zich expliciet richt op het reduceren van recidiverisico. Het derde principe volgens Andrews en Bonta is het principe van responsiviteit: de behandeling moet gevoelig zijn voor de individuele eigenschappen van de delinquent.

aangetoonde effectiviteit van een behandeling binnen een vrijwillige setting als de ggz niet automatisch gegeneraliseerd kan worden naar de tbs.

In effectiviteitstudies in de tbs zou er daarom aandacht besteed moeten worden aan behandelmotivatie, tenminste als controlevariabele maar bij voorkeur als co-variabele voor behandel-effect. Ook studies naar manieren om behandelmotivatie bij patiënten te verhogen verdienen de aandacht (McMurrin & Ward, 2004). Het spreekt voor zich dat het concept daartoe wel goed geconceptualiseerd en geoperationaliseerd moet zijn (Drieschner, Lammers & Van der Staak, 2004; Sellen, McMurrin, Cox, Theodosi & Klinger, 2006).

*5.1.3.2 Biologische en neuropsychologische responsiviteitsfactoren.* De laatste jaren wordt er steeds meer bekend over de relatie tussen het functioneren van het brein (in de vorm van neurobiologische en neuropsychologische maten) en impulsief, agressief, en antisociaal gedrag. De uitkomsten van dergelijk onderzoek zijn relevant voor effectonderzoek in de tbs (de Kogel, in voorbereiding). Neurobiologische en neuropsychologische variabelen zijn potentieel waardevolle interne responsiviteitsfactoren. We noemen een paar hypothesen. Een tbs-gestelde met een verminderd reactieve HPA-as (de biologische basis van het stresssysteem) kan mogelijk pas baat hebben bij psychosociale interventies als eerst de HPA-as-reactiviteit is genormaliseerd door middel van een farmacologische interventie (Van Wiel et al., 2004). Een tbs-gestelde met een deficiënte aanmaak van de hormonen oxytocine en vasopressine (bijvoorbeeld ten gevolge van een jeugd die was gekenmerkt door verwaarlozing en mishandeling) is minder in staat tot positief sociaal gedrag, hechting en vertrouwen (Insel & Fernald, 2004). Toediening van oxytocine creëert bij deze patiënten mogelijk de noodzakelijke voorwaarde om ontvankelijk te zijn voor therapeutische interventies. Een gedetineerde met een brein dat deficiënt is op neuropsychologische taken die impulscontrole en vermogen tot inhibitie meten, heeft minder baat bij een standaard behandeling met cognitieve gedragstherapie (Fishbein & Sheppard, 2006), maar zal mogelijk meer baat hebben bij het trainen van zelfcontroletechnieken. Het in mindere mate aanwezig zijn van spiegelneuronen bij een patiënt met een antisociale persoonlijkheidsstoornis stelt mogelijk grenzen aan de behandelbaarheid van deze psychopathologie (Iacoboni & Lenzi, 2002).

*5.1.3.3 Onderzoek naar responsiviteit.* Responsiviteit kan op grofweg twee manieren onderzocht worden. Ten eerste kunnen dergelijke variabelen als controlevariabelen opgenomen worden binnen een design dat als primair doel heeft de werkzaamheid of effectiviteit van een behandeling te vergelijken met een andere groep die niet wordt behandeld. De vraagstelling kan bijvoorbeeld zijn of het behandel-effect mede afhankelijk is van de aard van het delict waarvoor men veroordeeld werd. Het is echter niet altijd mogelijk om responsiviteitsvariabelen binnen een ander onderzoeksdesign in te passen, met name wanneer het externe factoren betreft. Zoals in het volgende hoofdstuk zal worden beschreven, heeft bijvoorbeeld een RCT vrij strikte eisen aan het controleren van zoveel mogelijk variabelen. Een bepaald type delict kan bijvoorbeeld een inclusiecriteria zijn om tot een homogene deelnemersgroep te komen. Echter, dit sluit uit dat type delict als responsiviteitsfactor onderzocht kan worden. Een ander voorbeeld is het format van toediening van een interventie (individuele of groepstherapie). Ook is het in het kader van een RCT niet toegestaan de setting van de behandeling al te zeer te variëren. Bovendien zorgt implementeren van extra variabelen voor een probleem



met de statistische power: hoe meer variabelen, hoe meer proefpersonen er nodig zijn om een mogelijk effect aan te kunnen tonen.

#### Kader 1

##### *Voorbeelden Responsiviteitsonderzoek*

Een voorbeeld van externe responsiviteitsfactoren die zijn onderzocht betreft de eigenschappen van de therapeut (zie Marshall [2005] voor een review). Marshall et al. (2002) onderzochten een groot aantal eigenschappen (gedrag en vaardigheden) van de therapeut in relatie tot gedragsveranderingen bij de delinquent. Daartoe werden twee studies opgezet. In de eerste studie werd op basis van videobeelden bepaald welke eigenschappen van de therapeut door getrainde beoordelaars betrouwbaar konden worden vastgesteld en in werkelijkheid voldoende voorkwamen. Dit leverde een lijst op van 17 eigenschappen, waaronder empathie, warmte, toepasselijke humor, toepasselijke lichaamstaal, het stellen van open vragen, directief en confrontatief. In de tweede studie werd deze lijst gebruikt om te bepalen welke eigenschappen het meest geassocieerd zijn met gedragsveranderingen bij de delinquent op in totaal 44 verandermaten. Ten behoeve van dit onderzoek werden door het Canadese gevangeniswezen 25 videobanden van therapie sessies (waarop alleen de therapeut zichtbaar was) beschikbaar gesteld die dusdanig geselecteerd waren dat de variatie in gedragsverandering van de delinquenten zo groot mogelijk was. Nadat deze banden waren gescoord op eigenschappen van de therapeut, werden de onderzoekers geïnformeerd over de (voor reguliere doeleinden gemeten) gedragsveranderingen. Hieruit bleek dat met name de eigenschappen empathie, warmte, belonend, directief en een niet-confrontatieve opstelling van de behandelaar tot gedragsverandering leidden.

Een ander voorbeeld van een onderzoek naar externe responsiviteit betreft het Canadese onderzoek van DiFazio, Abracen & Looman (2001) naar de invloed van de behandelmodaliteit. Zij onderzochten of het geven van therapie in groepsvorm effectiever is dan het geven van therapie in individuele vorm. Doorgaans wordt aangenomen dat groepstherapie effectiever is omdat, onder andere vanwege de lagere kosten, er meer contactmomenten mogelijk zijn en de delinquent in algemene zin dus meer therapietijd krijgt. Voor sommige delinquenten is groepstherapie minder geschikt, bijvoorbeeld vanwege laag cognitief functioneren, psychiatrische problematiek (psychoses) of problemen met algemene dagelijkse vaardigheden.

DiFazio, Abracen & Looman, onderzochten seksueel delinquenten die verwezen werden naar het Regional Treatment Centre (Ontario) Sex Offender program (RTCSOP). Van de 205 delinquenten kregen er 143 het gebruikelijke volledige programma aangeboden (groepstherapie) en 62 een alternatief individueel programma. Deze selectie behoorde tot de gebruikelijke praktijk en vond plaats op basis van de eigenschappen van de delinquent. De selectie was derhalve niet random, wat een grote tekortkoming van deze studie was. Uit de resultaten bleek dat beide groepen niet verschilden in het aantal recidieven. De auteurs concludeerden hieruit dat de effectiviteit van de behandelprogramma's ondanks het grote verschil in intensiteit niet voor elkaar onderdoen en dat het aanpassen van het behandelprogramma op de individuele eigenschappen van de delinquent geen nadelig effect heeft op het behandelresultaat.

Wanneer het niet mogelijk blijkt relevante responsiviteitsfactoren mee te nemen binnen ander onderzoek, kan er ook voor gekozen worden deze middels

aparte projecten te onderzoeken. In dat geval gaat het niet primair om de vraag of de behandeling an sich effectief is, maar bijvoorbeeld om de vraag of de behandeling effectiever is als groepstherapie dan als individuele therapie. Het spreekt voor zich dat dit type onderzoek pas zin heeft wanneer een behandeling überhaupt bewezen effectief is.

Wanneer we de verschillende onderzoeksdesigns waarderen (zie hoofdstuk 6) zullen we ons met betrekking tot responsiviteit dus een tweetal vragen moeten stellen: (1) in welke mate biedt het design mogelijkheden om responsiviteitsfactoren mee te nemen binnen het effectonderzoek?; (2) in welke mate biedt het design de mogelijkheid responsiviteit an sich te onderzoeken?

## 5.2 Eisen van designs

Onderzoeksdesigns stellen verschillende eisen aan de wijze waarop het onderzoek praktisch moet worden opgezet. Hieronder volgt een korte opsomming van eisen die aan de orde kunnen zijn. De implementatie daarvan is in minder gecompliceerde settings vaak al bijzonder complex en realisering ervan binnen de tbs kan voor extra problemen zorgen. Bij de beoordeling van de onderzoeksdesigns in het volgende hoofdstuk zullen we met onderstaande aspecten rekening houden.

### 5.2.1 Randomisatie

Veel onderzoeksdesigns stellen als eis dat proefpersonen aselekt worden toegewezen aan een experimentele en controlebehandeling. Deze random toewijzing voorkomt dat er op voorhand verschillen zijn tussen beide groepen proefpersonen, waardoor het onmogelijk is het effect van de experimentele manipulatie (behandeling) gecontroleerd vast te stellen.

### 5.2.2 Placebobehandeling

In medisch onderzoek bestaat de controleconditie idealiter uit een placebomiddel. Dit is een medicijn zonder farmacologische of biologische werkzaamheid dat op het oog niet is te onderscheiden van het werkzame middel. Een placebo wordt ingezet om verwachtingseffecten bij patiënten, behandelaars en onderzoekers in de experimentele en controleconditie gelijk te houden. Voor psychologische behandelingen zijn ook placebocondities ontwikkeld. Naar analogie van de farmacologische placebo's zijn dergelijke "behandel"condities op het oog gelijk aan werkzame interventies (bijvoorbeeld een gelijk aantal contacturen met een therapeut), maar zij hebben op theoretische gronden geen invloed op de psychopathologie.

### 5.2.3 Groepsgrootte en statistische power

De statistische power<sup>4</sup> van een onderzoek is de kans om een effect aan te tonen indien het er daadwerkelijk is. Bij de statistische power wordt meestal gedacht aan de omvang van de steekproef, maar de power is feitelijk afhankelijk van drie grootheden: de grootte van het aan te tonen effect (bijvoorbeeld 10% reductie in

---

<sup>4</sup> Basaal aan wetenschappelijk onderzoek is dat er bij het trekken van een conclusie twee fouten gemaakt kunnen worden. De fout van de eerste soort is dat men ten onrechte de nulhypothese verwierpt en aanneemt dat er verschil is tussen de experimentele en de controle conditie. Meestal wordt het risico om zo'n fout te maken op minder dan 5% gesteld. De fout van de tweede soort is het omgekeerde: het ten onrechte aannemen dat de nulhypothese waar is en dat de experimentele behandeling even effectief is als de controlebehandeling. De statistische power is één min de kans op het maken van de "fout van de tweede soort" en wordt meestal op 80% gesteld. Een kans van 20% om geen effect te vinden van de experimentele manipulatie (behandeling) wordt dus acceptabel geacht.

recidive door een interventie, cf. McGuire, 2002), het aantal patiënten dat deelneemt aan het onderzoek (N) en gewenste zekerheid (significantieniveau).

De power is afhankelijk van keuzes die gemaakt worden bij het design van de studie. Vanzelfsprekend heeft de aard of kracht van de interventie invloed. Indien een interventie een groot effect sorteert, zal dit effect gemakkelijker aangetoond kunnen worden en zijn daar minder proefpersonen voor nodig. Echter, de effectgrootte van de interventie zal vaak onbekend zijn. De onderzoeker kan dan vooraf stellen dat alleen het aantonen van een substantieel verschil tussen twee behandelingen klinisch relevant is. Bijvoorbeeld, men is alleen geïnteresseerd in een experimentele interventie wanneer die leidt tot minimaal 5% meer reductie in recidive dan de behandelingen die nu in gebruik zijn. Andere factoren zijn ook van invloed op de grootte van het effect. Het samenstellen van homogene patiëntgroepen is een manier om de variantie in de gegevens terug te brengen en zo het verschil in uitkomst tussen de experimentele en controle groep te vergroten (Kraemer, 1992). Een andere methode om de power te vergroten is zorgen voor voldoende deelnemers aan de studie. Bij therapie-effectonderzoek is dit, uitgaande van een gemiddeld tot sterk effect, minstens 50 deelnemers per arm in het design (Kazdin & Bass, 1989). Betrouwbaar en sensitief meten van de therapie-uitkomst is nog een manier om de power te optimaliseren. Het is uiterst belangrijk om bij de keuze van een meetinstrument aan te sluiten bij de doelstelling van de te evalueren interventie. Hoe specifiekere meetinstrumenten gericht zijn op wat de interventie beoogt te veranderen (bijvoorbeeld concreet observeerbaar gedrag of een specifieke vaardigheid), des te gevoeliger zijn ze voor het effect van de interventie. Ten slotte is gebruik van krachtige statistische analysemethoden, die optimaal recht doen aan de aard van uitkomstvariabelen of het design van de studie, een manier om de power te vergroten (Tabachnick & Fidell, 2001).

Er worden in de literatuur verder nog creatieve oplossingen gesuggereerd om beschikbare proefpersonen zo efficiënt mogelijk in een trial in te zetten (Cohen, 1988). Zo kan de onderzoeker ongelijke groepsgrootten nastreven, zodat er optimale power is om verschillende effecten te toetsen (Woods et al., 1998).<sup>5</sup>

#### 5.2.4 *Behandelformat*

Een behandeling wordt pas goed onderzoekbaar als er een draaiboek of protocol voor is opgesteld. In dit protocol wordt de aard en de duur van de behandeling vastgelegd, alsmede alle handelingen die de behandelaar en de patiënt dienen uit te voeren. Dit draaiboek zorgt ervoor dat de behandeling op dezelfde wijze wordt uitgevoerd door verschillende behandelaars die aan een trial participeren, dat de behandeling repliceerbaar is en dat de behandeling bij bewezen werkzaamheid ook verspreid kan worden onder collega behandelaars of andere instellingen.

#### 5.2.5 *Behandelintegriteit*

De interne validiteit van het onderzoek vereist een controle op toepassing van het protocol. Men noemt dit ook wel de vaststelling van “treatment integrity” of

---

<sup>5</sup> Stel bijvoorbeeld dat een onderzoeker twee actieve behandelingen (pillen en praten) wil vergelijken met een wachtlijstcontrolegroep. Er zijn twee onderzoeksvragen: (1) zijn de actieve behandelingen werkzaam (effectiever dan niets doen) en (2) welke van beide behandelingen werkt het best (pillen of praten). Het verschil tussen de actieve behandelingen zal kleiner zijn dan het verschil tussen de beide actieve behandelingen enerzijds en de controlegroep anderzijds. Dit houdt in dat er een grotere groep deelnemers nodig is om een verschil tussen beide actieve behandelingen aan te tonen dan om het verschil tussen enerzijds de actieve en anderzijds de wachtlijstgroep aan te tonen. Dit zou dan kunnen resulteren in een 1:2:2 verhouding, wat leidt tot een toedeling van bijvoorbeeld 30 patiënten aan de wachtlijstconditie en 60 patiënten aan allebei de actieve condities.

behandelintegriteit (Nezu & Nezu, 2008). Het gaat er hierbij vooral om te bepalen of de behandeling is uitgevoerd zoals hij bedoeld is. Heeft de behandelaar alles gedaan wat is aangegeven in het protocol of draaiboek van de behandeling (geboden handelingen)? Heeft de behandelaar niet iets gedaan wat niet beschreven staat (verboden handelingen)? Is de behandeling voltooid? Heeft de patiënt zich gehouden aan het protocol, huiswerkopdrachten uitgevoerd, pillen geslikt, etc.?

