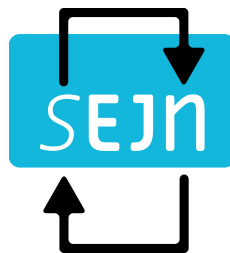


Over verandering gesproken

Op weg naar meer zicht op effectiviteit van de jeugdzorgpraktijk



Esther Geurts
Lianne Lekkerkerker
Tom van Yperen
Jan-Willem Veerman

Nederlands Jeugdinstituut en Praktikon
namens het Samenwerkingsverband Effectieve Jeugdzorg Nederland / SEJN
Utrecht, september 2010

© 2010 Nederlands Jeugdinstituut / Praktikon

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op enige andere wijze zonder voorafgaande schriftelijke toestemming.

Het Nederlands Jeugdinstituut en Praktikon hebben deze studie - onder de werktitel 'Project Nulmeting' - verricht namens het Samenwerkingsverband Effectieve Jeugdzorg Nederland (SEJN), in opdracht en met projectsubsidie van het ministerie voor Jeugd en Gezin.

Het SEJN is een groep van jeugdzorginstellingen en ondersteuningsorganisaties (Praktikon, Nederlands Jeugdinstituut, PI Research en PIONN) die een samenhangende meet- en verbeterbeweging in de jeugdsector wil realiseren. Voor meer informatie: www.sejn.nl.

Auteurs

Esther Geurts (Nederlands Jeugdinstituut)

Lianne Lekkerkerker (Nederlands Jeugdinstituut)

Tom van Yperen (Nederlands Jeugdinstituut ,1^e projectleider)

Jan-Willem Veerman (Praktikon, 2^e projectleider)

Begeleidingsgroep

Anja Verburg (MOgroep Jeugdzorg)

Eefje Strijbosch (Tender)

Erik Knorth (RijksUniversiteit Groningen)

Karin van der Ham (ministerie voor Jeugd en Gezin)

Ruud van Herk (ministerie voor Jeugd en Gezin)

Nederlands Jeugdinstituut (penvoerder)

Catharijnesingel 47

Postbus 19221

3501 DE Utrecht

Telefoon (030) 230 63 44

Website www.nji.nl

Infolijn (030) 230 65 64, e-mail info@nji.nl

Praktikon

Montessorilaan 10

Postbus 9104

6500 HE Nijmegen

Telefoon (024) 361 11 50

Website www.praktikon.nl

Inhoudsopgave

VOORWOORD	5
SAMENVATTING	7
1. DOELEN EN OPZET PROJECT	9
1.1 INLEIDING	9
1.2 DOELEN EN DOMEIN	10
1.3 FASERING VAN HET PROJECT	11
2. UITVOERING PROJECT NULMETING.....	12
2.1 CONTACT ONDERZOEKERS EN INSTELLINGEN JEUGD & OPVOEDHULP	12
2.2 VERSLAG WERKBEZOEKEN JEUGD & OPVOEDHULP.....	13
2.3 RESPONS DATAVERZAMELING.....	14
2.4 PRIVACY, EIGENDOM EN RECHTEN	15
2.5 DATABEWERKING.....	15
3. RESULTATEN TAZ-BESTAND	17
3.1 INLEIDING.....	17
3.2 KENMERKEN VAN JEUGDIGEN	17
3.3 INSTRUMENTEN, INFORMANTEN EN INDEXEN.....	20
3.4 STEP: VERANDERING ERNST PROBLEMATIEK VOLGENS HULPVERLENER.....	21
3.5 YSR: VERANDERING PROBLEMATIEK JEUGDIGE VOLGENS JEUGDIGE	23
3.6 SDQ-J: VERANDERING PROBLEMATIEK JEUGDIGE VOLGENS JEUGDIGE.....	24
3.7 CBCL: VERANDERING PROBLEMATIEK JEUGDIGE VOLGENS MOEDERS	25
3.8 SDQ: VERANDERING PROBLEMATIEK JEUGDIGE VOLGENS OUDERS.....	27
3.9 SDQ: VERANDERING PROBLEMATIEK JEUGDIGE VOLGENS LEERKRACHT	29
3.10 NOSI: VERANDERING OPVOEDINGSTRESS VOLGENS MOEDERS.....	30
3.11 BEËINDIGING EN DUUR VAN DE HULP	31
3.12 DISCUSSIE	33
4. GEBRUIK TAZ-BESTAND ALS BRON CONTROLEGROEPEN	37
4.1 INLEIDING.....	37
4.2 DE PRINCIPES.....	37
4.3 OPZET VAN HET SIMULATIE-ONDERZOEK.....	39
4.4 EERSTE SIMULATIE-ONDERZOEK: MATCHING STEP.....	40
4.5 TWEDE SIMULATIE-ONDERZOEK: MATCHING CBCL.....	42
4.6 DISCUSSIE	45
5. CONCLUSIES.....	47
5.1 INLEIDING.....	47
5.2 BEELD VAN EFFECTIVITEIT	47
5.3 DATAPOOL MATCHING MET EXPERIMENTELE GROEPEN	48
5.4 DE GATEN ZICHTBAAR	49

5.5	TOT SLOT.....	50
6.	LITERATUURVERWIJZINGEN.....	52
	BIJLAGE 1. VERSLAG PROEFBESTANDEN.....	53
	DOEL PROEFBESTANDEN.....	53
	AANPAK.....	53
	BEVINDINGEN.....	53
	BETEKENIS VOOR TAZ-BESTAND	55
	OVERZICHT VARIABLEN IN PROEFBESTANDEN	55
	BIJLAGE 2. RESPONS	57
	BIJLAGE 3. INHOUD TAZ-BESTAND	59
	BIJLAGE 4. VERSCHILLEN IN BEREKENING EFFECT SIZES ES	60
	BIJLAGE 5. DE MATCHINGPROCEDURE	61
	INLEIDING	61
	VOORWAARDEN	61
	VARIANT 1. SIMPLE PROFILE MATCHING (SPM).....	62
	VARIANT 2. CLUSTER ANALYSIS MATCHING (CAM).....	63
	VARIANT 3. PROPENSITY SCORE MATCHING (PSM).....	64
	REFERENTIES	65

Voorwoord

In april 2009 is het zogeheten ‘Project Nulmeting’ van start gegaan. In dit project is een bestand opgebouwd waarin bestaande onderzoeksgegevens, betreffende de ‘thans aangeboden zorg’ in de jeugdzorg (ook wel TAZ-bestand genoemd), bijeen zijn gebracht. Met deze gegevens moet een schets geboden worden van de effectiviteit van de jeugdzorg in de alledaagse praktijk. Tevens moet het een ‘pool’ zijn van data waaruit men kan putten om vergelijkingsgroepen te maken voor de effectbepaling van experimentele behandelvormen. Voorts zou het project duidelijk moeten maken over welke behandelgroepen er nog weinig gegevens zijn, zodat daar vervolgonderzoek naar geëntameerd zou kunnen worden. De bedoeling was databestanden bijeen te brengen van zoveel mogelijk onderzoekers en praktijkinstellingen die gegevens hebben verzameld over de provinciaal gefinancierde jeugdzorg (voorheen jeugdhulpverlening genoemd, tegenwoordig aangeduid als jeugd- en opvoedhulp).

In oktober 2010 verscheen het eerste tussenverslag voor dit project (Van Yperen e.a., 2009). Hierin hebben we verslag gedaan van onze ervaringen en bevindingen in de eerste fase van het project. Het verslag is besproken met de begeleidingscommissie. In overleg met de begeleidingscommissie hebben we besloten om geen oproep te doen aan alle instellingen voor Jeugd & Opvoedhulp om data te leveren voor het bestand, maar ons te richten op de tien instellingen voor Jeugd & Opvoedhulp waarvan verwacht werd dat daarvan de kans het grootst is dat zij gegevens voor het project kunnen leveren.

In maart 2010 verscheen het tweede tussenverslag (Lekkerkerker e.a., 2010). Hierin hebben we gerapporteerd over de ervaringen en de bevindingen die zijn opgedaan in het contact met de instellingen voor Jeugd & Opvoedhulp en de onderzoekers die we hebben benaderd voor het project. Aan de hand van een aantal proefbestanden is de verdere haalbaarheid van het project bepaald. De eerste resultaten waren veelbelovend.

In het voor u liggende eindverslag bieden we de wetenschappelijke rapportage over de dataverzameling, databewerking en de uiteindelijke opbouw van het TAZ-bestand. Vervolgens gaan we in op de drie doelstellingen van het project: nulmeting, matching en vervolgonderzoek. Tenslotte eindigen we met een conclusie over de kwaliteit en bruikbaarheid van het TAZ-bestand. Het onderhavige rapport vormt de basis voor afgeleide artikelen, presentaties en andere producten voor praktijkonderzoekers, managers en beleidsmakers.

Wij zijn ervan overtuigd dat met het project een nieuwe stap is gezet is de ontwikkeling van een beeld van de effectiviteit van de jeugdzorg in de praktijk. Dat beeld is nog incompleet, maar de opgebouwde systematiek is nuttig. Onze dank gaat uit naar personen en instellingen (De Hoenderloogroep, Entréa, Jarabee, Stichting Jeugdformaat, Pactum, Vitree, Tender en Yorneo) die aan het project hebben meegewerkt. Zij waren bereid de gegevens te leveren die in dit project zijn gepresenteerd.