Er worden in therapie-effectonderzoek vaak instrumenten ingezet om te bepalen in hoeverre de behandelaar zich aan het protocol voor de behandeling heeft gehouden, bijvoorbeeld op basis van een aselechte greep uit audio of video-opnames van alle therapiezittingen. De eis van behandelintegriteit geldt overigens niet alleen voor de experimentele behandeling, maar ook voor de controlebehandeling. In de controleconditie dienen geen elementen van de experimentele behandeling opgenomen te zijn.

Naast behandelintegriteit kan ook nog de behandelkwaliteit worden vastgesteld. Het gaat er dan niet zozeer om of alle regels van de behandeling zijn toegepast maar meer om de kwalitatieve beoordeling van de wijze van toepassing van de behandeling. Hierbij speelt de competentie van de therapeut een belangrijke rol.

#### *5.2.6 Meetcultuur*

Voor het welslagen van therapie-effectonderzoek is bij alle betrokkenen een attitude vereist die gekenmerkt wordt door een positieve opstelling ten opzichte van de toepassing van meetinstrumenten (“meten is weten”). Hiertegenover staat een cultuur waarin de inzet van meetinstrumenten als te reductionistisch wordt gezien. Een holistisch mensbeeld doet mogelijk meer recht aan de klinische werkelijkheid, maar brengt wetenschappelijk onderzoek niet verder. Meten leidt overigens pas tot meer weten wanneer goede beslissingen worden genomen over *wat* men wil meten en *hoe*. Bij therapie-effectonderzoek is de juiste keuze van de uitkomstmaat essentieel. Verder moet er zoveel mogelijk gebruik worden gemaakt van evidence-based meetinstrumenten. Meetinstrumenten verdienen het predikaat evidence-based pas als betrouwbaarheid, validiteit en nauwkeurigheid ruim gedocumenteerd zijn.

#### *5.2.7 Logistieke infrastructuur*

Voor het doen van therapie-effectonderzoek is een infrastructuur noodzakelijk ten behoeve van de logistieke eisen die dit soort onderzoek met zich mee brengt. Er moet worden toegezien op een juiste randomisatie, op de uitvoering van metingen voor en vooral na afloop van de behandeling en eventueel bij follow-up (zelfrapportage maar ook onafhankelijke beoordelaars), monitoring van de behandeling (de juiste toepassing van het behandeldraaiboek), er moeten beslissingen worden genomen over deelname van patiënten (juiste toepassing van inclusie en exclusiecriteria) en voortijdige beëindiging van behandeling. Zonder onderzoeksmedewerkers die met bovenstaande taken belast kunnen worden is methodologisch verantwoord therapie-effectonderzoek onmogelijk.

## **6. Specifieke onderzoeksdesigns voor therapie-effectonderzoek**

Om causale relaties te kunnen onderzoeken, waarbij een duidelijke relatie is tussen oorzaak en gevolg, is de Randomised Controlled Trial (RCT) dé voorgeschreven methode in klinische studies naar effect van behandelingen. Goede tweede is outcomes onderzoek, feitelijk cohort-studies waarin een groep patiënten wordt behandeld en gedurende lange tijd gevolgd. Een derde vorm van therapie-effectonderzoek zijn single case designs. In dit hoofdstuk worden deze drie typen onderzoek beschreven.

### **6.1. Randomised Controlled Trials (RCT's)**

RCT's komen voort uit een onderzoekstraditie in de medische wetenschappen waar patiënten aselekt worden toebedeeld aan een experimentele conditie die een naar verwachting werkzame behandeling bevat of een controleconditie die een niet werkzame behandeling (placebo) ontvangt. Het RCT design met de meest rigoureuze methodologische controle is het zogenoemde dubbelblind placebo gecontroleerde design. In een dubbelblinde studie weet noch de onderzoeker noch de patiënt aan wie actieve medicatie of placebo gegeven wordt. Dubbele blindering is wenselijk omdat gebleken is dat verwachtingseffecten over de behandeling van invloed zijn op het resultaat. Het onderzoeksdesign kan ook enkelblind of open zijn. In een enkelblind design weet de patiënt niet maar de behandelaar wel wie er in de experimentele of controle conditie zit. In een open studie weten zowel behandelaar als patiënt wie actieve en wie placebo behandeling krijgt.

Als de onderzochte behandeling niet medicamenteus maar psychosociaal van aard is, is een dubbelblinde opzet praktisch niet uitvoerbaar: de behandelaar zal altijd weten welke behandeling geboden wordt. Enkelblind (alleen de behandelaar weet welke behandeling controle en welke behandeling experimenteel is) is bij een psychosociale behandeling ook vaak moeilijk te realiseren, omdat het op gespannen voet staat met de plicht tot volledig informeren van de patiënt die aan wetenschappelijk onderzoek deelneemt (informed consent). Om deze problemen te omzeilen zal bij onderzoek naar een psychosociale behandeling de controlegroep meestal bestaan uit treatment as usual (TAU) of zullen twee experimentele behandelingen met elkaar worden vergeleken. Je weet dan echter niet wat het effect van de behandeling in vergelijking met niets doen is. Een ander probleem bij het onderzoeksdesign met een TAU conditie is dat positieve verwachtingen van de patiënt over de baten van de nieuwe behandeling een ongewenste positieve invloed kunnen hebben op de resultaten van de trial. Verwachtingseffecten van de patiënt kunnen in ieder geval zoveel mogelijk tegengegaan worden door in de voorlichting over de trial te benadrukken dat het nog niet bekend is of de onderzochte behandeling een positief effect zal hebben.

#### *6.1.1 Eisen die bij RCT's gelden*

*6.1.1.1 Randomisatie.* Om de werkzaamheid van een interventie aan te tonen is het experimentele design waarbij patiënten aselekt worden toegewezen aan de behandelcondities een sine-qua-non (Fink, 2005; Shadish, Cook & Campbell, 2002). Quasi-experimentele designs (bijvoorbeeld matched controls of ongecontroleerde cohort studies) zijn methodologisch inferieur, maar second best (Campbell & Stanley, 1966; Shadish, 2000). Een alternatief voor randomisatie op patiëntniveau is randomisatie op instellingsniveau. Patiënten van instelling A, D en F ontvangen de experimentele behandeling en patiënten van instelling B, C en E ontvangen de

controlebehandeling, waarbij toewijzing van instellingen aan de experimentele of controle behandeling at random heeft plaatsgevonden. Voorwaarde is dan wel dat deelnemende instellingen onderling niet sterk verschillen in behandel aanbod en effectiviteit.

*6.1.1.2 Placebobehandeling.* Een RCT vereist een vergelijkingsconditie waartegen de resultaten van de experimentele behandeling afgezet kunnen worden. Dat kan een wachtlijstcontrolegroep zijn, maar daarmee wordt niet gecontroleerd voor verwachtingseffecten van patiënt en behandelaar. De verwachting dat een behandeling een positief effect heeft, kan op zich al therapeutisch werken. Een placebobehandeling kan voor deze verwachtingseffecten controleren. Ook kan gekozen worden voor een TAU conditie, maar deze biedt minder garanties tegen verwachtingseffecten.

*6.1.1.3 Statistische power.* Omwille van de power moet men er voor zorgen dat de verschillende groepen in een RCT voldoende groot van aantal zijn. Wat voldoende groot is hangt vervolgens onder meer af van het te verwachten effect. In het meeste klinisch effectonderzoek, waar we te maken hebben met middelgrote effecten, zijn minstens 50 personen per experimentele arm of conditie noodzakelijk om voldoende power te hebben om een verschil aan te tonen (Kazdin & Bass, 1989). In de tbs-sector zullen de klinische effecten van interventies niet groter zijn, eerder kleiner. Voor het aantonen van een recidiveverschil zal een aanzienlijk groter aantal personen nodig zijn per conditie dan wanneer men een effect op het niveau van behandel doelen, bijv. psychotische symptomen, wil aantonen. Zo zijn bij een verschil in recidive tussen de experimentele en de controlegroep van 10% er 252 deelnemers per conditie nodig die de interventie afmaken (Wartna, 2005).

*6.1.1.4 Behandelformat.* Voor de evaluatie van de werkzaamheid van een experimentele interventie moet elke deelnemer aan de experimentele poot van de trial dezelfde interventie hebben ondergaan. Dat betekent voor de praktijk van het onderzoek dat voor een RCT een tijdsduur van de interventie wordt gekozen die niet langer is dan wat bij de gemiddelde patiënt vereist is. Als een veel langere tijdsduur van de interventie wordt gekozen, die gebaseerd is op een inschatting van de benodigde therapieduur bij de meest veranderingsresistente groep, zou dat impliceren dat patiënten die snel herstellen daarna nog geruime tijd de interventie ondergaan, terwijl ze al hersteld zijn. De tijdsduur van het pretest-posttest interval in de controle conditie van de trial (de vergelijkinggroep) moet bovendien bij benadering even lang zijn als de tijdsduur van de interventie in de experimentele conditie. Gevolg van het bovenstaande is dat in een RCT doorgaans een wat verdunde of ingekorte interventie wordt onderzocht in vergelijking met wat de behandeling in de klinische praktijk zou behelzen. In de klinische praktijk omvat de behandeling doorgaans twee keer zo veel zittingen (de Beurs et al., 2001). Een tweede consequentie is dat de uitkomst van een RCT weinig zegt over de effectiviteit van de interventie bij meer complexe patiënten.

*6.1.1.5 Vaststelling van behandelintegriteit.* Voor een RCT is het van eminent belang dat de geëvalueerde behandeling zuiver is toegepast. De specifieke kenmerken of interventies van een behandeling worden in een protocol of draaiboek vastgelegd. Zo'n protocol vergemakkelijkt een gestandaardiseerde toepassing van de behandeling tijdens de trial en maakt, bij gebleken werkzaamheid van de interventie, toekomstige verspreiding van de behandeling mogelijk. De behandelingen die geëvalueerd worden dienen dus vastgelegd te zijn in protocollen. Ook aandacht verdienen de

ondersteunende materialen bij een therapie-uitkomststudie, zoals informatiemateriaal voor de experimentele en controle behandeling en informatie over de trial.

*6.1.1.6 Responsiviteitsonderzoek.* RCT's en responsiviteitsonderzoek staan in zekere zin op gespannen voet met elkaar. In een RCT zal een onderzoeker variantie ten gevolge van patiëntkenmerken of variaties in de omstandigheden van de interventie zoveel mogelijk beperkt willen houden ten behoeve van de interne validiteit van de studie. Dat komt tot uiting in het nastreven van een zo homogeen mogelijke groep van deelnemers aan de trial (via uitgebreide inclusie- en exclusiecriteria) en zoveel mogelijk controle over de interventie middels protocollen en gestandaardiseerde meetprocedures. Toch zal er variantie zijn in de trial en met de uitkomstgegevens van een trial kan exploratief onderzocht worden voor wie en onder welke omstandigheden de interventie het beste werkte. Voor een formele toetsing van deze onderzoeksvraag is echter weer een nieuwe trial vereist, bijvoorbeeld met een case-control design.

### *6.1.2 Toepasbaarheid en realiseerbaarheid van eisen RCT's binnen tbs*

*6.1.2.1 Randomisatie.* Blinding zal om voor de hand liggende redenen alleen bij medicatie-onderzoek een optie zijn. Aselect toewijzen (randomisatie) van patiënten aan verschillende therapiecondities binnen eenzelfde instelling is eveneens moeilijk. Het zal al snel bekend zijn binnen de instelling dat deelname aan een nieuwe experimentele behandeling mogelijk is.<sup>6</sup>

Een andere optie is te randomiseren over instellingen. We stellen ons hierbij een multi-center trial voor waarbij de helft van de deelnemende instellingen een nieuwe interventie implementeert en de andere helft treatment-as-usual aanbiedt. Bij gebleken werkzaamheid van de interventie kunnen de controle instellingen tot implementatie overgaan. In zo'n opzet wordt aselect bepaald welke instelling wel en welke niet de te onderzoeken behandeling inzetten. We moeten er dan wel vanuit kunnen gaan dat er geen relevante systematische verschillen zijn tussen de deelnemende instellingen (bijvoorbeeld in afdelingsklimaat en bejegening door behandelaars of in hun opleidingsniveau), wat meestal niet het geval zal zijn. Tevens moeten er voldoende instellingen deelnemen aan het onderzoek om succesvol en effectief te kunnen randomiseren. Ten slotte zullen de uitkomstgegevens met gepaste statistische technieken moeten worden geanalyseerd om recht te doen aan de onderlinge afhankelijkheid van gegevens van patiënten die uit dezelfde instelling afkomstig zijn. Multi-level analyse is de aangewezen methode wanneer er systematische onderlinge afhankelijkheid in de data zit. We moeten concluderen dat blinding en randomisering beperkt toepasbaar en realiseerbaar is in de tbs.

*6.1.2.2 Placebobehandeling.* Behandeling met een medicamenteus placebo is voor een beperkte tijd mogelijk; echter niet voor lange duur omdat het aanbieden van een placebobehandeling op gespannen voet staat met de verplichting een actieve behandeling aan te bieden aan tbs-gestelden (zie ook noot 6). Ontwikkeling van een

---

<sup>6</sup> Eerder (zie 4.2.1) is opgemerkt dat patiënten op grond van toeval deelname aan de nieuwe behandeling ontzeggen, terwijl zij in de tbs verblijven in het kader van een maatregel ter behandeling, rechtsongelijkheid met zich meebrengt ten opzichte van de deelnemers aan de experimentele conditie. Een wachtlijstcontrolegroep biedt mogelijk enige uitkomst, maar dat kan alleen als de experimentele behandeling (en daarmee de wachtlijst) van beperkte duur is en niet langer dan ongeveer drie maanden duurt, een situatie die zich niet snel voordoet in de tbs.

psychotherapieplacebo dat op het oog niet gemakkelijk te onderscheiden is van een actieve behandeling, is moeilijk.

*6.1.2.3 Behandelintegriteit.* In hoeverre kan bij therapie-effectonderzoek in de tbs-sector behandelintegriteit van de onderzochte interventie in een trial gewaarborgd worden? Behandelintegriteit kan een heikel punt zijn, omdat een experimentele behandeling niet los kan worden onderzocht van het “groepstherapeutisch” klimaat dat in een tbs-setting bestaat: isolatie van de interventie van het therapeutisch klimaat kan lastig zijn. Bovendien zijn voor RCT's vanwege de vereiste groepsgrootte meestal meerdere instellingen nodig. Over meerdere instellingen een voldoende behandelkwaliteit blijven garanderen is een extra complicerende factor.

*6.1.2.4 Behandelformat.* Een RCT vereist dat de uitkomst van de behandeling met dezelfde tijdsintervallen wordt gemeten in de experimentele en controleconditie. Als de controleconditie een wachtlijstcontrolegroep is, dient zich de vraag aan hoe lang een potentieel werkzame interventie uitgesteld kan worden. In de praktijk van therapie-effectonderzoek is een therapieduur van drie tot vier maanden gebruikelijk en wordt een wachtlijstconditie van die duur nog net acceptabel geacht. Designs met een wachtlijstcontrolegroep kunnen in de tbs-setting dus alleen gebruikt worden voor de evaluatie van relatief kortdurende interventies. Bij een langere therapieduur, bijvoorbeeld een behandeling voor persoonlijkheidsstoornis van 18 maanden, wordt als vergelijkingsgroep een andere behandeling ingezet, zoals wordt geïllustreerd met de studie van Bernstein (zie Kader 2).

*6.1.2.5 Statistische power.* De tbs-sector is met 1800 patiënten relatief klein en bestaat uit een heterogene populatie (qua psychopathologie, delict en duur waarvoor veroordeeld, mate van intelligentie, etc.). De te verwachten effecten van interventies zullen over het algemeen niet groot zijn. Onder optimale omstandigheden zijn 50 personen per behandelarm noodzakelijk om voldoende power te hebben, maar voor een RCT in de tbs zal waarschijnlijk een grotere groep deelnemers nodig zijn. RCT's doen derhalve een zeer groot beroep op de tbs-sector. In de praktijk betekent dit dat er gelijktijdig slechts een zeer beperkt aantal RCT's kan worden opgezet.

Ten slotte dient opgemerkt te worden dat het verblijf in de tbs-kliniek niet vrijwillig is. Deelname aan een onderzoekstrial tijdens de tbs-behandeling is dat in principe wel, maar in de context van de tbs-maatregel kan men zich afvragen in hoeverre er toch niet sprake is van dwangbehandeling. Naast ethische bezwaren heeft dit potentieel een negatieve invloed op de uitkomst van behandeling en kan het leiden tot kleinere behandel-effecten dan momenteel gevonden worden in de reguliere geestelijke gezondheidszorg (en daarmee tot een geringere statistische power van het onderzoek).



## Kader 2

### *Illustratief voorbeeld van RCT onderzoek in de tbs-setting*

Door David Bernstein en collega's werd in samenwerking met de Universiteit Maastricht (Prof. Dr A. Arntz) in 2006 een multi-center gerandomiseerde gecontroleerde studie opgezet bij patiënten met cluster B persoonlijkheidsstoornissen in verschillende forensische psychiatrische instellingen in Nederland. Het onderzoek verkeert momenteel in de uitvoeringsfase. Doel van de studie is een antwoord te verkrijgen op de vraag of schema-focussed therapy een persoonlijkheidsverandering en lager recidive kan bewerkstelligen bij forensische patiënten. Uit eerder onderzoek in Nederland van Giessen-Bloo et al. (2006) is gebleken dat deze therapievorm werkzaam is bij ambulante patiënten met een Borderline Persoonlijkheidsstoornis in de ggz.

In de studie van Bernstein et al. worden 120 patiënten met cluster B persoonlijkheidsstoornissen (Antisociale, Narcistische, Borderline, of Paranoïde Persoonlijkheidsstoornis) gerandomiseerd naar twee behandelcondities: schematherapie of de gebruikelijke forensische behandeling (TAU). Patiënten krijgen drie jaar lang individuele behandeling. In de schematherapie-conditie worden de patiënten twee keer per week gezien, conform de richtlijnen voor schematherapie bij ernstige persoonlijkheidsstoornissen (vgl. Young, Klosko & Weishaar, 2003). Het hulpverleningscontact in de gebruikelijke behandelingconditie is vastgesteld op een keer per week (wat de gebruikelijke frequentie is voor individuele behandeling in Nederlandse forensische instituten). De gebruikelijke behandeling in forensische instituten is multi-modaal: patiënten krijgen doorgaans zowel individuele als groepstherapie, samen met educatie, rehabilitatie en andere diensten.

Primaire uitkomstmaten in deze studie zijn verandering in persoonlijkheidsstoornissen en risico op recidive, wat iedere zes maanden gemeten wordt gedurende de behandeling. Algemene psychopathologie, en veranderingen in onderliggende vroege maladaptieve schema's en schemamodi zijn de secundaire uitkomstmaten. Nadat het drie jaar durende onderzoek afgerond zal zijn, zullen de patiënten nog drie jaar gevolgd worden om hun feitelijke recidive vast te stellen. Bij degenen die nog steeds gedwongen in een forensisch instituut verblijven, zal geweld binnen de instelling onderzocht worden.

Op dit moment nemen vijf forensische instellingen deel aan het onderzoek. De drie jaar durende behandelstudie wordt naar verwachting beëindigd in 2011, de follow-upstudie in 2014.

Ons inziens zijn een tweetal bedenkingen bij het studie design te plaatsen. Ten eerste is er enige twijfel over de adequaatheid van de controleconditie TAU. Deze bestaat uit wekelijks een uur psychotherapie, terwijl de experimentele interventie uit twee maal per week een uur schema focussed therapie bestaat. Dit verschil laat straks de mogelijkheid open dat grotere effectiviteit van de schema-focussed therapie te danken is aan meer aandacht voor de patiënt en niet interventie specifieke elementen en zo beperkt het de interne validiteit van de studie. Ten tweede wordt er in de studie niet gerandomiseerd over behandelaars, waardoor het risico ontstaat dat de experimentele behandeling uitgevoerd wordt door de meest kundige en enthousiaste behandelaars van de instelling. Dit beperkt de externe validiteit van de studie.

## 6.2 Outcomes onderzoek

### 6.2.1 Beschrijving voornaamste kenmerken

Het voornaamste kenmerk van outcomes onderzoek is dat men het effect meet van klinisch handelen in de praktijk van alledag, hierbij gebruikmakend van in wetenschappelijk effectonderzoek ontwikkelde instrumenten en methodologie. In principe wordt bij alle patiënten die bij een instelling onder behandeling zijn periodiek de klachten of het functioneren gemeten en wordt vastgelegd wat de behandeling voorafgaande aan de meting was. Doel van dit soort onderzoek is ondermeer na te gaan in hoeverre behandelingen, die onder gecontroleerde omstandigheden van wetenschappelijk effectonderzoek werkzaam zijn gebleken, nog effectief zijn als ze in de klinische praktijk worden uitgevoerd. Hiertoe wordt alle patiënten in een instelling herhaaldelijk gedurende hun behandeling met dezelfde instrumenten gemeten en wordt vastgelegd wat de behandeling inhield, zodat vastgesteld kan worden welk effect hiermee wordt gesorteerd.

De Medical Outcomes Study (MOS; Wells et al., 1989) was een van de eerste studies naar therapie-effect in de klinische praktijk. Dit onderzoek evalueerde het effect van de behandeling van stemmingsstoornissen zowel in de eerstelijnszorg bij de huisarts als in de tweede lijn bij gespecialiseerde ggz-instellingen. Bij bijna 1800 patiënten met een stemmingsstoornis werd de mate van detectie van depressie en de uitkomst van de behandeling vastgesteld. De conclusie luidde dat veel stemmingsstoornissen niet onderkend werden en over het algemeen niet goed behandeld werden. Sinds de MOS is er nog een aantal studies verschenen waarbij gebruik werd gemaakt van gegevens die verzameld werden in het administratieve proces rond intake of behandeling, zoals biografische gegevens, diagnostische status en duur van de behandeling en uitkomst. De stand van zaken in Engeland in 2002 is samengevat door Gilbody, House en Sheldon (2002). Zij concluderen dat dergelijke administratieve gegevens eigenlijk ongeschikt zijn voor beantwoording van vragen omtrent de effectiviteit van de klinische praktijk. De kwaliteit van de data is vaak omgekeerd evenredig met de omvang van de database.

In de VS gebruiken Lambert en collega's (2001) de term "patient-focused" onderzoek voor routinematig en frequent meten van het beloop van klachten. Bij monitoring volgens het model van Lambert wordt na iedere zitting gemeten. Lambert en collega's hebben voor dit doel een korte vragenlijst ontwikkeld, de Outcome Questionnaire (OQ). Deze lijst van 45 vragen wordt voorafgaand aan iedere zitting door de patiënt ingevuld. Met dit instrument worden drie aspecten van de psychopathologie gemeten: symptomen van vooral depressie en angst, interpersoonlijk functioneren (partner, familie, vrienden) en het maatschappelijk functioneren (baan, opleiding). Bespreking van het beloop van scores op deze drie gebieden heeft een prominente rol in de behandeling. Op basis van het niveau van de klachten bij de eerste meting met de OQ kan een verwacht beloop van de klachten uitgezet worden. Wanneer de patiënt bij een vervolgmeting buiten de bandbreedte van het verwachte beloop scoort, gaat er een waarschuwingssignaal uit naar de behandelaar. Lambert et al. (2005) rapporteren gunstige resultaten van deze feedback aan behandelaar en patiënten op basis van de outcome monitoring. In een gecontroleerde studie knapten patiënten waarbij wekelijks het klachtniveau werd gemeten en teruggerapporteerd sneller op dan controle patiënten waarbij alleen een begin- en eindmeting plaatsvond zonder terugkoppeling van de bevindingen aan de behandelaar.

In Duitsland wordt sinds het midden van de jaren '90 in enkele instellingen therapie-effect gemeten volgens het Stuttgart-Heidelberg Model (Kordy, Hannover &

Richard, 2001). Klachten worden aan het begin van de behandeling en zes maanden later gemeten met een beperkte set meetinstrumenten (onder meer de Symptom Checklist, de Global Assessment of Functioning schaal en de Inventory for Interpersonal Problems). De gegevens worden opgenomen in een database en vervolgens op drie niveaus gebruikt: het systeem levert feedback aan de patiënt, de behandelaar en op geaggregeerd niveau aan het management van de instelling. Bijzonder aan het systeem is dat er een alarmeringsfunctie is ingebouwd: indien er aan bepaalde voorwaarde wordt voldaan (bijvoorbeeld sterke aanwijzingen voor suïcidaliteit of onvoldoende voortgang geboekt in een voorgaande periode) genereert het systeem een alarmsignaal aan de behandelaar.

*6.2.1.1 Outcome Monitoring in de Nederlandse ggz.* In Nederland wordt sinds 2001 Routine Outcome Monitoring (ROM) uitgevoerd in Leiden bij Rivierduinen, een grote instelling voor geestelijke gezondheidszorg in Zuid-Holland. ROM bij Rivierduinen houdt in dat de uitkomst van behandelingen in de klinische praktijk vastgesteld wordt bij patiënten met stemmings-, angst- en somatoforme stoornissen in ambulante zorg. Onderzoeksverpleegkundigen meten bij deze patiënten periodiek allerlei aspecten van hun psychisch functioneren met vragenlijsten en beoordelingsschalen. Aan de hand van de meetgegevens wordt bepaald of bij een patiënt de symptomen afnemen en het functioneren verbetert over de tijd. De uitkomstgegevens worden iedere drie maanden verzameld en er vindt terugkoppeling plaats aan de behandelaar en de patiënten.

*6.2.1.2 Doel van ROM.* Door toepassing van ROM wordt transparantie gecreëerd ten aanzien van de kwaliteit en efficiëntie van de behandeling met als uiteindelijk doel de kwaliteit van de behandeling in de ggz te vergroten (de Beurs en Zitman, 2007). Dezelfde functie zou ROM ook in de tbs-sector kunnen hebben. Door ROM wordt de effectiviteit van de behandeling bij individuele patiënten, de effectiviteit van behandelaars, teams of zelfs de gehele instelling zichtbaar. Door aan de hand van ROM over voldoende patiënten voormetinggegevens en therapieuitkomstgegevens te verzamelen wordt het mogelijk onderzoek te doen naar de klinische effectiviteit van specifieke interventies bij bepaalde gespecificeerde subgroepen patiënten.

Er zijn nog verschillende subdoelen bij ROM die relevant kunnen zijn voor de tbs. Ten eerste is dat feedback bieden aan de individuele behandelaar en diens patiënt over het verloop van de klachten ter ondersteuning van het klinisch handelen. Er wordt gezocht naar een optimale afstemming van meetmomenten met het klinisch handelen, zodat gegevens beschikbaar zijn op het moment dat er een beslissing genomen moet worden over voortzetting, verandering of afsluiting van de behandeling. Een tweede doel is om het management van een instelling meer inzicht te bieden op de effectiviteit van handelen in (sub)divisies van de instelling of van individuele therapeuten. Een derde doel is een “meetvriendelijke” cultuur in de instelling tot stand te brengen, waarin uitspraken over het klinische functioneren van patiënten ondersteund worden met de uitkomsten van gestandaardiseerde en gevalideerde meetinstrumenten.

*6.2.1.3 ROM en RCT's.* Het voornaamste verschil tussen ROM en RCT onderzoek is dat het bij ROM het geen voorwaarde is dat er een controlegroep wordt gecreëerd. Ook worden er bij ROM geen stringente inclusie en exclusie criteria gehanteerd. Verder bedient ROM zich van dezelfde onderzoeksmethoden en techniek als RCT's. Het is dan mogelijk om binnen een infrastructuur die voor ROM is opgezet RCT's te

organiseren. Randomisatie naar experimentele en controlebehandeling wordt dan een aanvullende voorwaarde bij het onderzoek.

*6.2.1.4 ROM en Responsiviteitsonderzoek.* Outcomes onderzoek is bij uitstek geschikt voor de evaluatie van responsiviteitsvariabelen. Namelijk, doordat i.t.t. bij RCT's geen stringente inclusiecriteria worden gehanteerd, zal een grote(re) variatie aan patiënten, behandelcondities etc. geïnccludeerd worden. Dit maakt het mogelijk om de prognostische waarde van patiëntkenmerken of kenmerken van het therapeutisch proces te onderzoeken. Voor ROM zijn de strenge waarborgen ten behoeve van de interne validiteit van RCT's niet strikt noodzakelijk. Eigenlijk is alleen vereist dat de relevante variabelen in de monitoring worden meegenomen. Na verloop van tijd zijn er dan voldoende gegevens om te onderzoeken of er een betekenisvol verband is tussen een responsiviteitsvariabele en de uitkomst van een interventie.

## *6.2.2 Eisen die bij outcomes onderzoek gelden*

*6.2.2.1 Randomisatie en placebobehandeling.* Randomisatie en het vergelijken met een placebobehandeling zijn geen eisen die outcomes onderzoek stelt.