De auteurs

Utrecht, september 2010

Samenvatting

In dit project is een gegevensbestand bijeen gebracht uit diverse studies en registraties die zicht bieden op de effectiviteit van de Thans Aangeboden Zorg in de provinciaal gefinancierde jeugdzorg (tegenwoordig Jeugd- & Opvoedhulp genoemd). Dit bestand noemen we hier het TAZ-bestand. De gegevens zijn afkomstig uit reeds uitgevoerde en lopende studies bij niet-experimentele groepen en uit lopende cliëntregistraties. Het resulterende TAZ-bestand dient – conform de opdracht van het project - drie doelen:

- *Doel 1: Het bestand geeft een beeld van de effectiviteit van de dagelijkse praktijk bij een aantal hulpvormen in de jeugd- en opvoedhulp. Het biedt een soort nulpunt voor de te behalen winst met (nieuwe) verbeterlagen in de sector.*

Resultaat: Het TAZ–bestand blijkt een selectie van cliënten te beslaan die niet geschikt is om een goed beeld te geven van de mate van verandering in de problematiek bij de cliënten die jeugd- en opvoedhulp krijgen. We hebben vanwege de selectie immers onvoldoende zicht op de representativiteit van de in het bestand opgenomen groep. Beperken we ons tot deze selectie, dan blijkt de overgrote meerderheid van de cliënten op de geplande wijze de hulp te hebben beëindigd. De hulpverleners en de jeugdigen zelf rapporteren over het algemeen een kleine verandering (de effect size ligt tussen de 0.24 en 0.30) en de moeders over het algemeen een (bijna) middelgrote verandering (de effect size ligt tussen de 0.49 en 0.61). Over alle vragenlijsten en informanten gemiddeld is de mate van verandering klein (effect size 0.38).

De mate van verandering hangt samen met de ernst van de problematiek bij het begin van de hulp. Zo blijkt er in het bestand een aanzienlijke groep die op een vragenlijst bij aanvang in het bereik van geen tot matige problemen scoort en daarna geen of weinig verandering laat zien. Hoe hoger de probleemscore aan het begin, hoe groter over het algemeen de verandering. Bij de ernstige problematiek zien we dan ook vaak veranderingen die groot tot zeer groot zijn (effect size > 0.80). Voor een deel is hier een statistisch artefact in het spel, maar afgezien daarvan roept de gevonden relatie veel vragen op. Hoe dan ook maakt het duidelijk dat effectgroottes van instellingen of van behandelvormen nooit voor zich spreken; ze moeten altijd in relatie worden gebracht met de kenmerken van de doelgroep – waaronder de ernst van de problematiek aan het begin van de hulp – en de doelen, zwaarte en duur van de hulp.

Overigens blijkt de groep met ernstige problematiek bij aanvang ook bij de afsluiting gemiddeld nog steeds aanzienlijke tot ernstige problematiek te hebben. De verandering bij ernstige problematiek is dus wel groot, maar niet voldoende om de groep uit de ‘gevaarzone’ te halen.

- *Doel 2: Het bestand biedt een verzameling van demografische en effectgegevens over een groot aantal cliënten in diverse hulpvormen in de sector. Hieruit zijn voor wetenschappelijke onderzoeksdoeleinden TAZ-controlegroepen samen te stellen om de effectiviteit van experimentele interventies te bepalen.*

Resultaat: Op basis van simulatie-onderzoek blijkt het TAZ-bestand zich redelijk goed te lenen voor een matching van experimentele groepen. Groot nadeel is dat van belangrijke matchingsvariabelen zoals etniciteit en reden beëindiging zorg er veel gegevens missen. Dat beperkt de trefkans bij matchingprocedures sterk, zeker als er wordt voorgeselecteerd op

bijvoorbeeld hulpvorm. Twee van de drie uitgeteste matchingprocedures bleken goed hanteerbaar en leidden tot een 100% matching van de experimentele groep (N=40) aan casussen uit het TAZ-bestand. Het selectieprobleem van het TAZ-bestand is bij een matching met een experimentele groep uit de praktijk kleiner dan bij gebruik van het bestand voor de bepaling van de effectiviteit van de thans aangeboden zorg, maar nog niet helemaal verdwenen. Voorts geldt dat men bij gebruik van de matchingprocedure een quasi-experimenteel onderzoek doet. Dat biedt minder zekerheid dan een experimentele opzet. Maar het is nog altijd meer dan alleen een voor- en een nameting bij een experimentele groep zonder enig vergelijkingsmateriaal.

- *Doel 3: De gegevens die bijeen worden gebracht zullen 'gaten' vertonen. Het bestand laat de 'gaten' zien en toont welke vervolgstudies nodig zijn om deze te dichten.*

Resultaat: Gezien het selectieve karakter van het TAZ-bestand is niet te zeggen welke gaten er in het zicht op de effectiviteit van de jeugd- en opvoedhulp bestaan. Wel is met het project een goede indruk verkregen van de lacunes die er optreden, zodra men resultaatgegevens van instellingen bijeen brengt. Die lacunes wijzen onder meer op de noodzaak om

- de verzameling van gegevens die relevant zijn voor de resultaatbepaling landelijk verder te standaardiseren;
- probleemkenmerken van doelgroepen in kaart te brengen (eigenlijk zou er per cliënt een probleemprofiel moeten worden gemaakt over alle relevante terreinen: gedrag jeugdige, vaardigheden en ontwikkeling, lichamelijke problemen, problemen in gezin of opvoeding, problemen in of met de bredere omgeving);
- uitwisselbaarheid en koppelmogelijkheden van bestanden te vergroten;
- problemen omtrent de berekening van de effectgrootte (ES) op te lossen;
- de bereidheid van instellingen om gegevens voor onderzoek beschikbaar te stellen te vergroten.

Het project 'Nulmeting' was een eerste poging om bestanden bijeen te brengen voor de doelen die hier zijn beschreven. Op basis van de wetenschappelijke rapportage die in dit rapport is geboden, zullen er afgeleide publicaties en presentaties worden gemaakt voor praktijkonderzoekers, managers en beleidsmakers. Hoewel het project eindig is, gaat de ontwikkeling door met uitbreiding van het TAZ-bestand met nieuwe gegevens, een studie naar een betere berekening van de effectgrootte bij een voor- en nameting zonder controlegroep en de beschikbaarstelling van de uitgeteste matchingprocedures aan instellingen voor jeugd- en opvoedhulp.

1. Doelen en opzet project

1.1 Inleiding

Er is weinig bekend over de effectiviteit van de jeugdzorg in de alledaagse praktijk. Schattingen lopen uiteen van nul effect (Weisz, Donenberg, Han & Kauneckis, 1995) tot kleine of matige effecten (Van Yperen, 2005). De roep om de implementatie van interventies die *evidence-based* zijn – en in wetenschappelijk onderzoek aanzienlijke effecten laten zien - klinkt in de jeugdzorg de laatste jaren steeds luider. De nadruk op het zoveel mogelijk werken met interventies van bewezen effectiviteit stelt de sector echter wel voor een probleem. De meeste interventies die zij tot hun beschikking hebben, zijn nog onvoldoende *evidence-based*. Veerman en Van Yperen (2007) schatten dat hooguit 10% van de in de praktijk uitgevoerde interventies dit predicaat kunnen krijgen. Er is dan ook veel onderzoek nodig om te bekijken welke interventies het goed doen.

Veel beleidsmaatregelen willen het gebruik van effectieve interventies stimuleren. De vraag is of dit helpt de sector effectiever te maken. Systematische beleidsevaluatie moet hier meer licht op werpen. Dergelijke evaluaties zouden eigenlijk moeten plaatsvinden door bij aanvang van de implementatie van een beleidsmaatregel een nulmeting te doen, gevolgd door een hermeting na verloop van tijd. Dit gebeurt echter maar weinig.

Systematische evaluatie vraagt om wetenschappelijk effectonderzoek. Dit is ervoor bedoeld meer specifiek vast te stellen of een nieuwe interventie effectiever is dan de thans gebruikelijke zorg. Daarbij is het gebruikelijk om een experimentele groep te vergelijken met een controlegroep. In de experimentele groep krijgen cliënten de interventie waarvan men het effect wil vaststellen. De controlegroep krijgt dan meestal de gebruikelijke, in de dagelijkse praktijk uitgevoerde zorg. Hier noemen we dit de Thans Aangeboden Zorg (TAZ)¹. Door de experimentele en TAZ-groep door toeval ('at random') samen te stellen is een vrij zuivere schatting te verkrijgen van de toegevoegde waarde van de interventie. Deze methode bestaat bekend als de RCT (Randomized Controlled Trial) en geldt bij wetenschappelijke onderzoekers doorgaans als de meest krachtige methode om het effect van een interventie vast te stellen. Een RCT blijkt in de praktijk van de jeugdzorg echter bijzonder moeilijk realiseerbaar. Het indelen van de groepen verloopt vaak niet goed, hulpverleners willen geregeld niet meewerken aan de randomisering, tijdens het onderzoek blijken de groepen toch niet voldoende vergelijkbaar etc. Zowel voor onderzoekers als praktijkwerkers werkt dit vaak frustrerend: het onderzoek verloopt vaak met hangen en wurgen. Het alternatief is dat men de controlegroep weglaat. Voor een voorstudie is dat acceptabel. Echter, voor de uiteindelijke bepaling of een nieuwe interventie een meerwaarde oplevert is een controlegroep wetenschappelijk gezien een *must*. Het ontbreken van dit soort onderzoek vormt dan ook vaak het spreekwoordelijke gat tussen wetenschap en praktijk.