*6.2.2.2 Voldoende groepsgrootte.* Outcomes onderzoek is vooral gericht op externe validiteit of generaliseerbaarheid van bevindingen uit wetenschappelijk onderzoek in de praktijk. Specifiek aan outcomes onderzoek is dan ook dat het belangrijk is om zo veel mogelijk patiënten die onder behandeling zijn te includeren en te meten. Dit is vooral om vertekening van de resultaten vanwege selectieve weigering van bijvoorbeeld “moeilijke mensen” tegen te gaan. Alleen dan zijn de resultaten representatief voor bijvoorbeeld de instellingspopulatie waaruit zij afkomstig zijn. Naast de generaliseerbaarheid neemt de betrouwbaarheid van de bevindingen ook toe wanneer de uitkomsten afkomstig zijn van groepen patiënten van voldoende omvang. Tenslotte vereist responsiviteitsonderzoek voldoende deelnemers om replicerbare resultaten te genereren. Als vuistregel voor het aantal deelnemers geldt dat 5 tot 10 keer het aantal te onderzoeken responsiviteitsvariabelen aan personen geïnccludeerd moet worden (Stevens, 2002).

*6.2.2.3 Behandelformaat en behandelintegriteit.* Outcomes onderzoek stelt geen specifieke eisen aan het behandelformaat, in die zin dat de behandeling van beperkte duur dient te zijn of “aan- en uitgezet” moet kunnen worden. De behandeling kan gewoon conform de klinische praktijk worden toegepast. Dit is echter wel afhankelijk van de voorliggende onderzoeksvraag en de doelstelling waarmee ROM gebruikt wordt. Die doelstelling kan zijn monitoring van de effectiviteit van de bestaande behandelpraktijk, maar kan ook zijn de (ongecontroleerde) evaluatie van nieuwe experimentele behandelingen. Ongeacht de doelstelling is een goede monitoring waaruit de behandeling heeft bestaan in de periode tussen meetmomenten essentieel. ROM bestaat altijd uit de combinatie van uitkomstgegevens en documentatie van de behandeling.

*6.2.2.4 Meetcultuur.* Een aantal van de eisen die gelden bij goed RCT onderzoek gelden ook bij outcomes onderzoek, zoals het gebruik van betrouwbare, valide en genormeerde meetinstrumenten. De behandeling die tussen de meetmomenten wordt gegeven moet goed gedocumenteerd worden om te weten wat geëvalueerd wordt. Net als bij RCT onderzoek moeten instrumenten ingezet worden om behandelintegriteit vast te stellen. Verder is het van belang met een voldoende hoge

frequentie te meten en/of op klinisch relevante momenten te meten. Meetinstrumenten moeten zijn toegesneden op patiëntengroepen en op de inhoud van interventies en kunnen variëren per meetmoment.

*6.2.2.5 Infrastructuur voorhanden.* Dataverzameling, databeheer en softwarematige ondersteuning zijn belangrijke punten van aandacht. Voor het doen van effectonderzoek dient men te beschikken over een infrastructuur om periodiek gegevens over het functioneren (op klacht- of symptoomniveau of over gedrag of incidenten) van de patiënt te meten en deze gegevens te beheren. De ervaring leert dat hiervoor het beste medewerkers aangesteld kunnen worden die speciaal met deze taak belast worden. Voorts dient het proces van gegevensverwerking met goede ICT ondersteund te worden, bijvoorbeeld door software in te zetten, waarmee meetinstrumenten kunnen worden afgenomen, die direct feedback biedt aan de behandelaar in de vorm van genormeerde uitslagen. Voor software ondersteuning is een Elektronisch Patiënten Dossier nodig dat flexibel ingericht kan worden met meetinstrumenten.

### *6.2.3 Toepasbaarheid en realiseerbaarheid van eisen outcomes onderzoek binnen tbs*

*6.2.3.1 Randomisatie en placebobehandeling.* Deze eisen zijn, zoals eerder vermeld, voor outcomes onderzoek niet van toepassing en dus geen belemmering voor het toepassen van deze methode in het tbs-kader.

*6.2.3.2 Voldoende groepsgrootte.* Om betrouwbare resultaten te verwerven zijn ook bij outcomes onderzoek een voldoende aantal deelnemers nodig, zeker wanneer resultaten van subgroepen (naar type delict, type stoornis of behandeling) in ogenschouw genomen worden. Dit is echter goed te realiseren in outcomes onderzoek, omdat er geen strenge in- en exclusiecriteria gelden zoals bij RCT's en geen vergelijkingsgroep hoeft te worden samengesteld.

*6.2.3.3 Behandelformat en behandelintegriteit.* Outcomes onderzoek heeft geen specifieke eisen aan behandelformat en behandelintegriteit anders dan de gebruikelijke praktijk. De tbs-praktijk levert op dit vlak dan ook geen belemmeringen. Monitoring wat op het gebied van behandeling voorvalt, is wel een belangrijk integraal onderdeel van outcomes onderzoek.

*6.2.3.4 Meetcultuur.* Het tot stand brengen of versterking van een reeds aanwezige meetcultuur wordt de laatste jaren in de tbs-sector al nagestreefd en heeft in ieder geval geleid tot tbs-brede implementatie van routinematige risicotaxatie. Behandel-evaluatie met instrumenten die qua meetpretentie dichter bij interventies staan, wordt wel op sommige instellingen maar nog niet sectorbreed toegepast.

In de tbs-sector zijn en worden projecten reeds gerealiseerd die outcomes onderzoek benaderen of in ieder geval het uitvoeren van outcomes onderzoek mogelijk maken. Elke instelling is wel op de een of andere manier bezig met het monitoren van hun patiënten. Het Monitoring Informatiesysteem TBS (MITS) project van DJI valt in dit kader ook te noemen. De ervaring met MITS leert ook dat het niet eenvoudig is om een goed systeem van monitoring te implementeren en te onderhouden. Verzamelen van “meta-data”, gegevens over de kwaliteit van de monitoring data en de dekkingsgraad van de monitoring kunnen helpen om zicht te krijgen en te houden op hoe succesvol de implementatie verloopt. Hoeve Boschoord heeft recent, in samenwerking met twee andere tbs-instellingen, een systeem van

monitoring opgezet waarbij onder meer agressie-incidenten worden bijgehouden. Daarnaast wordt halfjaarlijks de Adult Behavior Checklist en jaarlijks de HKT-30 afgenomen.

Een succesvolle implementatie van routine outcome monitoring leidt tot een “meetcultuur” in de instelling. Om zo’n meetcultuur tot stand te brengen is coöperatie van behandelaars noodzakelijk en zij dienen dan ook hun voordeel te kunnen doen met de gegevens die gegenereerd worden. Zij zullen vooral baat hebben bij uitkomstgegevens van individuele patiënten die direct betrekking hebben op hun interventie en die sturend kunnen zijn bij de verdere planning van behandeling. Een systeem als MITS, wat alleen mogelijkheden tot registratie achteraf heeft, biedt onvoldoende garantie voor een complete gegevensverzameling. Wanneer eenmaal een meetcultuur tot stand is gebracht, kan die naast outcomes onderzoek ook andere vormen van onderzoek ondersteunen. Een positieve bijkomstigheid van outcomes onderzoek is dat een RCT zich eenvoudiger laat implementeren als men al gewend is aan uitgebreide metingen in de instelling. Outcomes onderzoek faciliteert ook gegevensverzameling voor single case design.

Samenwerking tussen meerdere tbs-instellingen is wenselijk om in staat te zijn relatief snel voldoende gegevens te verzamelen omtrent de effectiviteit van specifieke interventies en te onderzoeken voor wie die interventies vooral geschikt zijn (responsiviteit). Ten behoeve van de onderlinge vergelijkbaarheid van gegevens en samenwerking tussen instellingen is dan een voorwaarde dat er een kernset wordt samengesteld van meetinstrumenten die door iedereen gedeeld wordt, zo mogelijk sector-breed. Een voordeel van onderlinge afstemming is dat niet elke instelling het wiel opnieuw hoeft uit te vinden. Ook maakt afstemming van instrumenten onderlinge vergelijking van effectiviteit mogelijk. Ten slotte voorkomt afstemming informatieverlies bij overplaatsing van een tbs-gestelde. Een voorbeeld dat illustreert dat vruchtbare samenwerking mogelijk is in het tbs-veld, is het tot stand komen van een inventarisatie van de geboden behandelingen ten behoeve van het creëren van een behandelingsactiviteiten-module in het MITS. Voortbouwend op MITS zou er een registratiesysteem voor Routine Outcome Monitoring gemaakt moeten worden met een kernset van meetinstrumenten waarover in de tbs-sector consensus is dat dit nuttige “prestatie-indicatoren” zijn van tbs-behandeling.

*6.2.3.5 Infrastructuur voorhanden.* In de tbs is nog geen specifieke infrastructuur voorhanden ten behoeve van het routinematig meten. Hierbij moet gedacht worden aan getraind personeel dat metingen kan uitvoeren en ICT ondersteuning. Wel zijn bij veel instellingen enkele wetenschappelijk onderzoekers actief die hierbij betrokken kunnen worden.

#### *6.2.4 Bedreiging van de interne validiteit*

Uitspraken doen over de werkzaamheid van een therapie(vorm) op grond van observationele gegevens is niet het primaire doel van outcomes onderzoek, maar is niet onmogelijk. Het grootste probleem dat optreedt wanneer er geen willekeurige toewijzing aan therapiecondities heeft plaatsgehad, zoals bij outcomes onderzoek het geval is, is vertekening van de uitkomsten door zogenaamde confounders. Confounders zijn variabelen die een effect hebben op de therapie-uitkomst en die niet evenredig zijn verdeeld over de verschillende behandelingen (bijvoorbeeld, de assertieve patiënten hebben vooral behandeling X gekregen, de non-assertieve vooral behandeling Y.) Er zijn voorstellen gedaan in de methodologische onderzoeksliteratuur over hoe met potentiële confounders om te gaan bij uitkomstgegevens uit een observationele studie, zoals de “propensity score

modelling” strategie. De “propensity score modelling” strategie is een elegante statistische methode voor het analyseren van data verkregen met observationeel onderzoek waarbij aselecte toewijzing ontbreekt (Hullsieck & Louis, 2002).<sup>7</sup>

Ten slotte noemen we nog het onderzoek van Ferriter and Huband (2005). Zij vergeleken de uitkomsten van RCT’s en gecontroleerde, maar niet gerandomiseerde, studies in het forensische veld en concludeerden dat de effectgroottes die in beide typen onderzoek gerapporteerd worden aan elkaar gelijk waren. Hun conclusie is dat goed uitgevoerde gecontroleerde studies waarbij randomisatie ontbreekt toch een belangrijke bijdrage kunnen leveren aan de evidence base van het forensisch veld.

#### *6.2.5 Conclusie outcomes onderzoek*

Eerder werd in de beschrijving van outcomes onderzoek gesteld dat er een aantal methodologische eisen gelden voor dit type onderzoek. Dit betreft onder meer een hoog insluitingpercentage, uitgebreide documentatie van behandeling en een goed meetinstrumentarium. Al deze eisen zijn in principe realiseerbaar in het tbs-veld. Outcomes onderzoek in de tbs-sector zal vooral neerkomen op juiste assessment en registratie. Het opzetten van een systeem voor het routinematig verzamelen van therapie-uitkomstgegevens maakt ook ander onderzoek zoals RCT’s en single case studies mogelijk. Bij gebruik van beoordelingsschalen of event sampling is medewerking van de patiënt niet strikt noodzakelijk. Een insluitingpercentage van 100% is dan in principe haalbaar. Uitgebreide en gedetailleerde documentatie van de behandeling kan tot stand komen gelijktijdig met de invoering van zorgprogramma en de wettelijke verplichting om Diagnose Behandelcombinaties te registreren. Het toenemende gebruik van geprotocolleerde behandeling zal ook bijdragen aan een betere documentatie van de aard van de behandeling die wordt geëvalueerd.

### **6.3 Single case designs**

#### *6.3.1 Beschrijving voornaamste kenmerken.*

Hoewel RCT’s en outcomes onderzoek de meest aangewezen methoden zijn voor effectstudies, kennen deze methoden ook een aantal praktische bezwaren. Met name randomisatie bij het toewijzen van behandelingen aan proefpersonen lijkt, hoewel niet per definitie onmogelijk, in een tbs-kader zeer lastig. Een ander probleem waar zowel RCT’s als outcomes onderzoek mee te kampen hebben, is het feit dat de behandeling in de tbs zeer specifiek op het individu is toegesneden. Daardoor is het vaak onmogelijk groepen samen te stellen van voldoende patiënten die een vergelijkbare combinatie van behandelmodules heeft ondergaan.

Om tegemoet te komen aan deze praktische problemen zijn een aantal alternatieve experimentele onderzoeksdesigns ontwikkeld, die bekend staan onder de naam single case designs. In plaats van grootschalige studies met grote onderzoekspopulaties, kenmerken single case designs (ook wel N=1 studies genoemd) zich door een sterke focus op het individu. In plaats van groepen wordt

---

<sup>7</sup> Er wordt bij de propensity score methode rekening gehouden met potentiële confounders of matchingsvariabelen door voor iedere deelnemer te berekenen wat de kans is om tot de behandelde of de controlegroep te behoren op basis van de confounding variabelen. Dit resulteert in een propensity score. Bij de uiteindelijke vergelijking van de resultaten wordt rekening gehouden met deze score, bijvoorbeeld door het als een covariaat mee te nemen in de analyse. Een gedetailleerde uitwerking van de methode valt buiten het bestek van dit rapport en we verwijzen de geïnteresseerde lezer naar de publicatie van Hullsieck en Louis (2002).

het gedrag binnen één individu longitudinaal gevolgd en vergeleken op verschillende momenten, namelijk tijdens non-interventieperioden en interventieperioden.

Het is een misverstand de single case methode gelijk te stellen aan casusbeschrijvingen waarin anekdotisch veranderingen bij een patiënt worden geschetst. Bij een casusbeschrijving is geen sprake van een systematische observatie, waardoor een mogelijke conclusie dat gedragsverandering het gevolg is van een behandeling niet gestaafd kan worden. Het single case design, daarentegen, gaat uit van een experimentele onderzoeksopzet met een systematische manipulatie van variabelen. Hiermee wordt getracht de interne validiteit te waarborgen, zodat oorzakelijke verbanden wel aangetoond kunnen worden. Een ander misverstand is dat er bij single case studies slechts één proefpersoon wordt gebruikt. Vaak worden er echter meerdere proefpersonen gebruikt, al ligt de nadruk altijd op de gedragsveranderingen bij het individu.

Er zijn grofweg twee soorten single case designs te onderscheiden: demonstratiedesigns en vergelijkingsdesigns (Tervo, Estrem, Bryson-Brockmann & Symons, 2003). Demonstratiedesigns hebben als doel het effect aan te tonen van (verschillende componenten van) één specifieke behandeling. Vergelijkingsdesigns zijn met name bruikbaar voor het vergelijken van verschillende behandelingen of het bestuderen van de interactie tussen behandeling en context (zoals responsiviteitsfactoren). Er is in de loop der jaren een rijke schakering aan specifieke onderzoeksdesigns binnen de single case methodologie ontwikkeld, variërend van eenvoudig tot zeer verfijnd.