In 2008 is daarom het project Nulmeting opgezet om het genoemde gat tussen wetenschap en praktijk te dichten met een voor betrokkenen aanvaardbare oplossing (voor het projectvoorstel zie Van Yperen & Veerman, 2008a). De kern van deze oplossing is dat het uitvoeren van een RCT weliswaar het streven mag zijn om een zo sterk mogelijke bewijskracht voor effectiviteit te leveren, maar dat minder

¹ In het engels spreekt men hier vaak van 'Care As Usual' / CAU of 'Treatment As Usual' / TAU.

sterke onderzoeksopzetten in opklimmende gradaties eveneens bewijskracht kunnen leveren. Een van die onderzoeksopzetten is een zogeheten quasi-experimentele design waarbij gebruikt gemaakt wordt van een reeds bestaande groep als controlegroep. Dit design kent geen willekeurige toewijzing. De bestaande groep kan uit een andere afdeling van een instelling komen waar de experimentele behandeling niet plaats vindt, kan uit een afdeling van een andere instelling komen, of uit verschillende afdelingen van verschillende instellingen. Een lastig punt hierbij blijft dan de selectie van die bestaande groep, de kans is aanwezig dat die op belangrijke aspecten anders is samengesteld dan de experimentele groep. Dit zou de vergelijking met de experimentele groep mank doen gaan. In het project Nulmeting worden hiervoor oplossingen gezocht.

Het project vormt een initiatief van het zogeheten Samenwerkingsverband Effectieve Jeugdzorg Nederland (SEJN), een groep van ondersteuningsorganisaties en jeugdzorginstellingen die streven naar het vestigen van een meetcultuur en een verbeterbeweging die erop is gericht de effectiviteit van de jeugdzorg in beeld te brengen en verder te ontwikkelen (voor meer informatie zie www.sejn.nl). Het Nederlands Jeugdinstituut en Praktikon voerden het project namens het SEJN uit.

1.2 Doelen en domein

Het project Nulmeting wil een gegevensbestand bijeen brengen uit diverse studies en registraties die zicht bieden op de effectiviteit van de thans uitgevoerde jeugdzorg. Dat gebeurt door de opbouw van een bestand met gegevens over de Thans Aangeboden Zorg. Dit bestand noemen we hier het TAZ-bestand. De gegevens zijn afkomstig uit reeds uitgevoerde en lopende studies bij niet-experimentele groepen en uit lopende cliëntregistraties. Het TAZ-bestand dient drie doelen:

- Ten eerste geeft het een beeld van de effectiviteit van de dagelijkse praktijk bij een aantal hulpvormen in de jeugd- en opvoedhulp. Het biedt een soort nulpunt voor de te behalen winst met (nieuwe) verbeterlagen in de sector.
- Ten tweede biedt het bestand een verzameling van demografische en effectgegevens over een groot aantal cliënten in diverse hulpvormen in de sector. Hieruit zijn voor wetenschappelijke onderzoeksdoeleinden TAZ-controlegroepen samen te stellen om de effectiviteit van experimentele interventies te bepalen.
- De gegevens die bijeen worden gebracht zullen ‘gaten’ vertonen. Het bestand laat de ‘gaten’ zien en toont welke vervolgstudies nodig zijn om deze te dichten.

Het domein waarop de studie betrekking heeft is als volgt te karakteriseren. In de eerste plaats gaat het om in Nederland aangeboden jeugdzorg die voorheen werd aangeduid als jeugdhulpverlening (i.e. de provinciaal gefinancierde jeugdzorg, tegenwoordig Jeugd & Opvoedhulp genoemd). Voorbeelden van de hulpvormen waar het hierbij om gaat zijn de intensieve pedagogische thuishulp, de ambulante jeugd- en opvoedhulp, de (semi-) residentiële hulp, de pleegzorg, zowel in een vrijwillig als gedwongen kader. De studie gaat niet over de via de AWBZ of zorgverzekeraars gefinancierde zorg (jeugd-ggz, kinder- en jeugdpsychiatrie, hulp aan licht verstandelijk gehandicapten), de justitiële hulp (justitiële jeugdinrichtingen, jeugdreclassering en overige hulp aan jeugdige delinquenten), de gemeentelijk gefinancierde preventie en de verslavingszorg. Ten tweede betreft het hulp die in de alledaagse praktijk is aangeboden. Hieronder wordt verstaan: het betreft een niet-experimentele behandeling die in meerdere instellingen wordt uitgevoerd, en/of die algemeen gezien wordt als gebruikelijk aanbod. In de derde plaats gaat het om recente studies, dat wil zeggen, gepubliceerd na 2000 (2001-2009). De veronderstelling is dat deze rapportages een beeld geven van de jeugdzorg zoals die in de afgelopen

tien jaren (1999-2008) is aangeboden. Blijkt uit de rapportage dat de gegevens betrekking hebben op een periode die voor meer dan de helft van vóór 1998 dateert, dan is de studie uitgesloten. Ten vierde zijn databestanden ingesloten waarin de afname van de problematiek bij de cliënten in kaart is gebracht door een meting aan het begin en het einde van de hulp.

1.3 Fasering van het project

Oorspronkelijk stond de uitvoering van het project gepland voor de periode januari 2009 tot en met februari 2010. Doordat het project later van start is gegaan dan gepland en het instellingen meer tijd kostte om gegevens te leveren dan beoogd, is het project – in overleg met het ministerie van Jeugd en Gezin - uitgevoerd in de periode van april 2009 tot en met juli 2010. Het project is opgedeeld in drie fasen, die hierna kort worden toegelicht.

Fase 1. Voorbereiding van het traject

In deze fase is helderheid verschaft over de te beantwoorden vragen van beleidsmakers, praktijkinstellingen en wetenschappers, de in beeld te brengen variabelen en de eisen ten aanzien van de op te nemen bestanden uit reeds uitgevoerde en lopende studies, de gegevensopslag en rapportage. Tevens is in deze fase een eerste inventarisatie gemaakt van reeds uitgevoerd onderzoek in de afgelopen zes jaren (2003-2008) waaruit in principe gegevens te gebruiken zijn. Voorts zijn er afspraken gemaakt over privacy-aspecten van het project, het eigendom over de gegevens en de gebruiks- en publicatierechten. Voorstellen voor dit alles zijn vastgelegd in een startdocument. Dit startdocument is per post en email verspreid onder een brede groep onderzoekers en kwaliteitsfunctionarissen van instellingen. Zij zijn gevraagd commentaar te leveren op het startdocument. Resultaat hiervan was driedelig: (1) het project is door wetenschappers en praktijkwerkers expliciet gesteund, (2) de opzet van het project is geaccepteerd en (3) er is een modelovereenkomst opgesteld met afspraken over de levering van gegevens.

Fase 2. Start en uitvoering dataverzameling

In de tweede fase is de daadwerkelijke gegevensverzameling opgestart. In het volgende hoofdstuk wordt uitgebreid verslag gedaan van deze fase. Globaal bestond deze fase uit de volgende activiteiten: Allereerst zijn verschillende onderzoekers en instellingen voor Jeugd & Opvoedhulp benaderd voor deelname aan het TAZ-bestand. Vervolgens hebben werkbezoeken plaatsgevonden bij instellingen die deel wilden nemen. Daarna zijn de gegevensbestanden verzameld. Op de verzamelde gegevens zijn de nodige technische en inhoudelijke bewerkingslagen uitgevoerd om ze naar het gemeenschappelijke gegevensmodel te converteren. Vervolgens zijn analyses uitgevoerd om de kenmerken van de verzamelde gegevens in beeld te brengen. Deze kenmerken zijn in hoofdstuk 3 gerapporteerd.

Fase 3. Demonstratie gebruik en eindrapportage

In de laatste fase is het onderhavige rapport opgemaakt. Het project is afgerond met een afsluitende bijeenkomst van de deelnemers en andere genodigden. In deze bijeenkomst is een demonstratie vertoond van het gebruik van het bestand als bron voor het matchen van cliënten uit een experimentele groep met een TAZ-controlegroep (zie hoofdstuk 4). Na verspreiding van het rapport wordt het bestand bij een overeengekomen instantie neergelegd. De instantie is daarbij verantwoordelijk voor het – onder voorwaarden – beschikbaar stellen van de gegevens aan onderzoekers die een TAZ-controlegroep nodig hebben. Verder zal het TAZ-bestand na afsluiting van het project nog een diverse publicaties opleveren.

2. Uitvoering project Nulmeting

2.1 Contact onderzoekers en instellingen Jeugd & Opvoedhulp

Voor de opbouw van het TAZ-bestand is gebruik gemaakt van zowel gegevens uit reeds uitgevoerde studies bij niet-experimentele groepen, als van gegevens uit lopende cliëntregistraties. De eerstgenoemde gegevens betreffen verzamelde data van onderzoekers, de tweede betreffen monitorbestanden van instellingen voor Jeugd & Opvoedhulp (hierna af te korten tot J & O).

De onderzoekers

Zoals in de rapportage fase 1 vermeld staat is er een inventarisatie verricht van reeds uitgevoerd en lopend onderzoek in de afgelopen tien jaren. Belangrijkste doel van de inventarisatie was onderzoekers te identificeren die mogelijk over bruikbare databestanden beschikken. Hierbij zijn we als volgt tewerk gegaan.

- Eerst is op basis van bestaande overzichtsstudies een lijst gemaakt van potentieel bruikbare publicaties over onderzoeken waarin de effectiviteit van de dagelijkse hulp is onderzocht, hetzij als primair object, hetzij als controleconditie voor de effectbepaling van een experimentele behandeling. Deze gegevens zijn voor de meest recente onderzoeken aangevuld met informatie uit de Databank Effectieve Jeugdinterventies (zie www.nji.nl/jeugdinterventies) en de Databank Nederlands Onderzoek Jeugd en Opvoeding van het Nederlands Jeugdinstituut. In de laatstgenoemde is (onder meer) recent afgesloten en lopend onderzoek dat gefinancierd wordt door ZonMw opgenomen.
- Ten tweede zijn als extra controle op de bestaande overzichten en voor aanvulling met recente studies alle recente jaargangen (2001-2009) doorgenomen van Nederlandse tijdschriften waarin mogelijk gerapporteerd wordt over de effecten van de jeugd- en opvoedhulp: Kind en Adolescent, Kind en Adolescent Praktijk, Maandblad Geestelijke Volksgezondheid, Jeugd en Co Kennis, Pedagogiek, Tijdschrift voor Orthopedagogiek, Tijdschrift voor Orthopedagogiek, Kinderpsychiatrie en Klinische Kinderpsychologie (TOKK), Pedagogische Studiën.

In overleg met de begeleidingscommissie is besloten de databestanden met een $N > 50$ te selecteren en de onderzoekers schriftelijk te benaderen voor deelname aan het project. Twee weken na verzending van een schriftelijk verzoek zijn de onderzoekers telefonisch benaderd om inhoudelijke vragen over het project te beantwoorden en te peilen of ze wilden deelnemen aan het project.

De instellingen voor Jeugd & Opvoedhulp

Aan de leden van het Platform Effectiviteit van het Samenwerkingsverband Effectieve Jeugdzorg Nederland (zie www.sejn.nl) is gevraagd of zij gegevens hebben die relevant zijn voor het project en of die gegevens beschikbaar zijn te stellen. In overleg met de begeleidingscommissie werd daarnaast besloten tevens de tien grootste instellingen voor J&O te benaderen die praktijkonderzoek verrichten in BergOp². Een van de onderzoekers bij dit project (JWV) heeft vervolgens de directeurs van de desbetreffende instellingen schriftelijk benaderd met een verzoek tot deelname aan het project

² BergOp is een databaseprogramma voor de opslag van onderzoeksgegevens (zie www.bergop.info). BergOp is ontwikkeld door Praktikon, Nijmegen. BergOp biedt mogelijkheid om per locatie alle gewenste instrumenten die in het kader van een onderzoek

'Nulmeting Jeugdzorg' en levering van data. Op het moment dat een instelling te kennen gaf mee te willen werken, heeft het Nederlands Jeugdinstituut telefonisch contact gezocht. Het telefonisch gesprek werd gebruikt om meer uitleg te geven over het doel, de inhoud, de opzet en de aanpak van het project om uiteindelijk een afspraak te maken voor een werkbezoek.

Doel van het werkbezoek was enerzijds de instelling de mogelijkheid te geven tot het stellen van vragen over het project en anderzijds had het werkbezoek tot doel helderheid te krijgen over de manier waarop onderzoek een invulling krijgt in de instelling. Tevens werd concreet gekeken welke variabelen door de instelling geleverd konden worden en op welke wijze de data bij het Nederlands Jeugdinstituut terecht zouden kunnen komen.

2.2 Verslag werkbezoeken Jeugd & Opvoedhulp

In de werkbezoeken werd duidelijk dat onderzoek steeds meer is ingebed in het primaire proces bij de verschillende instellingen voor J&O. Het blijkt dat bij het begin van het deeltraject J&O³ en bij einde van het deeltraject J&O vragenlijsten worden ingevuld door ouders/ jongere/ leerkracht/ hulpverlener om meer zicht te krijgen op het gedrag van het kind en de kwaliteit van de omgeving. In sommige instellingen voor J&O vinden daarnaast tussenmetingen plaats gekoppeld aan het behandelplan en worden follow-up metingen afgenomen zes maanden en één jaar na einde deeltraject J&O.

Om uitkomsten van de vragenlijsten te benutten worden de scoreprofielen bij de ene instelling opgenomen in het behandelplan, bij de andere instelling worden ze gebruikt in een cliënt contact en in een enkel geval worden de uitkomsten op module- en organisatieniveau benut.

Kijkend naar de financieringsbronnen van de instellingen voor J&O blijken de instellingen naast de provinciaal gefinancierde hulp, ook te maken te hebben met de AWBZ, de zorgverzekeringswet, de gemeente en het ministerie van Jeugd en Gezin. Daarbij geldt dat de hulpvorm niet altijd verschillend van aard is indien deze gefinancierd is vanuit de AWBZ, de zorgverzekeringswet of de provinciaal gefinancierde hulp. Voor dit project is afgesproken dat zolang de hulp van J&O wordt geboden binnen de Jeugdzorg, de gegevens meegenomen worden voor de opbouw van het databestand.

In de werkbezoeken werd duidelijk dat de wijze van registratie in de ene instelling gebeurt op basis van de hulpvorm (bijv. ambulant, residentieel) en in de andere instelling op basis van zorgaanspraken (volgens de indeling van Kaiser). Dit zorgt voor verschillen in de structuur en in de opbouw van de

gebruikt worden, in één en dezelfde database op te slaan. Van de vragenlijsten kunnen vervolgens individuele rapporten of scoreprofielen afgedrukt worden ten behoeve van de hulpverlening.

3 Met behulp van de prestatie-indicatoren worden de resultaten vastgelegd van de jeugdzorg die geboden is aan een jeugdige binnen één cliënttraject. Het cliënttraject betreft altijd de jeugdige zoals opgenomen in de Wet op de Jeugdzorg. Zijn ouder(s) of stiefouder(s) of andere die de jeugdige als behorend tot hun gezin verzorgen en opvoeden (artikel 1, onderdeel d, zoals opgenomen in de Wet op de Jeugdzorg) zijn altijd een onderdeel van het cliënttraject. Een cliënttraject kan starten in het vrijwillig of gedwongen kader.

Een cliënttraject in het vrijwillig kader start op het moment dat Bureau jeugdzorg de cliënt accepteert. Dit is het moment dat 'Vervolg binnen Bureau Jeugdzorg' is bepaald. De datum start cliënttraject is gelijk aan de datum dat Bureau Jeugdzorg de cliënt accepteert;

Een cliënttraject in het gedwongen kader start als een maatregel Jeugdbescherming of Jeugdreclassering start. Een maatregel jeugdbescherming start op de datum waarop de eerste beschermingsmaatregel start (aanvang eerste maatregel) c.q. de datum opgenomen in de beschikking en een maatregel jeugdreclassering start op de dag dat de Raad voor de Kinderbescherming of de Officier van Justitie de jeugdige bij bureau jeugdzorg aanmeldt voor een maatregel jeugdreclassering.

Één cliënttraject kan bestaan uit een of meer van de volgende soorten deeltrajecten (J&O, Jeugdbescherming, Jeugdreclassering). Een J&O (zorg) - traject is: de periode waarin voortdurend – op grond van één of meerdere indicatiebesluiten - provinciaal gefinancierde geïndiceerde jeugdzorg wordt geleverd aan de cliënt door één organisatie Jeugd & Opvoedhulp. Ook als deze jeugdzorg, geleverd door één organisatie Jeugd & Opvoedhulp meer dan één zorgvorm omvat, spreken we over één J&O (zorg) traject.

gegevensbestanden. De structuur en de opbouw van de gegevensbestanden verschilt ook door het gebruik van uiteenlopende registratiesystemen (CARE4, IJZA, REGAS). Daarbij komt dat een koppeling tussen de registratiesystemen en BergOp lastig te maken is, omdat een eenduidig identificatienummer (zoals het BSN) nog niet in gebruik genomen is.