*6.3.1.1 Demonstratiedesigns.* Het meest eenvoudige single case onderzoeksdesign is het AB design, waarbij bij een individu de periode voorafgaand aan de behandelinterventie wordt vergeleken met de periode tijdens of na de interventie. Bij dit design kunnen gemeten gedragsveranderingen toegeschreven worden aan de interventie indien er aan een aantal condities wordt voldaan. Deze condities zijn onder andere stabiliteit van het gedrag, objectieve meting, continue en herhaalde meting, een duidelijke verandering in de trend op het moment van de interventie en repliceerbaarheid bij verschillende individuen. Helaas wordt slechts zelden aan alle condities voldaan waardoor een oorzakelijk verband alsnog lastig aan te tonen is. Het AB design functioneert daardoor vooral als hypothesegenererend.

Het ABA design, ook wel 'withdrawal' design genoemd, voegt na de interventieperiode een periode toe waarin de interventie aan het individu wordt onttrokken. Dit design maakt het mogelijk de werking van een interventie aan te tonen, wanneer de gedragsverandering optreedt in de interventieperiode, maar weer verdwijnt in de nieuwe non-interventieperiode. Dit design is niet geschikt wanneer de behandeling een blijvend effect heeft zoals bij het aanleren van vaardigheden. Vaak wordt er overigens gekozen voor een ABAB design, aangezien het om ethische redenen niet verantwoord is het experiment af te sluiten met een non-interventie periode (zie ook Kader 3 over het onderzoek van Burns et al., 2003).

Een multiple baseline design is een specifieke vorm van het AB design. In tegenstelling tot het AB design wordt er bij een multiple baseline design niet slechts één pre-interventieperiode geïntroduceerd, maar meerdere, bijvoorbeeld bij meerdere individuen. In dit geval meet men bij verschillende individuen gedurende een pre-interventieperiode. Vervolgens vangt de behandeling voor individu 1 aan, maar blijven de andere individuen nog in de baselineperiode. Pas wanneer er een gedragsverandering gemeten wordt voor individu 1 start de behandeling bij individu 2. Het idee achter het multiple baseline design is dat een gedragsverandering bij individu 1 pas dan aan de behandeling kan worden toegeschreven wanneer deze



verandering niet ook tegelijkertijd bij individu 2 optreedt, die nog in de baseline conditie verkeert.

*6.3.1.2 Vergelijkingsdesigns.* Bij een cross-over design worden twee behandelingen (A en B) onderling met elkaar vergeleken door bij een individu met een vastgestelde regelmaat de behandelinterventies af te wisselen. Daarvoor kan een ABAB variant gekozen worden, hoewel het de voorkeur heeft de sequentie van behandelingen te randomiseren (bijvoorbeeld ABBABAAB) of twee deelnemers de behandelingen in omgekeerde volgorde aan te bieden (deelnemer 1 ABAB en deelnemer 2 BABA). Voor dit design geldt net als voor het ABA design dat de effecten van de behandeling reversibel moeten zijn en is daardoor vooral geschikt om medicamenteuze behandelingen te evalueren. Dit gebeurt momenteel in de forensische context in FPK De Kijvelanden.

Het alternating treatments design houdt in dat twee soorten interventies zeer snel achter elkaar worden afgewisseld. Wanneer behandeling A effectiever is dan behandeling B, verwacht men dat de gemeten uitkomstmaten tijdens perioden A een andere patroon of trend laten zien dan tijdens perioden B. Een vereiste voor dit design is wel dat de interventie een snel effect dient te hebben. Het voordeel van dit design boven een cross-over design (welke een minder snelle afwisseling kent) is dat het alternating treatments design minder gevoelig is voor veranderingen in de omgeving aangezien aangenomen wordt dat de context minder snel varieert dan de experimentele conditie. Hiermee wordt weliswaar de interne validiteit verbeterd, maar de externe validiteit is problematisch. In een normale klinische setting vindt een dergelijke snelle afwisseling namelijk niet plaats en bovendien is niet denkbeeldig dat de twee behandelingen elkaar beïnvloeden.

*6.3.1.3 Methodologische aspecten.* Goed opgezette single case studies zijn wetenschappelijk gerespecteerde onderzoeksdesigns. De American Psychological Association (APA) bijvoorbeeld, heeft criteria opgesteld voor empirisch gevalideerde behandelingen en accepteert daarbij expliciet bewijs verkregen uit single case experimentele designs (Chambless, 1998). De APA onderscheidt “well established treatments” waarvoor ondermeer geldt dat de effectiviteit moet zijn aangetoond door middel van ófwel ten minste twee between groups experimentele designs ófwel een groot aantal (meer dan negen) single case designs. Daarnaast definieert de APA “probably efficacious treatments” waarvoor een kleiner aantal single case studies voldoet (drie of meer). Ook de eerdergenoemde Cochrane Collaboration accepteert bewijs uit single case designs, zij het als een lager niveau van bewijs (Davies, Howells & Jones, in press; Cochrane website)

In vergelijking met niet-experimentele designs zoals outcomes onderzoek zijn single case designs relatief goed in staat oorzakelijke verbanden aan te tonen (interne validiteit). Problematisch is wel dat hierbij enkele aannames gemaakt moeten worden waar niet altijd aan voldaan wordt (zoals beschreven bij het AB design, te weten stabiliteit van het gedrag, objectieve meting, continue en herhaalde meting, een duidelijke verandering in de trend op het moment van de interventie en replicerbaarheid bij verschillende individuen). De complexiteit van de gemaakte aannames maakt een juiste interpretatie soms extra ingewikkeld. Een ander aandachtspunt is de juiste en gedisciplineerde uitvoering van het onderzoek, waar de interne validiteit van met name single case designs zeer gevoelig voor is.

Hoewel de interne validiteit van single case designs, mits zorgvuldig opgezet, acceptabel genoemd mag worden, kan de externe validiteit problematisch zijn. Om de effectiviteit van een behandeling gemeten bij slechts enkele subjecten te kunnen

generaliseren over een grotere groep, wordt onder andere aangeraden de context en setting van de onderzochte behandeling zo gedetailleerd mogelijk te beschrijven, en de resultaten zoveel mogelijk te interpreteren in een bestaande theorie.

*6.3.1.4 Responsiviteitsfactoren.* Responsiviteitsfactoren, factoren die (mede) bepalen hoe effectief een behandeling is bij een specifiek individu (zie par. 5.1.3), zijn moeilijk te onderzoeken met single case designs. Single case designs zijn in feite (soms gerandomiseerde) gecontroleerde experimentele trials. De beperkingen die bij RCT's genoemd worden met betrekking tot het includeren van externe responsiviteitsfactoren gelden evenzeer voor single case designs. Daarbij komt dat het intra-subject karakter van de single case methodologie het onmogelijk maakt interne factoren zoals behandelmotivatie te variëren. Gegeven deze beperkingen zijn single case designs niet de eerstaangewezen methoden voor het bestuderen van responsiviteit. Wel kunnen externe responsiviteitsfactoren eventueel middels een single case design onderzocht worden door een behandeling afwisselend onder verschillende condities aan te bieden (bijvoorbeeld een cross-over design waarin een interventie afwisselend als groepstherapie en als individuele therapie wordt aangeboden).

### *6.3.2 Eisen die bij single case designs gelden*

Single case designs zijn in essentie een experimenteel onderzoeksdesign. In tegenstelling tot outcomes onderzoek, waar de nadruk ligt op het periodiek meten van de klinische praktijk, wordt er bij single case designs actief ingegrepen in de behandelpraktijk. Er zullen dus patiënten en behandelaars gevonden moeten worden die bereid zijn de behandeling aan experimentele voorwaarden aan te passen. Ethische bezwaren treden hier overigens zelden op, zolang de patiënt niet een behandeling onthouden wordt en een eventuele verlengde baselineperiode kort duurt. De ethische toelaatbaarheid kan zelfs als een van de sterke punten worden beschouwd in relatie tot andere experimentele onderzoeksdesigns zoals RCT's, waar een controlegroep geformeerd moet worden die geruime tijd een potentieel superieure behandeling wordt onthouden.

*6.3.2.1 Randomisatie.* Single case designs vereisen geen randomisatie van proefpersonen. Soms wordt echter wel geëist dat behandelcondities worden gerandomiseerd, bijvoorbeeld bij het afwisselen van behandelcondities in een cross-over design.

*6.3.2.2 Placebobehandeling.* Evenals bij RCT's geldt dat een behandeling idealiter vergeleken wordt met een niet-werkende placebobehandeling, door deze behandelingen altemeer toe te passen bij dezelfde persoon (zie ook paragraaf 6.1).

*6.3.2.3 Groepsgrootte.* Single case designs zijn ontwikkeld om met weinig proefpersonen valide uitspraken te kunnen doen. Echter, er zijn daarom wel een flink aantal herhaalde metingen nodig, ongeveer 5 á 10 tijdens zowel de baseline periode als de interventie periodes. In tegenstelling tot wat wel wordt beweerd kan er niet worden volstaan met één voormeting en enkele nametingen, aangezien het beperkte aantal proefpersonen ( $N=1$ ) gecompenseerd dient te worden met veel meetmomenten teneinde het effect van meetfouten te beperken. Een goede single case studie vereist daarom een goede logistiek, vergelijkbaar met RCT's.

6.3.2.4 *Behandelformat, behandelintegriteit, meetcultuur en infrastructuur.* De eisen in gevallen van single case designs zijn gelijk aan die bij RCT's.

### 6.3.3 *Realiseerbaarheid eisen binnen tbs-kader*

Op bijna alle onderdelen is de realiseerbaarheid van single case designs vergelijkbaar met RCT's, met uitzondering van de vereiste randomisatie, groepsgrootte en behandelintegriteit.

6.3.3.1 *Randomisatie.* Omdat er geen randomisatie over proefpersonen vereist wordt, vervallen de problemen die bij RCT's op dit punt gelden. Het gerandomiseerd afwisselen van behandelinterventies met alternatieve interventies of periodes zonder interventies is op individueel niveau ook binnen een tbs-kader goed mogelijk.

6.3.3.2 *Groepsgrootte.* Het grote voordeel van de single case designs is natuurlijk dat er geen tot nauwelijks eisen gesteld worden aan de groepsgrootte. Het is praktisch goed uitvoerbaar meerdere studies binnen één instelling op te zetten en uit te voeren (hoewel een multi-center aanpak tot een betere externe validiteit leidt). Single case designs kunnen derhalve een plaats hebben naast RCT's en outcomes onderzoek bij het komen tot een evidence-based behandelprogramma, mits talrijk en goed opgezet. Gegeven alle genoemde beperkingen, echter, zullen single case designs niet in de plaats kunnen komen van RCT's en outcomes onderzoek.

6.3.3.3 *Behandelintegriteit.* Het in stand houden van een voldoende behandelintegriteit is bij RCT's binnen het tbs-kader moeilijk omdat er vanwege de vereiste groepsgrootte doorgaans multi-center onderzocht zal moeten worden. Dit nadeel heeft een single case design niet omdat een dergelijk onderzoek zich juist heel goed leent om binnen één instelling opgezet te worden.

### Kader 3

#### *Voorbeeld van Single Case Design: Anger Management Training*

In een Engels onderzoek door Burns, Bird, Leach en Higgins (2003) werd een single case design gevolgd teneinde het effect van Anger Management Training aan te tonen onder een populatie forensische patiënten met leerproblemen. Onderzocht werd zowel het effect op de subjectieve woedebeleving als op het gedrag van de proefpersonen.

De studie gebruikte in totaal vijf proefpersonen, waarvan er twee gaandeweg het onderzoek afvielen. De resterende drie werden onderworpen aan een ABA design. Deze keuze werd door de onderzoekers gerechtvaardigd door te wijzen op het feit dat dit de aangewezen methode is wanneer men het effect van een behandeling op een specifieke stoornis wil onderzoeken en er hierdoor te weinig deelnemers voorhanden zijn voor een klassieke groepsvergelijkende methode. Een multi-center benadering werd door hen afgewezen omdat daarmee problemen ontstaan met het matchen van therapeutische omgeving, behandelfilosofie en benadering door behandelaars.

De totale onderzoeksduur was 28 weken, verdeeld over drie perioden van 8 weken pretest assessment, 12 weken behandeling en 8 weken post-test assessment. De proefpersonen werden wekelijks gemeten op een drietal schalen (Novaco Anger Scale, Spielberger State-Trait Anger Expression Inventory en Modified Overt Aggression Scale). Tijdens de behandelperiode werd alleen gemeten in week 3 tot 10 om voor alle drie de perioden acht meetmomenten te creëren. De resultaten werden statistisch geanalyseerd door middel van time series analysis en de Combined S test, zowel op de groep als geheel als op de drie patiënten individueel.

De resultaten voor de Novaco Anger Scale werden grafisch overgenomen in een figuur (zie Burns et al., 2003), waaruit op het oog sprake lijkt te zijn van een lichte verbetering in scores tijdens de behandelperiode. Dit werd bevestigd door de statistiek die liet zien dat er geen trend is binnen de periodes, maar wel significante verschillen tussen de periodes. De auteurs concludeerden dat Anger Management Training bruikbaar kan zijn voor de onderzochte populatie, hoewel zij meer en langduriger follow-up onderzoek nodig achten voordat er verdergaande conclusies getrokken kunnen worden omtrent de blijvendheid van veranderingen en generaliseerbaarheid van de resultaten.

## 6.4 Samenvatting en conclusie

In de voorgaande twee hoofdstukken zijn zowel algemene eigenschappen van onderzoeksdesigns (hoofdstuk 5) als specifieke onderzoeksdesigns aan de orde gekomen (hoofdstuk 6). In deze paragraaf worden de verschillende punten nogmaals kort besproken en in hun onderlinge samenhang samengevat (zie ook Tabel 6.1 en 6.2).

De tegenstelling tussen interne en externe validiteit heeft geleid tot twee typen therapie-effectonderzoek die elkaar aanvullen: experimenteel onderzoek en outcomes onderzoek (Well, 1999; De Beurs et al., 2001). Bij RCT's worden patiënten door het lot aan twee of meer behandelingen toegewezen, een nieuwe experimentele behandeling en een controlegroep, bijvoorbeeld een wachtlijstgroep of een groep die een placebobehandeling krijgt. Voorafgaande en na afloop van de behandeling worden de klachten, symptomen of gedragingen gemeten en naderhand wordt onderzocht welke groep gemiddeld de grootste klachtenreductie of gedragsverandering heeft behaald. De randomised controlled trial is de voorkeursmethodiek bij "efficacy"-onderzoek, d.w.z. onderzoek naar de werkzaamheid van behandelingen. Een goed uitgevoerde RCT heeft een hoge interne validiteit en staat toe dat er conclusies kunnen worden getrokken m.b.t. de experimentele conditie(s) zonder dat er alternatieve verklaringen mogelijk zijn. Single case designs zijn evenals RCT's experimentele designs, waarbij proefpersonen onder gecontroleerde omstandigheden een behandeling krijgen aangeboden. In tegenstelling tot RCT's worden bij single case designs geen groepen met elkaar vergeleken, maar wordt gekeken naar veranderingen binnen het individu. De empirische bewijskracht van single case designs is minder, omdat niet is aangetoond dat de resultaten zijn te generaliseren naar andere patiënten.