Voor de benaderde instellingen geldt dat zij sinds 2007 onderzoek doen. In 2007 (en soms 2008) werden de vragenlijsten verwerkt in het oude programma van BergOp. Sinds 2008 werken de instellingen J&O met een nieuw BergOp programma. Het oude en het nieuwe programma BergOp bleken niet aan elkaar gekoppelde, maar twee afzonderlijke databestanden op te leveren. Dit had voor de datalevering tot gevolg dat de focus kwam te liggen op onderzoeksgegevens uit de jaren 2008 en 2009.

Gedurende de werkbezoeken werd duidelijk dat een aantal variabelen door instellingen voor J&O niet of moeilijk te leveren zijn. Dit heeft te maken met de keuze om deze variabele niet te registreren of met het feit dat deze variabele slecht door hulpverleners geregistreerd wordt. Het gaat om de volgende variabelen:

1. kenmerken van cliënten: gezinssamenstelling, juridisch kader, culturele achtergrond, aard van de problematiek, kwaliteit opvoedingsomgeving, onderwijs/dagbesteding;
2. kenmerken van de uitgevoerde hulp: intensiteit van de hulp;
3. kenmerken van de uitvoerende organisatie en hulpverleners: niet leverbaar op cliëntniveau;
4. resultaat van de hulp in termen van uitkomstvariabelen: kwaliteit opvoedingsomgeving, doelrealisatie, reden beëindiging hulp en cliënttevredenheid.

Wij waren nieuwsgierig naar wat de instellingen voor J&O noemden als meerwaarde om deel te nemen aan het project. Dit leverde de volgende punten op:

- zicht krijgen op effectiviteit van de jeugdzorg;
- meer transparantie creëren in de activiteiten en resultaten van de jeugdzorg;
- mogelijkheid voor gebruik bij vergelijkingsonderzoek;
- mogelijkheid voor gebruik bij validatie van instrumenten;
- zicht krijgen op activiteiten van andere instellingen voor J&O;
- uitwisseling en leren van andere instellingen voor J&O;
- bekendheid om als voorloper betrokken te zijn bij het project.

Om meer zicht te krijgen op de haalbaarheid en de kwaliteit van het TAZ-bestand hebben we voorafgaand aan de dataverzameling enkele proefbestanden verzameld en bekeken. Doel hiervan was op kleine schaal na te gaan in hoeverre en op welk niveau harmonisatie van gegevens mogelijk is en welke stappen daarvoor nodig zijn. Een verslag hiervan is te vinden in Bijlage 1.

2.3 Respons dataverzameling

In tabel 1 op de volgende pagina is het aantal onderzoekers en instellingen voor J&O weergegeven die zijn benaderd voor deelname, het aantal dat uiteindelijk deel wil nemen aan het project, het aantal waarvoor een overeenkomst is opgesteld en retour is ontvangen en het aantal dat uiteindelijk data geleverd heeft. Meer informatie over de hierbij betrokken onderzoekers en instellingen voor J&O is opgenomen in bijlage 2.

Tabel 1. Deelname project

	Benaderd	Deelname	Overeenkomst		Data levering
			Opgesteld	Retour	
Onderzoekers	13	8	8	6	6
Jeugd & Opvoedhulp	11	9	9	9	8
	24	17	17	15	14

NB. In de begeleidingscommissie is vastgesteld dat tien instellingen voor Jeugd & Opvoedhulp benaderd zouden worden. Dit is ook gebeurd. Er is één instelling toegevoegd die al deelnam aan de pilot van dit project.

In Tabel 1 is te zien dat dertien onderzoekers en elf instellingen voor J&O zijn benaderd. Daarvan namen vijf onderzoekers en drie instellingen voor J&O niet deel aan het project, zij leverden geen data (zie rechterkolom). Twee onderzoekers hadden wel interesse, maar hadden de pech dat hun computer met daarop de databestanden gestolen was. Twee onderzoekers wilden niet deelnemen omdat ze de visie en de inhoud van het project niet ondersteunden. Van één onderzoeker konden we helaas de adresgegevens niet achterhalen. Twee onderzoekers hadden de intentie om deel te nemen, maar zijn door persoonlijke omstandigheden afgehaakt. Twee instellingen voor J&O wilden niet deelnemen gezien de hoeveelheid informatie die geleverd moest worden en het gebrek aan tijd en mankracht om deze gegevens te leveren. Eén instelling voor J&O had wel de intentie deel te nemen aan het project, maar moest op het laatste moment afhaken door gebrek aan tijd en mankracht om data aan te leveren.

2.4 Privacy, eigendom en rechten

Om privacy, eigendom en rechten over de datagegevens vast te leggen is, zoals in rapportage fase 1 vermeld staat, een ‘Overeenkomst levering’ en een ‘Overeenkomst gebruik’ door het Nederlands Jeugdinstituut met de desbetreffende onderzoeker en de directeurs van de instellingen voor J&O opgesteld. In de ‘Overeenkomst levering’ geeft de onderzoeker of instelling toestemming voor het opnemen van de gegevens in het TAZ-bestand en voor gebruik van de gegevens door een derde, mits deze derde een ‘Overeenkomst gebruik’ heeft getekend. De ‘Overeenkomst gebruik’ geeft een derde toestemming om gebruik te maken van het TAZ-bestand.

In Tabel 1 is weergegeven of een ‘Overeenkomst levering’ is opgesteld en retour is ontvangen. Bij twee onderzoekers is de overeenkomst niet retour ontvangen wegens persoonlijke omstandigheden van een van de onderzoekers. Van in totaal zes onderzoekers en negen instellingen voor J&O hebben we de overeenkomsten ondertekend ontvangen.

2.5 Databewerking

De gegevens die werden aangeleverd door de instellingen voor J&O bestonden enerzijds uit achtergrondgegevens afkomstig van registratiesystemen zoals Care4, IJza, etc. en anderzijds uit vragenlijstgegevens van het programma BergOp. Daarnaast waren er gegevens afkomstig van de onderzoekers.

De grootste klus was het samenvoegen van de verschillende bestanden in één groot TAZ-bestand. Als eerste stap hierin zijn per bestand de variabelennamen en bijbehorende waarden geharmoniseerd. De harmonisatie van de variabelen met betrekking tot doelrealisatie en cliënttevredenheid bleek erg

ingewikkeld. Dit heeft te maken met verschil in terminologie, scoring en gebruik van verschillende vragenlijsten. Met het oog op de beschikbare tijd binnen dit project is besloten deze twee aspecten voor latere analyses (buiten dit project) te bewaren. Een lijst van alle gegevens in het TAZ-bestand en de bijbehorende N zijn te vinden in Bijlage 3.

Nadat de gegevens waren geharmoniseerd, zijn alle losse databestanden samengevoegd tot één groot TAZ-bestand. Enerzijds moesten gegevens binnen een aantal instellingen aan elkaar gekoppeld worden (achtergrondgegevens en vragenlijstgegevens) en anderzijds moesten gegevens van verschillende instellingen aan elkaar gekoppeld worden. Het samenvoegen van gegevens binnen instellingen leverde in sommige gevallen problemen op, omdat in de bestanden geen unieke kenmerken waren opgenomen om de verschillende bestanden aan elkaar te matchen. Hierdoor waren we genoodzaakt de gegevens van die cliënten uit het bestand te verwijderen.

Buiten de reductie van gegevens door moeilijkheden met het koppelen van bestanden, hebben we het bestand verder gereduceerd tot een bruikbaar bestand door een aantal criteria te hanteren voor de opbouw van het TAZ-bestand. Dit heeft er toe geleid dat:

- alleen die cliënten zijn opgenomen waarvan een aanvang- en eindmeting aanwezig is;
- alleen die metingen zijn opgenomen waarvan de voor- en eindmeting verricht is door dezelfde informant (moeder, vader, jongere, leerkracht of hulpverlener);
- per cliënt per informant maximaal één record opgenomen is (zogeheten duplicates zijn verwijderd)⁴;
- alleen die cliënten zijn opgenomen die op indicatie J&O-hulp hebben ontvangen (verwijderd zijn cliënten die vrij toegankelijke hulp ontvingen);
- alleen die cliënten zijn opgenomen waarvan een of meerdere gegevens over de geboden zorg bekend zijn.

Het hanteren van bovenstaande criteria leverde een reductie op van het TAZ-bestand van ongeveer 50%. Het uiteindelijke TAZ-bestand bevat in totaal 6596 records met 5906 unieke casussen.

⁴ Per cliënt kunnen meerdere records voorkomen. Deze hebben dan betrekking op vragenlijstgegevens van verschillende informanten. Analyses en matchingprocedures zijn alleen met het TAZ-bestand adequaat uit te voeren als er wordt geselecteerd op vragenlijst en informant.

3. Resultaten TAZ-bestand

3.1 Inleiding

In dit hoofdstuk rapporteren we over de resultaten die het TAZ-bestand oplevert. In paragraaf 3.2 gaan we eerst in op de kenmerken van de jeugdigen. In paragraaf 3.3 komt aan bod op welke instrumenten in het bestand de uitkomsten van de hulp worden bepaald. De daarop volgende paragrafen 3.4 tot en met 3.10 behandelen de uitkomsten van de hulp per themagebied, instrument en informant. In paragraaf 3.11 geven we een impressie van de wijze van beëindigen en de duur van de hulp. Het hoofdstuk besluit met een discussie.