In outcomes onderzoek wordt de behandeling en het verloop van de pathologie gemonitord, door op gezette tijden te meten wat de behandeling beoogt te veranderen (klachten, psychopathologie, ongewenst gedrag, recidiverisico). Bij outcomes onderzoek staat de externe validiteit centraal. De vraag staat centraal in hoeverre de bevindingen uit experimenteel onderzoek te generaliseren zijn naar de praktijk, waarbij de omstandigheden minder gecontroleerd zijn dan bij experimenteel onderzoek. Aan outcomes onderzoek kleven een aantal methodologische bezwaren, maar wanneer aan bepaalde kwaliteitseisen wordt voldaan, levert deze onderzoeksmethode waardevolle gegevens. Bij deze kwaliteitseisen valt te denken aan een hoog insluitingpercentage (er wordt getracht van zoveel mogelijk patiënten gegevens te verkrijgen om vertekening vanwege selectieve weigering van bijvoorbeeld "moeilijke mensen" tegen te gaan), uitgebreide documentatie van behandeling en toepassing van de juiste meetinstrumenten voor het vaststellen van kenmerken van de deelnemers, hun behandeling en de uitkomst van de behandeling (betrouwbaar, valide en sensitief).

Dat er strenge eisen worden gesteld aan meetinstrumenten voor het vaststellen van het effect van de behandeling (ergo aan het instrumentarium om de afhankelijke variabelen te meten) geldt voor alle typen onderzoek.

### 6.4.1 Externe validiteit/effectiviteit

Doordat outcomes onderzoek nauwkeurig de alledaagse praktijk registreert, is dit het aangewezen design wanneer het aankomt op het onderzoeken van de effectiviteit van behandelingen in de praktijk. Outcomes onderzoek heeft derhalve een hoge externe validiteit. Uitkomsten van experimentele onderzoeken, RCT's en single case designs, zijn beperkter generaliseerbaar naar behandelcondities en patiënten die men aantreft

inde praktijk door de nadruk op de strenge criteria waaraan de behandeling (bijvoorbeeld zeer goed getrainde behandelaren) en de proefpersonen (alleen patiënten zonder co-morbiditeit) moeten voldoen.

#### *6.4.2 Interne validiteit/werkzaamheid*

Op basis van outcomes onderzoek kunnen tot op zekere hoogte causale uitspraken gedaan worden. Echter, voor het aantonen van de werkzaamheid van een behandeling is het essentieel dat proefpersonen onder gecontroleerde en experimentele condities aan een behandeling worden onderworpen. Single case designs zijn hier ook toe in staat, maar de bevinding is in principe niet generaliseerbaar. Uiteindelijk blijft de RCT op dit gebied de sine qua non, omdat dit design optimale experimentele controle toestaat én de resultaten generaliseerbaar zijn naar vergelijkbare patiënten.

#### *6.4.3 Responsiviteit*

RCT's en single case designs stellen strikte eisen aan de uitvoering van het onderzoek, bijvoorbeeld in hoe de behandeling wordt uitgevoerd of dat een relatief homogene patiëntengroep wordt geïncludeerd. Hierdoor zijn deze designs relatief slecht in staat om responsiviteitsfactoren mee te nemen in therapie-effectonderzoek. In het design tracht men juist variatie in responsiviteitsvariabelen terug te brengen onder controle te brengen van de onderzoeker. Dit geldt met name voor externe responsiviteitsfactoren waarin vanwege behandelintegriteit niet gevarieerd kan worden. Outcomes onderzoek is, zolang men voldoende en de juiste meetinstrumenten inzet om de responsiviteitsfactoren te meten, bij uitstek in staat het effect van dergelijke factoren te onderzoeken.

#### *6.4.4 Realiseerbaarheid van eisen in tbs-kader*

De tbs is een relatief kleine sector, waarbinnen zich een heterogene populatie bevindt. Dit brengt beperkingen met zich mee wat betreft de mogelijkheden tot onderzoek. Met name onderzoeksdesigns die veeleisend zijn wat betreft de uitvoeringscondities en bovendien vereisen dat de praktijk zich aan het onderzoek aanpast zijn hierdoor lastiger uitvoerbaar.

Randomisatie van tbs-gestelden over interventiecondities ten behoeve van een RCT kan lastig zijn. Ook het aantal proefpersonen nodig voor een goede RCT is problematisch; er zijn al snel meer dan 100 deelnemers nodig met een bepaalde pathologie of die een bepaalde interventie ondergaan (50 per onderzoeksconditie in een twee-armig design). Op het eerste gezicht lijkt 100 deelnemers op een totaal van 1800 tbs-gestelden niet veel, maar de praktijk wijst uit dat vanwege selectie van geschikte deelnemers, weigering na aanbod deel te nemen aan de trial of drop-out tijdens de trial de aanvankelijke pool van potentiële deelnemers een veelvoud moet zijn van het benodigde aantal patiënten dat de behandeling afrondt. Veel meer dan één RCT op hetzelfde moment binnen de tbs zal dan ook niet mogelijk zijn. Daarnaast stellen RCT's eisen aan de behandeling, met name beperkingen aan de duur ervan, wat moeilijk te realiseren zal zijn in tbs-verband. Een goede behandelintegriteit is realiseerbaar, maar het feit dat RCT's multi-center zullen moeten worden opgezet om voldoende patiënten te kunnen includeren, vormt hiervoor een bedreiging. Het is immers moeilijk ervoor te zorgen dat verschillende behandelaars op verschillende locaties de behandeling op de voorgeschreven wijze uitvoeren, en de controle door de onderzoeksleider daarop is bij een multi-center studie ingewikkeld.

Single case designs zijn qua opzet vergelijkbaar met RCT's (beide zijn experimentele onderzoeksopzetten). De hierboven genoemde punten van RCT's gaan dus ook op voor dit type onderzoek, met uitzondering van de randomisatie, behandelintegriteit en groepsgrootte. De toepassingsmogelijkheden van single case designs zijn echter beperkt. Het behandelformat (met name het in korte episodes herhaaldelijk en afwisselend variëren in behandelconditie of de behandeling “aan” en “uit” zetten) ligt ver van de praktijk en is bij veel behandelingen, zoals groepstherapie, niet realiseerbaar. Het feit dat het mogelijk is waarlijk experimenteel onderzoek te doen bij enkele proefpersonen is één van de sterke punten.

Outcomes studies liggen qua inhoud zeer dicht tegen de praktijk en zijn daardoor zeer goed te realiseren. Om binnen een afzienbare periode over voldoende tbs-gestelden (die bijvoorbeeld een bepaalde interventie hebben ondergaan of behandeld zijn voor een specifieke pathologie) is samenwerking tussen de verschillende tbs-klinieken wel een voorwaarde. Praktische belemmeringen die te verwachten zijn binnen het veld hebben betrekking op het tot stand brengen van consensus met betrekking tot te gebruiken meetinstrumenten. Consensus is belangrijk om over voldoende onderling vergelijkbare gegevens te kunnen beschikken. Ook op het gebied van de infrastructuur zal geïnvesteerd moeten worden, onder andere in software en uitvoerend personeel.

Samenvattend concluderen wij dat door de implementatie van outcomes onderzoek waardevolle gegevens verworven zullen worden omtrent de effectiviteit van interventies in de tbs. Met outcomes onderzoek wordt een eerste belangrijke stap gezet naar een evidence-based benadering. Outcomes studies zijn superieur op het gebied van externe validiteit en komen daarmee het dichtst bij het beantwoorden van de vraag naar de effectiviteit van de tbs-behandelingen. Bovendien leidt het implementeren van outcomes onderzoek tot een verbetering van de meetcultuur en dat kan ertoe bijdragen dat in de toekomst RCT's en single case designs makkelijker uitgevoerd kunnen worden. Uiteindelijk zal voor het invullen van de evidence base ook experimenteel onderzoek noodzakelijk zijn dat daadwerkelijk de werkzaamheid aantoonst van interventies onder gecontroleerde onderzoeksomstandigheden. Dergelijk onderzoek vergt een lange adem en zal op korte termijn (de komende jaren) geen substantiële bijdrage aan de evidence base kunnen leveren.

Tabel 6.1. Overzicht van eigenschappen van drie typen therapie-effectonderzoek

	RCT	Outcomes onderzoek	Single Case
<b>Eigenschappen</b>			
Externe validiteit / ( <i>klinische</i> ) effectiviteit	-	+	-
Interne validiteit / werkzaamheid	+	-	+
Responsiviteitsfactoren als onderdeel van effectiviteitsstudie			
Interne factoren	+/-	+	+/-
Externe factoren	-	+	-
Responsiviteit als zelfstandige studie	+	+	+

*Opmerkingen.* + betekent eigenschap van het design; - betekent geen eigenschap van het design.

Tabel 6.2. Realiseerbaarheid van eisen aan onderzoeksdesign in tbs-kader

	RCT	Outcomes onderzoek	Single Case
Randomisatie	-	n.v.t.	+
Placebobehandeling	-	n.v.t.	-
Voldoende groepsgrootte	- <sup>1</sup>	+/- <sup>1</sup>	+
Behandelformat	- <sup>2</sup>	+	- <sup>2</sup>
Behandelintegriteit	+/- <sup>3</sup>	+	+
Infrastructuur voorhanden?	+	+/- <sup>4</sup>	+

*Opmerkingen.* + betekent makkelijk realiseerbaar in de tbs; - betekent lastig realiseerbaar in de tbs.

Noten:

1. Om een effect van een interventie aan te tonen zijn zowel bij RCT's als outcomes onderzoek een voldoende aantal proefpersonen nodig (met een minimale ondergrens van 50 personen per onderzochte conditie). Voldoende deelnemers is in outcomes onderzoek makkelijker te realiseren, omdat er geen strenge inclusie en exclusie criteria gelden.
2. Experimentele studies zijn vooral geschikt voor kortdurende interventies (wanneer gebruik wordt gemaakt van een wachtlijst of placebobehandeling, bij TAU (controle) conditie is de duur minder van belang). Het is vaak moeilijk in de tbs een behandeling hierop aan te passen.
3. RCT's vereisen een hoge behandelintegriteit (behandelaars houden zich aan het protocol en er komen in de controlebehandeling geen elementen van de experimentele behandeling aan bod). Op zich is hier goed aan te voldoen. Binnen de tbs echter dienen RCT's multi-center te worden opgezet, wat een extra belemmering kan vormen voor gelijkwaardige kwaliteit en behandelintegriteit tussen de verschillende instellingen.
4. Onder andere op ICT gebied en meetinstrumenten is er nog onvoldoende structuur voorhanden.



## **7. Conclusie en aanbevelingen**

### **7.1 Outcomes onderzoek**

In hoofdstuk 6 hebben we drie onderzoeksdesigns besproken en geconcludeerd dat, gezien de specifieke kenmerken van de tbs, outcomes onderzoek het best realiseerbaar is en waardevolle therapie-uitkomst gegevens oplevert voor het opbouwen van de evidence base binnen deze sector. Outcomes onderzoek of de implementatie van Routine Outcome Monitoring (ROM) betekent dat in een quasi-experimenteel design interventies worden geëvalueerd, zoals die worden toegepast in de praktijk van alledag. Daarnaast leent outcomes onderzoek zich bij uitstek voor onderzoek naar responsiviteitsvariabelen. De eerste aanbeveling luidt dan ook:

#### **Aanbeveling 1: Implementeer outcomes onderzoek (ROM)**

Om tot ROM te komen zal de bestaande meetcultuur in de tbs verder ontwikkeld moeten worden. De tbs-sector zal baat hebben bij het opzetten van infrastructurele voorzieningen voor het systematisch monitoren van therapie-relevante variabelen. Zo'n infrastructuur bestaat minimaal uit gestandaardiseerde diagnostiek en indicatiestelling, monitoring van het effect en van responsiviteitsvariabelen en het vastleggen van de aard van de behandeling.

### **7.2 Kleinschalige RCT's en Single Case Designs**

Kleinschalige RCT's en single case studies zijn een waardevolle aanvulling op outcomes onderzoek. Zij bieden meer empirische ondersteuning voor de werkzaamheid van interventies dan ongecontroleerde studies kunnen bieden. Single case studies vormen daarbij een eerste test waarmee, met de inzet van weinig middelen, toch een indruk gekregen kan worden of een interventie potentiële werkzaamheid heeft. Uiteindelijk leveren experimentele designs, en dan met name RCT's, de meest overtuigende empirische ondersteuning voor de werkzaamheid van behandelingen en daarmee de benodigde evidence-base voor zorgprogrammering. De tweede aanbeveling is dan ook om, gebruik makend van de infrastructuur van outcomes onderzoek, single case designs en RCT's uit te voeren:

#### **Aanbeveling 2: Voer kleinschalig experimenteel onderzoek uit (RCT's en Single Case Designs)**

RCT's in de tbs-sector zullen niet te groot en niet te talrijk kunnen zijn. De omvang van de tbs-sector is in vergelijking met de ggz te beperkt voor meerdere grootschalige RCT studies.<sup>8</sup> Omdat maar een zeer beperkt aantal studies tegelijkertijd ondernomen kunnen worden, moet er een weloverwogen keuze gemaakt worden met betrekking tot de te evalueren interventie en met betrekking tot de aard van de pathologie waarop deze zich richt of de beoogde doelgroep van de behandeling.

---

<sup>8</sup> De wenselijkheid van relatief kleinschalige RCT's blijkt ook uit de ervaring met de trial die op dit moment wordt uitgevoerd in de tbs-sector naar de werkzaamheid van Schema Focused Therapy van David Bernstein en collega's (zie kader bij 6.1) Deze studie illustreert een aantal van de problemen die de uitvoering van een grootschalige RCT in de tbs-sector met zich mee kan brengen. De opzet van de studie is uitstekend, maar bij de uitvoering moeten vanwege de klinische realiteit in instellingen toch hier en daar concessies aan het design worden gedaan.

Single case designs zijn relatief snel en kleinschalig uit te voeren. De uitkomsten geven een eerste idee wat potentieel werkzaam is en wat niet. Zij kunnen dus een goed voortraject voor een uitgebreidere RCT zijn. Voorts is bij heel specifieke onderzoeksgroepen (bijvoorbeeld iemand met een psychotische stoornis die een seksueel delict heeft gepleegd) het single case design feitelijk het enige wat praktische uitvoerbaar is. De methodologie van single case designs gaat uit van interventies die “aan” en “uit” gezet kunnen worden. Dat impliceert dat voor veel problematiek en interventies in de tbs-sector dit design niet bruikbaar is.

Therapie-effectonderzoek is vooral geschikt om de werkzaamheid of effectiviteit van specifieke interventies vast te stellen, afgemeten aan concrete uitkomsten die direct betrekking hebben op de doelstelling van de onderzochte interventie. Wat betreft te onderzoeken interventies verdient het dan ook aanbeveling dat men zich richt op specifieke behandelingen met een welomschreven doelstelling, bijvoorbeeld gericht op toename van sociale vaardigheden, impulscontrole en assertiviteit.