Omdat de instellingen voor J&O niet van alle cliënten in zorg gegevens hebben geregistreerd en vervolgens naar ons hebben aangeleverd, is het TAZ-bestand niet representatief⁵ voor de totale populatie in die werkveld. De uitkomsten kunnen hierdoor niet gezien worden als *de* kenmerken van de cliëntèle en *de* resultaten van de thans aangeboden zorg in de jeugd- en opvoedhulp. We komen daar later nog op terug.

3.2 Kenmerken van jeugdigen

In Tabel 2 op de volgende pagina zijn de belangrijkste demografische gegevens van de jeugdigen en de ouders in de unieke casussen (zie toelichting paragraaf 2.5) weergegeven. Leeftijd, sekse en hulpvorm zijn in het bestand vaak bekend. Dat geldt niet voor de meeste andere variabelen, waarvan steeds voor ongeveer de helft van de cases de kenmerken bekend zijn. Het is opvallend dat veel relatief eenvoudig te registeren gegevens, die ook voor de behandeling niet onbelangrijk zijn, ontbreken. Zo zijn de gezinssituatie en de leefsituatie maar in ongeveer de helft van de gevallen bekend. Het juridisch kader is slechts bij 13% bekend. Gaan we uit van de gegevens die bekend zijn, dan laat de tabel het volgende beeld zien.

Kenmerken jeugdigen

De gemiddelde leeftijd van de jeugdige is 11,1 jaar. Meer dan de helft van de jeugdigen bestaat uit jongens. Jeugdigen met een autochtone achtergrond zijn oververtegenwoordigd. Bijna evenveel jeugdigen komen uit een éénouder- als uit een tweeoudergezin. De meeste jeugdigen gaan naar school. Een overgrote meerderheid komt vanuit de thuissituatie bij start hulp. Ongeveer driekwart van de jeugdigen ontvangt een vorm van ambulante hulp. Een klein percentage van de jeugdigen ontvangt daghulp, residentiële hulp of pleegzorg. Meer dan de helft van de jeugdigen start de hulp vrijwillig. De rest (40%) heeft een maatregel in de vorm van een OTS of Voogdij.

Kijkend naar de aard van de problematiek, dan valt op dat bij meer dan driekwart het accent ligt op gezin en opvoeding en psychosociaal functioneren.

⁵ Zoals in hoofdstuk 2 vermeld staat zijn data ontleend aan onderzoekers en aan instellingen voor Jeugd & Opvoedhulp. De data ontleend van onderzoekers (zie Bijlage 2) geeft wel zicht op de respons en non-respons. Daarentegen geven de data ontleend van instellingen voor J&O geen zicht op respons en non-respons.

Hierna worden kort vermeldenswaardige verschillen in achtergrondkenmerken tussen jongens en meisjes, allochtone en autochtone jeugdigen en jeugdigen met en zonder maatregel in relatie tot elkaar besproken. Alleen significante verschillen worden genoemd. Aangezien het een groot bestand betreft, betekent een significant verschil niet per definitie dat het om een groot verschil gaat.

Tabel 2. Demografische kenmerken jeugdigen en ouders in unieke casussen (N = 5906)

Kenmerken	Aantal (en %) cases kenmerk bekend	Rubrieken	
Kenmerken van de cliënten			
LEEFTIJD	5730 (97%)	Gemiddelde leeftijd (en standaarddev.) Range	11,1 (4,4) 0 – 19
SEKSE	5893 (99%)	Meisje Jongen	42% 58%
ETNISCHE ACHTERGROND	3161 (54%)	Nederlands Marokkaans Antilliaans Turks Surinaams Anders	62% 4% 5% 5% 9% 16%
GEZINSSITUATIE	3206 (54%)	Eénouder Tweeouder Niet relevant vanwege leeftijd	47% 47% 6%
LEEFSITUATIE BIJ AANVANG	3119 (53%)	Thuis Familie/vrienden Zelfstandig wonend Anders	83% 3% 1% 13%
ONDERWIJS / DAGBESTEDING	3192 (54%)	School Werkend Anders	86% 2% 13%
AARD PROBLEMATIEK (% VOOR ZOVER AANWEZIG)	2723 (46%)	Psychosociaal functioneren Lichamelijke gezondheid Vaardigheden en cogn. ontwikkeling Gezin en opvoeding Jeugdige en omgeving Overige problemen Niet-gespecificeerde problematiek	71% 16% 38% 88% 51% 31% 24%

Vervolg Tabel 2.

Hulpverleningskenmerken

JURIDISCH KADER		766 (13%)	
	Vrijwillige hulp		60%
	OTS/VOTS		16%
	Voogdij/voorlopige voogdij		7%
	Jeugdreclassering		3%
	Anders ⁶		15%
HULPVORM*		5609 (95%)	
	Pedagogische thuishulp		40%
	Ambulante hulp		39%
	Dagbehandeling		11%
	Pleegzorg		3%
	Residentieel		6%
	Crisisopvang		<1%
	Niet gespecificeerd verblijf		<1%

* Hulpvorm is als volgt geoperationaliseerd:

- *Pedagogische thuishulp* bevat: IOG(-EV), Hulp aan huis, Crisishulp aan huis, IAG/VHT, IPT, FF;
- *Ambulante hulp* bevat: ambulante hulp, OVG, HEO, MFA, MST, FFT, PMTO, MDFT, AVL, AIJK, ASH, begeleide omgang, gezinsbehandeling, gezin op maat, Stop 4-7, crisis ambulante, kokos, coaching (op school), triple P, Vertrektraining, zelfstandigheidstraining;
- *Dagbehandeling* bevat: daghulp (onderwijs), MKD, Boddaert;
- *Pleegzorg* bevat: pleegzorg
- *Residentieel* bevat: leefgroep, fasegroep, KTC, OHV, logeershuis, observatie behandelgroep, 16+, cursusgroep;
- *Crisisopvang* bevat: crisisopvang (time-out);
- *Niet gespecificeerd verblijf* bevat: pedagogisch verblijf

Sekse

Kijken we naar het kenmerk 'sekse' dan gelden de volgende nuances:

- Meisjes zijn bij opname gemiddeld 11.9 jaar, jongens 10.5 jaar ($t(5719)=-13.15, p<.00$).
- Jongens wonen vaker dan meisjes in een tweeoudergezin. Meisjes wonen dus vaker dan jongens in eenoudergezin ($\chi^2(2)=8.21, p=0.016$).
- Jongens komen vaker dan meisjes uit de thuissituatie bij aanvang hulp ($\chi^2(3)=29.35, p=0.00$).
- Jongens hebben vaker dan meisjes problemen met vaardigheden ($\chi^2(1)=39.92, p=0.00$) en hebben vaker gezins- en opvoedingsproblemen ($\chi^2(1)=8.77, p=0.003$).
- Jongens hebben vaker dan meisjes ambulante hulp, daghulp, niet gespecificeerd verblijf of pedagogische thuishulp. Daarentegen hebben meisjes vaker dan jongens crisishulp of residentiële hulp ($\chi^2(6)=45.04, p=0.00$).

Etniciteit

Er is een verschil tussen allochtone en autochtone jeugdigen en het hebben van wel of geen maatregel: autochtone jeugdigen hebben vaker dan allochtone jeugdigen geen maatregel ($\chi^2(1)=24.66, p=0.00$).

Maatregel

Het al dan niet hebben van een maatregel is bij veel cases onbekend. Kijken we in de groep waarbij dit kenmerk wel is geregistreerd, dan zien we dat de jeugdigen met een maatregel:

- gemiddeld ouder zijn dan degenen zonder maatregel ($t(764)=-2.8, p<.005$);
- vaker problematiek op het gebied 'omgeving van de jeugdige' hebben ($\chi^2(1)=7.44, p=0.006$);

⁶ De categorie anders is zo aangeleverd door instellingen voor J&O.

- vaker residentiële hulp ontvangen ($\chi^2(4)=23.06, p=0.00$).

Jeugdigen zonder een maatregel:

- komen vaker uit een eenoudergezin ($\chi^2(2)=27.79, p=0.00$);
- komen vaker als leefsituatie bij aanvang van de hulp van 'thuis' ($\chi^2(3)=30.74, p=0.00$);
- ontvangen vaker pedagogische thuishulp, ambulante hulp of daghulp ($\chi^2(4)=23.06, p=0.00$);
- zijn vaker van autochtone afkomst ($\chi^2(1)=24.66, p=0.00$).

3.3 Instrumenten, informanten en indexen

Over welke vragenlijsten en informanten zijn de resultaten te tonen? In Tabel 3 is het aantal complete voor- en nametingen per vragenlijst vermeld waarop de analyses zijn uitgevoerd, met daarbij weergegeven door welke informant deze vragenlijst is ingevuld.

Tabel 3. Aantal ingevulde voor- en nametingen per vragenlijst uitgesplitst per informant

	Moeder	Vader	Beiden	Jeugdige	Ander	Hulpverlener	Onbekend	% totaal (N=5906) *
STEP	0	0	0	0	0	2877*	0	49 %
YSR	0	0	0	224*	0	0	0	4%
SDQj	0	0	0	0	0	0	795*	11%
CBCL1½-5	93*	53	0	0	0	0	0	2 %
CBCL6-18	2438*	340	17	0	57	14	14	41%
SDQ-ou	82	46	1	0	4	0	1473*	26 %
SDQ-L	0	0	0	0	0	0	451*	7%
TRF 0-5/8-18	0	0	0	0	0	0	4	< 0%
NOSI	676*	267	1	0	9	0	7	11 %

STEP = Standaard Taxatie Ernst Problematiek; YSR = Youth Self Report; SDQj = Strengths and Difficulties Questionnaire – Jongerenversie; CBCL 1½-5 = Child Behavior Check List (age 1½-5); CBCL 6-18 = Child Behavior Check List (age 6-18); SDQ-ou = Strengths and Difficulties Questionnaire – ouderversie; SDQ-L: Strengths and Difficulties Questionnaire – leerkrachtversie; TRF 0-5/6-18 : Teacher Report Form (age 0-5/6-18); NOSI = Nijmeegse Opvoeder Stress Index.

* = gebruikt bij rapportage uitkomsten hulp; % totaal = percentage van totaal aantal cases in het bestand (N=5906) gebruikt bij rapportage uitkomsten hulp.

Bij de STEP rapporteren we de gegevens zoals ingevuld door de hulpverlener. Bij de CBCL en de NOSI maken we gebruik van de informant 'moeder'. Ondanks het feit dat bij de SDQ-jeugdige/ ouders/ leerkracht de informant vaak onbekend is (we weten niet of vader, moeder of een ander de SDQ heeft ingevuld), hebben we besloten toch analyses uit te voeren. Tabel 3 laat zien dat het aantal cases dat voor de analyses is te gebruiken om de uitkomsten van de hulp te bepalen niet boven de vijftig procent komt. Het meest komt nog de STEP voor (49%). Het kleinste aantal cases (0-2%) geldt voor de TRF en CBCL 1½-5.

Voor de TRF met leerkracht als informant is het aantal compleet ingevulde lijsten te laag (N=4) om zinvolle analyses op uit te voeren.

In de volgende paragrafen worden de uitkomsten gepresenteerd van de ingevulde voor- en nametingen op de vragenlijsten die in het bestand voorkomen. Belangrijk is daarbij in de gaten te houden dat de vragenlijstgegevens niet voor alle cases bekend zijn en van verschillende informanten afkomstig kunnen zijn.

De beschikbare tijd in het project stond ons niet toe uitgebreide analyses te doen. Voor zover mogelijk zijn per vragenlijst de gemiddelde scores op de begin- en de eindmeting vergeleken. De mate van verandering in de problematiek zoals gemeten op de vragenlijsten, is daarbij weergegeven met behulp van de *Effect Size / ES*. Er is veel discussie over de vraag hoe deze index in de situatie met alleen een voor- en nameting (zonder controlegroep) berekend moet worden. Na wikken en wegen hebben wij ervoor gekozen om een strenge variant te gebruiken, zodat we in ieder geval een schatting hebben van de mate van verandering waar men minstens van uit mag gaan. De gerapporteerde effectgroottes moeten hier als volgt worden geïnterpreteerd:

- -0,21 of lager staat voor een negatieve verandering;
- -0,20 t/m 0,19 staat voor een verwaarloosbare verandering;
- 0,20 t/m 0,49 staat voor een kleine positieve verandering;
- 0,50 t/m 0,79 staat voor een middelmatig positieve verandering;
- 0,80 of meer staat voor een grote positieve verandering (Cohen, 1992; zie ook Veerman, 2008).

Voor een nadere toelichting op de gekozen index zie bijlage 4.

Naast de algemene mate van verandering is er gekeken naar de mate waarin er een betrouwbare negatieve of positieve verandering op individueel niveau is. Dat gebeurde met behulp van de zogeheten *Reliable Change Index / RCI*. De RCI wordt gebruikt om vast te stellen bij hoeveel jeugdigen sprake is van een 'betrouwbare verandering' tussen aanvang en einde hulp.⁷ Tot slot is meer gedetailleerd geanalyseerd wat de mate van verandering is afhankelijk van de ernst van de problematiek bij de beginmeting.

De verschillende berekeningen bleken niet altijd voor elke vragenlijst mogelijk wegens het ontbreken van bepaalde gegevens in het bestand, c.q. wegens tijdgebrek om die gegevens alsnog aan te vullen. Ook waren bepaalde analyses niet zinvol vanwege de kleine groepsgrootte. Die zijn dan ook in de rapportage weggelaten.

Daarnaast is er per vragenlijsten gekeken naar de verschillen tussen de diverse hulpvormen. Daarbij is een onderscheid gemaakt tussen pedagogische thuishulp, ambulante hulp, daghulp en hulp bij uithuisplaatsing. Dit laatste omvat pleegzorg, residentiële hulp, crisishulp, en niet gespecificeerd verblijf.

3.4 STEP: verandering ernst problematiek volgens hulpverlener

Tabel 4 op de volgende pagina toont de gemiddelde scores van hulpverleners op de schalen van de STEP bij aanvang en einde van de hulp. Daarbij is een onderscheid gemaakt tussen de STEP-scores inclusief de items die gaan over de vraag hoe lang de problemen al bestaan en exclusief deze items.

⁷ De formule van de Reliable Change Index (RCI) luidt: $RCI = (\text{score einde} - \text{score aanvang}) / \text{standaardfout van het verschil}$. De standaardfout wordt bepaald door de standaard-meetfout en de test-hertest betrouwbaarheid van het betreffende meetinstrument. Negatieve verandering = $RCI < -1.64$; geen verandering = $-1.64 \leq RCI \leq +1.64$; positieve verandering = $RCI > +1.64$. Bij tweezijdige toetsing wordt een $p \leq .10$ gehanteerd (Veerman, 2008, p. 326). De RCI is een zeer strenge toets waarbij alleen substantiële of betrouwbare verandering als 'verandering' wordt aangemerkt.

Tabel 4. Gemiddelde scores (en standaarddeviatie SD) STEP begin en einde hulp

	Begin	Einde	t-toets	ES
Inclusief items duur problematiek				
Functioneren jeugdige ^A (N=2877)	17.68 (3.96)	16.03 (4.80)	T=22.8, p=0.00*	0.41
Kwaliteit omgeving ^B (N=2832)	14.14 (4.16)	12.96 (4.61)	T=16.3, p=0.00*	0.27
Exclusief items duur problematiek				
Functioneren jeugdige ^C (N=2877)	13.12 (3.49)	11.71 (3.94)	T=23.2, p=0.00*	0.38
Kwaliteit omgeving ^D (N=2832)	9.94 (3.30)	9.02 (3.49)	T=16.8, p=0.00*	0.27

^A 6-9 = functioneert goed; 10-14 = functioneert redelijk, 15-20 = matige problemen; 21-25 = aanzienlijke problemen; 26-30 = zware tot extreme problemen.

^B 5-7 = kwaliteit goed; 8-12 = kwaliteit redelijk, 13-17 = matige problemen; 18-22 = aanzienlijke problemen; 23-25 = zware tot extreme problemen.

^C 5-7 = functioneert goed; 8-12 = functioneert redelijk, 13-17 = matige problemen; 18-22 = aanzienlijke problemen; 23-25 = zware tot extreme problemen.

^D 3-5 = kwaliteit goed; 6-10 = kwaliteit redelijk, 11-14 = matige problemen; 15-18 = aanzienlijke problemen; 19-22 zware tot extreme problemen.

ES = Effect Size.

Volgens de hulpverleners hebben de jeugdigen matige problemen bij aanvang en einde van de hulp. Uit de gepaarde t-toets blijkt dat het functioneren van de jeugdigen bij einde hulp significant verbeterd is. Ook is de kwaliteit van de omgeving bij einde hulp significant verbeterd. De effect size laat een kleine verandering zien.

Die kleine verandering is voor een deel logisch vanwege de relatief lage beginscore van veel cases (en daarmee ook het lage gemiddelde op de beginscore). Bij een dergelijke score is er immers per definitie minder resultaat te boeken dan wanneer de beginscore heel hoog is. Meer in het algemeen geldt dat de overall effect size verschillen in de mate van verandering bij uiteenlopende groepen kan verbloemen. Bij Jeugdformaat, een instelling voor jeugdzorg in Haaglanden, zijn eerder aanwijzingen gevonden dat dit zeker het geval is als er onderscheid gemaakt wordt naar de mate van ernst van de problematiek bij de start van de hulp (Reith & Stals, persoonlijke communicatie). Hoe ernstiger de problematiek, hoe hoger de effect size, zo bleek. Dat blijkt ook uit het TAZ-bestand. Tabel 5 geeft daarvan een voorbeeld. Het laat zien wat de effect sizes zijn binnen de groepen met verschillende mate van ernst van de problematiek bij aanvang gemeten met de schaal Functioneren Jeugdige van de STEP (zonder item over de duur van de problematiek).