## 7.3 Systematische monitoring

### 7.3.1 Therapie-effect monitor

Ongeacht het gekozen onderzoeksdesign is het belangrijk uitgebreid stil te staan bij de selectie van instrumenten voor het meten van de uitkomst van de behandeling. Eerder werd vastgesteld dat er gekozen kan worden voor proximale of distale uitkomstmaten, waarbij de eerste korte termijn veranderingen meet die de interventie beoogt te bewerkstelligen (bijvoorbeeld toegenomen vaardigheden), terwijl de tweede bedoeld is voor effecten die verder verwijderd zijn in de tijd en de interventie (recidive reductie). Meetinstrumenten kunnen daarnaast onderscheiden worden op grond van wie de bron van informatie is (zelfrapportage, beoordeling door een behandelaar of beoordeling door een onafhankelijke, eventueel “blinde”, beoordelaar) en de aard van wat gemeten wordt (bijv. emoties, cognities en daadwerkelijk gedrag).

Zelfrapportage instrumenten voor het bepalen van therapie-effect, de gebruikelijke uitkomstmaat bij therapie-effectonderzoek in de ggz, lijken bij onderzoek in de tbs-sector beperkt inzetbaar. Zelfbeoordeling zal bij voorkeur moeten worden aangevuld door beoordelingsschalen en *event monitoring*. Een studie wordt methodologisch sterker wanneer de uitkomst van de behandeling vanuit verschillende gezichtspunten (patiënt, behandelaar, onafhankelijk beoordelaar) wordt gezien.

Overwegingen omtrent meetinstrumenten:

- generieke versus behandelingspecifieke instrumenten;
- (combinatie van) zelfrapportage, beoordelingsschalen en event monitoring;
- monitoring van uitval.

### 7.3.2 Behandelingen monitor

Een belangrijk onderdeel van de voorgestelde infrastructurele voorziening voor ROM is een goede registratie van de aard en toediening van behandeling. Hiervoor is een inventarisatie nodig van de behandelingen die geboden worden in de tbs-sector. Voor deze inventarisatie zal in de toekomst mogelijk gebruik gemaakt kunnen worden van het werk van de Erkenningscommissie Gedragsinterventies Justitie onder voorzitterschap van Prof. Dr. W. van Tilburg. Op patiëntniveau moet een registratie worden bijgehouden van welke interventies worden toegepast. Daarnaast moet bij de tbs-patiënt ook vastgesteld worden of en in welke mate een aangeboden behandeling daadwerkelijk uitgevoerd en ondergaan is (treatment integrity).

Vaststelling van de treatment integrity is een belangrijke methodologische eis aan therapie-effectonderzoek. Het gaat hierbij dan om grove indicatoren zoals wel of geen deelname aan de sessies, maar ook om subtielere variabelen zoals de (steekproefsgewijze) vaststelling van therapiekwaliteit (bijvoorbeeld ervaringsniveau van de behandelaar en opleidingsniveau in relatie tot de complexiteit van de behandeling) en therapietrouw (is de behandeling door de behandelaar uitgevoerd zoals bedoeld, te beoordelen door onafhankelijke beoordelaars aan de hand van therapieverslagen of registraties van therapiezittingen, heeft de patiënt zich gehouden aan de therapeutische opdrachten of het medicatie regime, heeft de patiënt de behandeling afgemaakt, etc.).

Overwegingen omtrent behandelingenmonitor:

- keuze relevante aspecten behandelintegriteit;
- implementeren monitoring van behandelverloop op basis van bovenstaande relevante aspecten.

### 7.3.3 Meten van responsiviteitsvariabelen

Ten behoeve van responsiviteitsonderzoek zullen zowel de relevante interne als externe responsiviteitsfactoren moeten worden geïnventariseerd. Relevante interne factoren, zoals intelligentie, motivatie voor behandeling, veranderbaarheid, kunnen bij binnenkomst van de patiënt, en eventueel gedurende de behandeling, vastgelegd worden. Externe factoren, welke betrekking hebben op de omstandigheden waaronder de interventie toegediend of ontvangen werd, zullen gedurende de behandeling kunnen worden vastgelegd. Bij responsiviteitsonderzoek kan het dus gaan om elementen van de behandeling (externe factor), maar ook om de attitude van de patiënt zoals getoonde inzet bij de behandeling (interne factor). Voor het laatste zijn instrumenten ontwikkeld (Drieschner, 2005). Elementen van de behandeling (bijvoorbeeld patiënt-therapeut relatie) kunnen ook gemeten worden met standaard instrumenten, maar kunnen ook afhangen van de specifieke interventie. In het laatste geval is bij het te gebruiken instrumentarium maatwerk geboden.

Overwegingen omtrent responsiviteitsvariabelen:

- inventarisatie van de meest relevante interne en externe factoren;
- verzamelen van meetinstrumenten.

## 7.4 Samenwerking

### 7.4.1 Onderlinge samenwerking

Het gezamenlijk uitvoeren van ROM in de tbs-sector biedt grote (schaal)voordelen. De omvang van de patiëntenpopulaties met overeenkomstige pathologie is in individuele tbs-instellingen te klein om onderzoek naar therapie-effect te doen en noopt tot landelijke samenwerking van instellingen in de sector (Multi-center trials). Bovendien levert een landelijke implementatie van ROM het voordeel op dat er geen informatieverlies optreedt bij overplaatsing van patiënten naar andere instellingen. Dat hoeft niet in de weg te staan dat instellingen eigen wensen en bestaande onderzoekslijnen onderbrengen in ROM.

Behalve samenwerking tussen tbs-klinieken onderling, kan samenwerking met de reguliere geestelijke gezondheidszorg en daaraan gelieerde academische centra een positief effect hebben op therapie-effectonderzoek in de tbs-sector. Voor de ggz zijn zorgprogramma's ontwikkeld voor stoornissen die ook veelvuldig in de tbs-setting voorkomen. Evaluatie van die zorgprogramma om hun toepasbaarheid voor de tbs-

setting te beoordelen, zal bijdragen tot de gewenste evidence base. Verder is er in de ggz ruime ervaring voorhanden op het gebied van therapie-evaluatie onderzoek.

De onderzoekers in de tbs zijn enthousiast over onderlinge samenwerking, wat onder meer tot uiting komt in bestaande samenwerkingsovereenkomsten en in de grote opkomst bij het halfjaarlijkse onderzoekerverleg dat wordt georganiseerd door Dr. J. Harte. Ook het Expertisecentrum Forensische psychiatrie (EFP) zou mogelijk een rol kunnen spelen bij het tot stand komen van samenwerkingsverbanden.

#### *7.4.2 Consensus meetinstrumenten*

Belangrijk voor het opzetten van landelijk outcomes onderzoek is dat in alle klinieken op dezelfde wijze gemeten wordt middels tenminste een basisset van dezelfde instrumenten. Consensus is wenselijk over gestandaardiseerde diagnostiek en een minimale set van meetinstrumenten voor indicatiestelling, die bij de aanmelding of binnenkomst worden afgenomen. Tevens is een basisset van uitkomstmaten wenselijk.

In de tbs-sector is de laatste jaren veel gedaan aan instrumentontwikkeling en er is ruime expertise opgebouwd, met name met risicotaxatie-instrumenten, beoordelingsschalen en de monitoring van (agressieve) incidenten. Het verdient aanbeveling te komen tot een inventarisatie van wat in gebruik is aan meetinstrumenten in de verschillende instellingen. Het is hierbij belangrijk om onderscheid te maken tussen (1) verschillende methoden (beoordeling, zelfrapportage, event monitoring), (2) verschillen in niveau van meten (dicht bij de te evalueren interventie of juist meer generiek, zoals risicotaxatie) en (3) verschillende subpopulaties in de tbs.

#### *7.4.3 Scholing*

Medewerkers, belast met de taak de uitkomstgegevens te verwerven, dienen hiertoe geschoold te worden. In de tbs-sector zal vooral gebruik worden gemaakt van beoordelingsschalen. Voor een betrouwbare en valide toepassing van de meeste instrumenten dienen beoordelaars vaak uitgebreid getraind te worden en moet de training periodiek herhaald worden om verschuiving van beoordelingen over de tijd te voorkomen (bekend als drift). De betrouwbaarheid kan verder onderhouden worden door van tijd tot tijd de overeenstemming tussen beoordelaars vast te stellen (interbeoordelaarsbetrouwbaarheid).

Van belang voor scholing:

- opstellen van een plan voor scholing van medewerkers voor ROM;
- regelmatig vaststellen van interbeoordelaarsbetrouwbaarheid binnen en tussen klinieken.

#### *7.4.4 Externe toetsing onderzoeksaanvragen*

Indien overwogen wordt onderzoeksgelden ter beschikking te stellen om onderzoek naar therapie-effect te ondersteunen, suggereren we een werkwijze zoals die toegepast wordt bij de toedeling van onderzoeksgelden door bijvoorbeeld Zon-MW of NWO. Hierbij wordt het veld uitgenodigd voorstellen in te dienen volgens een vaststaand format en wordt gebruik gemaakt van veelal buitenlandse referenten, die de onderzoeksvoorstellen op kwaliteit en haalbaarheid beoordelen.

De voordelen van een dergelijke toewijzing van onderzoeksgelden is dat er een borging van kwaliteit plaatsvindt middels toetsing vooraf. Hierdoor wordt de kans op een overall voldoende wetenschappelijke kwaliteit aanzienlijk vergroot.

Bovendien leidt een dergelijke praktijk tot een zekere coördinatie van onderzoeksactiviteiten waardoor voorkomen kan worden dat werk dubbelop gedaan wordt.

Om samenwerking tussen klinieken verder te stimuleren kan hierbij als eis voor financiering worden betrokken dat het onderzoeksvoorstel aantoonbaar een resultaat is van samenwerking van een aantal tbs-instellingen.



## Summary

The study “Options for treatment outcome research in the ‘terbeschikkingstelling’ (tbs)”, was aimed at: (1) listing research methods and designs for efficacy and effectiveness research and (2) describing whether these methods are usable for investigating mechanisms of change and responsivity research. The tbs order is a penal measure. It can be imposed by the court upon mentally disordered offenders who are considered not to be responsible or to have diminished responsibility for their offence. Effectiveness and responsivity research will yield the evidence base for tbs treatment guidelines, composed of empirically supported therapies. Guidelines include descriptions of what works best for whom under which circumstances. The present study yields a realistic estimate of possibilities for and limitations to treatment outcome research in the tbs.

This report comprises three sections. The first section specifies the research question and our approach to answering it (Chapter 1). In addition, the tbs setting in which the treatment outcome research will be conducted is described in words and numbers (Chapter 2). Tbs patients are a quite heterogeneous population regarding their criminal history as well as the psychopathology they suffer from. The psychopathology is often severe and complex, as evidenced by high comorbidity rates in most tbs patients.

The second section (Chapter 3 and 4) considers the background of evidence-based treatment. The evidence-based treatment concept stems from the medical field, first moved to the mental health profession, and more recently to the tbs setting. Three steps need to be taken before the tbs setting can claim to work according to evidence-based principles: (1) scientific research to obtain the empirical support for treatments, (2) formulation of treatment guidelines based on empirical data, and (3) implementation and dissemination of these guidelines in the field. Thus, for the tbs setting, a first requirement is to build an evidence base for specific interventions and treatment programs through scientific research. Which of these are efficacious and effective? The research methods described in this report are primarily aimed at evaluation of these interventions and treatments. The suitability of treatment outcome research to address questions regarding the effectiveness of the tbs method as a whole, including non-therapeutic measures such as appropriate selection for treatment termination and posttreatment services that are offered, is limited.

An important issue in treatment outcome research is the choice of a proper outcome measure. Regarding outcome measures for effectiveness research in the tbs, the following is concluded: although reduction of actual recidivism may be the primary goal of tbs treatment, this variable is ill-suited and of limited value as an outcome measure of controlled effectiveness research, as it is in time far removed from the evaluated intervention and may be determined by many other factors than the treatment alone. In addition, recidivism is a rather insensitive outcome measure, necessitating a high number of patients participating in a trial to ensure sufficient statistical power. Better outcome variables are the extent to which specific treatments goals have been attained or the reduction of *risk* for recidivism.

The third section of this report describes methods of research to evaluate treatment effect. Chapter 5 describes methodological characteristics of several study designs for treatment outcome research. Issues related to internal and external validity are discussed and we address the difference between demonstrating efficacy (Is treatment X superior to control treatment Y?) under highly controlled circumstances and effectiveness of treatments (Is treatment efficacious under less controlled circumstances and is the treatment sufficient?) in clinical practice. An

important feature of a study design is whether it can be informative regarding the influence of responsivity factors. Besides these defining characteristics of study designs, each type of study comes with specific requirements for its implementation in clinical practice. Several of these requirements are described, such as randomisation, placebo treatment, and statistical power. These requirements have to be taken into account when we decide on the feasibility of study designs for the tbs.

Chapter 6 presents three types of outcome research: *controlled research* with randomised controlled trials, *outcome monitoring* and *single case designs*. Randomised controlled trials are the gold standard for controlled treatment outcome research, and the method of preference when it comes to demonstrating the efficacy of interventions (i.e., the outcome for experimentally-treated patients as compared to no treatment controls). In its simplest form, an RCT implies that patients are randomly assigned to either an experimental treatment condition (comprising the active treatment ingredients) or a control treatment without active ingredients (a wait list or placebo condition). Comparison of these groups indicates whether the experimental treatment is superior to the control treatment. The RCT design is a truly experimental design and offers ample possibilities to control for threats to the internal validity of the study. However, at the same time this optimized control threatens the external validity of the study: Prototypical patients and intensely supervised treatments are hard to find in everyday clinical practice. Single case designs are truly experimental studies that investigate changes within a single person. These designs are especially suited for a quick first impression on the efficacy of a new treatment. The generalizability of the findings is, however, limited until findings are replicated in a sizeable group of patients. Furthermore, single case designs are only feasible with specific interventions, namely interventions that can be “turned on and off at will”. RCT’s and single case designs are both less suited to investigate responsivity factors, factors that potentially influence treatment outcome. A key ingredient of the methodology of experimental designs is to try to minimize variance in responsivity factors, which renders them less appropriated for the investigation of these very factors.

Outcome monitoring aims at demonstrating the effectiveness of treatments in the clinical reality of everyday practice. Outcome monitoring implies routine data collection regarding the type and quality of treatments delivered, and simultaneous monitoring of the functioning and psychological wellbeing of the patient. At set time points, standardised outcome measures are administered. With outcome monitoring, current practice and new experimental treatments can be evaluated under controlled or uncontrolled conditions. Outcome monitoring is very well-suited for an investigation into factors related to patients’ responsivity to treatment (which personality or crime characteristics predict good outcome) and which treatment conditions do so. Thus, to address the question who benefits most from therapeutic interventions in the tbs and which treatment conditions are optimal for a positive outcome this design is suitable.

The three research designs are compared regarding their methodological strengths and weaknesses and their feasibility in the tbs setting. We conclude that the implementation of an RCT, albeit the gold standard for demonstrating efficacy of an intervention, is complex because of its required randomisation and incorporation of a control group in the design. Because of power considerations, substantial numbers of patients are required, limiting the number of RCT’s that can be conducted in the tbs. The internal validity of outcome monitoring is inferior to RCT’s but its external validity is superior. Furthermore, outcome monitoring is very feasible for the tbs setting. Single case designs and RCT’s both have a high validity for demonstrating



efficacy of the evaluated treatment. Single case designs are easily implemented, but their external validity is, however, limited.

Research endeavours in the tbs should aim at generating the much-needed evidence base (What works best for whom?) and, wherever feasible, constraints to empirical research should be counteracted. The last chapter addresses these constraints by describing requirements for treatment outcome research in the tbs and providing recommendations on how to build an evidence-based foundation for treatment in the tbs. In addition, outcome monitoring is recommended as a means of enabling small scale randomised controlled trials, and single case studies. A plea is made for the creation of an infrastructure for outcome monitoring, which would be the practical implementation of outcome monitoring. The infrastructure that is required for outcome monitoring encompasses monitoring of the treatments that are delivered and monitoring of outcomes. Treatment outcome research is a complicated endeavour and, apart from single case designs, valid results are only obtained with sufficient numbers of patients. Thus, most study designs pose requirements that go beyond the means of single tbs clinics. Collaboration among clinics should be stimulated, by striving for a common set of outcome measures. Another form of collaboration is organising RCT's primarily through multi-centre trials. A multi-centred approach could be a condition for applications for financial support for outcomes research.



## Literatuur

- Andrews, D.A., Bonta, J. & Hoge, R.D. (1990). Classification for effective rehabilitation: rediscovering psychology. *Criminal Justice and Behavior*, 17, 19-52.
- Andrews, D.A. & Bonta, J. (1994). *The psychology of criminal conduct*. Cincinnati, OH: Anderson.
- Bickman, L. (2000). *Research Designs*. London: Sage
- Bowen, E. & Gilchrist, E. (2004). Do court- and self-referred domestic violence offenders share the same characteristics? A preliminary comparison of motivation to change, locus of control and anger. *Legal and Criminological Psychology*, 9, 279-294.
- Brand, E. & Van Emmerik, J. (2006). *Handboek Forensisch Psychiatrische Profielen. Handleiding fp40*. Den Haag: Ministerie van Justitie, Dienst Justitiële Instellingen.
- Burns, M., Bird, D., Leach, C. & Higgins, K. (2003). Anger management training: the effects of a structured programme on the self-reported anger experience of forensic inpatients with learning disability. *Journal of Psychiatric Mental Health Nursing*, 10, 569-577.
- Campbell, D.T. & Stanley, J.C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Boston, Sage
- Chambless, D. L., Baker, M. J., Boucom, D. H., Beutler, L. E., Calhoun, K. S., Crits-Christoph, P. et al. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist*, 51, 3-16.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, Lawrence Erlbaum.
- Davies, J., Howells, K. & Jones, L. (in press). Evaluating innovative treatments in forensic mental health: A role for single case methodology? *The Journal of forensic psychiatry & psychology*.
- de Beurs, E., Van Dyck, R., van Balkom, A. J. L. M. & Lange, A. (2001). De klinische effectiviteit van "evidence-based" behandelingen voor psychische klachten. *Tijdschrift voor Psychiatrie*, 43, 1, 41-48.
- de Beurs, E. en Zitman, F.G. (2007). Routine Outcome Monitoring: Het meten van therapie-effect in de klinische praktijk met webbased software. *Maandblad Geestelijke Volksgezondheid*, 62, 13-28.
- de Kogel, C.H. (in voorbereiding). *Biologisch onderzoek en Justitievraagstukken op het gebied van preventie en sancties*. Factsheet 2007-5. Den Haag: WODC, Ministerie van Justitie.
- DiFazio, R., Abracen, J. & Looman, J. (2001). Group versus individual treatment of sex offenders: A comparison. *Forum on Corrections Research*, 13, 56-59.
- Drake, R.E., Goldman, H.H., Leff, H.S., Lehman, A.F., Dixon, L, Mueser, K.T., Torrey WC. (2001). Implementing evidence-based practices in routine mental health service settings. *Psychiatric Services*, 52, 179-182.
- Drieschner, K. (2005). *Measuring treatment motivation and treatment engagement in forensic psychiatric outpatient treatment: Development of two instruments*. Academisch proefschrift, Radboud Universiteit Nijmegen.
- Drieschner, K. H., Lammers, S. M. & van der Staak, C. P. (2004). Treatment motivation: An attempt for clarification of an ambiguous concept. *Clinical Psychology Review*, 23, 1115-1137.
- Eddy, D.M. (1990). "Practice policies: where do they come from?". *Journal of the American Medical Association* 263 (9): 1265, 1269, 1272 passim
- Elstein AS (2004). "On the origins and development of evidence-based medicine and medical decision making". *Inflammatory Research* 53 Suppl 2: S184-9.

- Ferriter, M. & Huband, N. (2005). Does the non-randomized controlled study have a place in the systematic review? A pilot study. *Criminal Behaviour and Mental Health*, 15, 111-120.
- Fink, A. (2005). *Evaluation Fundamentals: Insights into Outcomes, effectiveness and Quality of Health Programs*. London, Sage
- Fishbein, D. & Sheppard, M. (2006). *Assessing the role of neuropsychological functioning in inmates' treatment response*. Baltimore: RTI International.
- Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., Kremers, I., Nadort, M. & Arntz, A. (2006). Outpatient psychotherapy for Borderline Personality Disorder: A randomized clinical trial of Schema Focused Therapy versus Transference Focused Psychotherapy. *Archives of General Psychiatry*, 63, 649-658.
- Gilbody, S. M., House, A. O. & Sheldon, T. A. (2002a). Outcomes research in mental health - Systematic review. *British Journal of Psychiatry*, 181, 8-16.
- Hanson, R. K. & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: a meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73, 1154-1163.
- Hanson, R. K., Gordon, A., Harris, A. J., Marques, J. K., Murphy, W., Quinsey, V. L. et al. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sex Abuse*, 14, 169-194.
- Hildebrand, M., Hesper, B. L., Spreen, M. & Nijman, H. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie; Een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R*. Utrecht: Expertisecentrum Forensische Psychiatrie.
- Howard, K.I., Moras, K., Brill, P.L., Martinovich, Z., Lutz, W. (1996). Evaluation of psychotherapy. Efficacy, effectiveness, and patient progress. *American Psychologist*, 51, 1059-1064
- Hullsiek, K.H. & Louis, T. (2002). Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics* 2, 179-193.
- Hunsley, J., Mash, E.J. (2007). Evidence-Based Assessment. *Annual Review of Clinical Psychology* 3, 1-29
- Iacoboni, M. & Lenzi, G.L. (2002). Mirror neurons, the insula, and empathy. *Behavioral and Brain Sciences*, 25, 39-40
- Insel, T.R. & Fernald, R.D. (2004). How the brain processes social information: Searching for the social brain. *Annual Review of Neuroscience*, 27, 697-722.
- Kazdin, A. E. & Bass, D. (1989). Power to detect differences between alternative treatments in comparative psychotherapy outcome research. *Journal of Consulting and Clinical Psychology*, 57, 138-147.
- Kennedy, S. M. (2000). Treatment responsivity: Reducing recidivism by enhancing treatment effectiveness. *Forum on Corrections Research*, 12, 19-23.
- Kordy, H., Hannover, W. & Richard, M. (2001). Computer-assisted feedback-driven quality management for psychotherapy: The Stuttgart-Heidelberg model. *Journal of Consulting and Clinical Psychology*, 69, 173-183.
- Kraemer, H.C., (1992). Measurement of reliability for categorical data in medical research. *Statistical Methods in Medical Research*, 1, 183-199.
- Lambert, M. J., Hansen, N. B. & Finch, A. E. (2001). Patient-focused research: using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69, 159-172.
- Landelijke Stuurgroep Multidisciplinaire Richtlijnen in de GGZ. (2003a). *Multidisciplinaire richtlijn Stemningsstoornissen*. Utrecht: Trimbos Instituut.

- Landelijke Stuurgroep Multidisciplinaire Richtlijnen in de GGZ. (2003b). *Multidisciplinaire richtlijn Angststoornissen*. Utrecht: Trimbos Instituut.
- Landelijke Stuurgroep Multidisciplinaire Richtlijnen in de GGZ. (2007). *Multidisciplinaire richtlijn Schizofrenie*. Utrecht: Trimbos Instituut.
- Lambert, M. J., Harmon, C., Slade, K., Whipple, J. L. & Hawkins, E. J. (2005). Providing feedback to psychotherapists on their patients' progress: clinical results and practice suggestions. *Journal of Clinical Psychology*, 61, 165-174
- Looman, J., Dickie, I. & Abracen, J. (2005). Responsivity issues in the treatment of sexual offenders. *Trauma Violence Abuse*, 6, 330-353.
- Marshall, W. L. (2005). Therapist style in sexual offender treatment: influence on indices of change. *Sex Abuse*, 17, 109-116.
- Marshall, W. L., Serran, G., Moulden, H., Mulloy, R., Fernandez, Y. M., Mann, R. et al. (2002). Therapist features in sexual offender treatment: Their reliable identification and influence on behaviour change. *Clinical Psychology and Psychotherapy*, 9, 395-405.
- McGuire, J. (2002). Criminal Sanctions versus psychologically based interventions with offenders: A comparative empirical analysis. *Psychology, Crime & Law*, 8, 183-208.
- McMurrin, M. & Ward, T. (2004). Motivating offenders to change in therapy: An organizing framework. *Legal and Criminological Psychology*, 9, 295-311.
- Miller, S., Sees, C. & Brown, J. (2006). Key Aspects of Psychological Change in Residents of a Prison Therapeutic Community: A Focus Group Approach. *The Howard Journal of Criminal Justice*, 45, 116-128
- Ministerie van Justitie. (2003). *Handleiding bkt-30 versie 2002, risicotaxatie in de forensische psychiatrie*. Den Haag: Ministerie van Justitie, Dienst Justitiële Inrichtingen.
- Nadort, M. (2006). *Schemagericht cognitieve therapie: de rol van dysfunctionele schema's in de ontwikkeling van persoonlijkheidsproblematiek*. Voordracht op het Voorjaarscongres van de Nederlandse Vereniging voor Psychiatrie, mei 2006
- Nezu, A.M. & Nezu, C.M. (2008). *Ensuring treatment integrity*. In Nezu, A.M. & Nezu, C.M. (Eds): Evidence-based Outcome research. New York: Oxford University Press
- Nijman, H.L.I., Allertz, W.F.F. á Campo, J.M.L.G. (1995). Agressie van patiënten: een onderzoek naar agressief gedrag van psychiatrische patiënten op een gesloten opnameafdeling. *Tijdschrift voor Psychiatrie* 37, 329-342
- Pawson, R. & Tilley, N. (1997). Realistic evaluation. London: Sage.
- Philipse, M., de Ruitter, C., Hildebrand, M., Bouman, Y., Webster, C. (2000). *HCR-20: beoordelen van het risico van gewelddadig gedrag. Versie 2* Amsterdam: Universiteit van Amsterdam.
- Prentky, R. A., Lee, A.E.S., Knight, R.A., Cerce, D. (1997) Recidivism Rates Among Child Molesters and Rapists: A Methodological Analysis. *Law and Human Behavior*, 21, 6, 635-659.
- Rossi, P.H. & Freeman, H.E. (1993) *Evaluation, a systematic approach*. London: Sage
- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996). "Evidence-based medicine: what it is and what it isn't". *British Medical Journal*, 312 (7023), 71-72.
- Sellen, J. L., McMurrin, M., Cox, W. M., Theodosi, E. & Klinger, E. (2006). The Personal Concerns Inventory (Offender Adaptation): measuring and enhancing motivation to change. *International Journal of Offender Therapy and Comparative Criminology*, 50, 294-305.
- Shadish, W., Cook, T., Campbell, D. (2002). *Experimental and Quasi-Experimental Designs*. Boston: Houghton Mifflin Company.

- Shadish, W.R. (2000). *The empirical program of Quasi-Experimentation*. In Bickman, L. (ed). *Validity and Social Experimentation*. Sage, London
- Snowdon, R.J., Gray, N., Taylor, J. & MacCulloch, M.J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine* 37, 1539-1549.
- Stevens J.P. (2002). *Applied Multivariate Statistics for the Social Sciences*. New York, Lawrence Erlbaum Associates
- Tabachnick, B.G. & Fidell, L.S. (2001). *Using multivariate statistics*. Reiter, Washington D.C.
- Tervo, R. C., Estrem, T. L., Bryson-Brockmann, W. & Symons, F. J. (2003). Single-case experimental designs: applications in developmental-behavioral pediatrics. *Journal of Developmental Behavioral Pediatr.*, 24, 438-448.
- Urbaniok, F. (2006). *Fotres: Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System*. Kempten, Zytglogge Verlag
- Urbaniok, F. (2007). The prediction of criminal recidivism. *European Archives of Psychiatry and Clinical Neuroscience*, 257, 129-134.
- Van Emmerik, J.L. (2001). *De terbeschikkingstelling in maat en getal. Een beschrijving van de tbs-populatie in de periode 1995-2000*. Interne publikatie van de Afdeling Informatieanalyse en Documentatie van Dienst Justitiële Inrichtingen, Ministerie van Justitie
- Van Fenema, E., van der Wee, N., de Beurs, E., Onstein, E. & Zitman, F., (2006). *Implementatie van zorgprogramma's, de stand van zaken in de behandelpraktijk*, *Maandblad voor de Geestelijke Volksgezondheid*, 61 817-826.
- Van Wiel, N.M., van Goozen, S.H.M., Matthys, W., Snoek, H., van Engeland, H., (2004). Cortisol and treatment effect in children with disruptive behavior disorders: a preliminary study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43, 1011-1018.
- Veen, V. C. & De Ruiter, C. (2005). De effectiviteit van behandelingen bij seksuele delinquenten; Een overzicht van de internationale literatuur. *Justitiële Verkenningen*, 31, 75-94.
- Wartna, B.S.J., Tollenaar, N. & Essers, A.A.M.(2005). *Door na de gevangenis: Een cijfermatig overzicht van de strafrechtelijke recidive onder ex-gedetineerden*. WODC, Den Haag: Boom
- Wartna, B.S.J., El Habachi, Essers, A. A.M.(2006). *Strafrechtelijke recidive van ex-terbeschikkingsgestelden*. WODC, Den Haag: Boom
- Wartna, B.S.J.. (2005). *Evaluatie van daderprogramma's: een wegwijzer voor onderzoek naar de effecten van strafrechtelijke interventies speciaal gericht op het terugdringen van recidive*. WODC, Den Haag: Boom
- Well, K.B. (1999). Treatment research at the crossroads: The scientific interface of clinical trials and effectiveness research. *The American Journal of Psychiatry*, 156, 5-10.
- Westen, D., Novotny, C. M. & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, 130, 631-663.
- Woods S.W., Sholomskas D.E., Shear M.K., Gorman J.M., Barlow D.H., Goddard A.W. & Cohen J. (1998) Efficient allocation of patients to treatment cells in clinical trials with more than two treatment conditions. *American Journal of Psychiatry*, 155, 1446-1448.
- Young, J. E., Klosko, J.S. & Weishaar, M.E. (2003). *Schema Therapy: A practitioner's guide*. New York: Guilford.