Tabel 5. Gemiddelde scores (en SD) STEP-FJ begin en einde hulp binnen ernstklassen (N=2877)^A

Ernstklasse begin FJ	Begin	Einde	t-toets	ES
Functioneert goed (N=183)	6.07 (0.84)	6.91 (2.50)	T=-4.41, p=0.00*	-0.45
Functioneert redelijk (N=1018)	10.58 (1.34)	9.95 (3.11)	T=7.01, p=0.00*	0.26
Matige problemen (N=1385)	14.59 (1.33)	12.72 (3.26)	T=22.41, p=0.00*	0.75
Aanzienlijke problemen (N=283)	19.26 (1.24)	16.07 (3.90)	T=14.11, p=0.00*	1.10
Zware tot extreme problemen (N=8)	24.00 (1.07)	17.00 (6.07)	T=3.29, p=0.00*	1.37

^A FJ = Functioneren jeugdige. Scores zonder item duur problematiek. Beginklassen: 5-7 = functioneert goed; 8-12 = functioneert redelijk, 13-17 = matige problemen; 18-22 = aanzienlijke problemen; 23-25 = zware tot extreme problemen. SD = Standaarddeviatie; ES = Effect Size.

Duidelijk is dat de effectgrootte sterk samenhangt met de ernstscore aan het begin. Hoe lager de ernstscore, hoe kleiner de verandering. Bij een aanvankelijk goed functioneren is er zelfs sprake van een achteruitgang, wellicht door een verbeterd diagnostisch beeld. De score blijft echter binnen 'het

goede gebied'. Bij ernstiger problemen bij aanvang is er sprake van een grote tot zeer grote verandering. Dat komt voor een deel door het feit dat bij ernstiger problemen gewoonweg meer verandering te behalen valt. Voor een ander deel is de hogere effect size te wijten aan een statistisch artefact: hoe extremer de beginscore, hoe groter de statistische kans om bij het einde lager te scoren. De scores komen immers steeds dichterbij het plafond van de schaal en kunnen daarom niet in dezelfde mate stijgen als dat ze kunnen dalen. In de discussieparagraaf van dit hoofdstuk komen we op de betekenis van dit soort cijfers terug.

Overigens zien we dat ondanks een grote verandering bij een hoge beginscore ook de eindscores nog betrekkelijk hoog zijn ('matige problemen'). De cliënten zijn dus – gemiddeld gesproken – nog niet uit de 'gevarezone'.

Er is een verschil in de mate van verandering van de ernstscores tussen de verschillende hulpvormen ($F(2,2581)=7.98, p<0.00$). Uit een post-hoc vergelijking volgens de Scheffé-methode blijkt dat er een significant verschil is tussen ambulante hulp en uithuisplaatsing ($p<0.01$). De verschillingscore is bij ambulante hulp groter dan bij uithuisplaatsing. De verschillingscore is bij daghulp vergelijkbaar met uithuisplaatsing.

3.5 YSR: verandering problematiek jeugdige volgens jeugdige

In Tabel 6 zijn de gemiddelde normatieve deviatiescores op de YSR bij aanvang en einde van de hulp weergegeven.

Tabel 6. Gemiddelde normatieve deviatiescores (en SD) op YSR bij begin en einde hulp

		Begin	Einde	t-toets	ES
Internaliseren	(N=215)	.61 (1.22)	.25 (1.18)	T=4.65, $p=0.00^*$	0.30
Externaliseren	(N=224)	.65 (1.09)	.39 (1.12)	T=3.73, $p=0.00^*$	0.24
Totale problemen	(N=224)	.69 (1.08)	.36 (1.14)	T=4.63, $p=0.00^*$	0.30

Noot: Normatieve deviatiescores geven de afwijking ten opzichte van de norm aan; interpretatie: <1.00: geen problemen; 1.00-1.28 = matige problemen; 1.29-1.64 = aanzienlijke problemen; 1.65-1.96 = ernstige problemen; >1.96 = zeer ernstige problemen (Veerman, 2008, p. 313). SD = Standaarddeviatie; ES = Effect Size.

De scores in Tabel 6 bij aanvang en einde hulp laten zien dat gemiddeld genomen jeugdigen zelf geen noemenswaardige problemen ervaren. Uit gepaarde t-testen komt desondanks naar voren dat er sprake is van een vermindering van de problemen: een vergelijking van de deviatiescores bij aanvang en einde hulp laten significante verschillen zien op de schalen internaliseren, externaliseren en totale problemen. De effectgrootte ES laat zien dat het om kleine veranderingen gaat.

Vervolgens is de Reliable Change Index (RCI) berekend. Tabel 7 toont het percentage jeugdigen waarbij sprake is van een negatieve, geen of een positieve verandering van probleemgedrag op basis van de RCI.

Tabel 7. Aantal jeugdigen waarbij sprake is van een verandering van probleemgedrag op de YSR

	Negatieve verandering	Geen verandering	Positieve verandering
Internaliseren (N=215)	13%	54%	33%
Externaliseren (N=224)	21%	35%	43%
Totale problemen (N=224)	3%	85%	12%

Meer dan de helft van de jeugdigen geeft aan geen verandering op internaliserend probleemgedrag te ervaren. Daarentegen ervaart een derde een positieve verandering met internaliserende problemen. Wat betreft externaliseren ervaart iets minder dan de helft positieve veranderingen en een derde geen verandering. De totale problemen zijn grotendeels niet veranderd. Ook hier geldt het fenomeen dat er weinig of zelfs een negatieve verandering is te verwachten bij een lage beginscore en veel (positieve) verandering bij een hoge beginscore.

Uitgaand van de RCI's hebben we bepaald of er een verschil is in de mate van verandering van totale problemen tussen de verschillende soorten hulp. Dat blijkt hier het geval ($F(3, 220)=3.30, p<0.05$). Uit een post-hoc vergelijking volgens de Scheffé-methode blijkt dat er een significant verschil is tussen ambulante hulp en uithuisplaatsing ($p<0.05$). De verschillscore is voor jeugdigen die ambulante hulp ontvangen groter in vergelijking met de jeugdigen die hulp ontvangen bij uithuisplaatsing.

3.6 SDQ-j: verandering problematiek jeugdige volgens jeugdige

Wegens het ontbreken van normatieve deviatiescores in de aangeleverde bestanden presenteren we hieronder de gemiddelde ruwe scores. In Tabel 8 staan de gemiddelde ruwe scores op de SDQ bij begin en einde van de zorgperiode weergegeven. Eveneens is de ES berekend. We gaan ervan uit dat de waarde van de ES niet verandert door het gebruik van de ruwe scores. Voorts moet bij de resultaten worden aangetekend dat de subschalen van de SDQ uit een gering aantal items bestaan. Dat maakt de veranderscores gevoelig voor toevallige fluctuaties en daarom minder betrouwbaar. Omwille van de vergelijkbaarheid met de YSR zijn de subschalen hier wel gemeld, maar feitelijk is alleen de totaalscore als een belangrijke indicator voor de mate van verandering te gebruiken.

Tabel 8. Gemiddelde ruwe scores (en SD) op de SDQ-jeugdige bij begin en einde hulp (N=795)

	Begin	Einde	t-toets	ES
Emotionele symptomen (N=795)	3.38 (2.39)	2.80 (2.28)	T=7.38, p=0.00*	0.25
Gedragsproblemen (N=795)	3.02 (1.89)	2.56 (1.74)	T=7.16, p=0.00*	0.25
Hyperactiviteit-Aandachtstekort (N=795)	4.71 (2.47)	4.38 (2.44)	T=4.37, p=0.00*	0.13
Problemen met leeftijdsgenoten (N=795)	2.31 (1.81)	2.17 (1.73)	T=2.22, p=0.00*	0.08
Prosociaal gedrag (N=795)	7.84 (1.74)	8.07 (1.69)	T=-3.66, p=0.00*	-0.13
Totale problemen (N=795)	13.41 (5.49)	11.91 (5.47)	T=8.52, p=0.00*	0.27

Noot: SD = Standaarddeviatie; ES = Effect Size.

Uit de gepaarde t-toets komt naar voren dat de totale problemen een positieve verandering over de tijd laten zien. Er is sprake van een kleine verandering. Op subschaalniveau is de grootste verandering te zien bij Gedragsproblemen en Emotionele symptomen en de kleinste bij problemen met leeftijdsgenoten.

In Tabel 9 op de volgende pagina laten we zien welke relatie er is tussen de hoogte van de beginscore en de effectgrootte. We hebben daarvoor de ruwe score aan de hand van percentielscores in klassen ingedeeld, om een vergelijkbare ordening te krijgen als bij de ernstindeling van de CBCL-scores.

Tabel 9. Gemiddelde scores (en SD) Totale problemen SDQ-jeugdige binnen ernstklassen (N=795)

Ernstklasse begin	Begin	Einde	t-toets	ES
Geen probleem (N=317)	8.14 (2.55)	8.61 (4.48)	T=-2.02, p=0.04*	-0.13
Matige problemen (N=106)	12.44 (.50)	11.69 (4.30)	T=1.80, p=0.08	0.25
Aanzienlijke problemen (N=149)	15.01 (.82)	13.34 (4.20)	T=4.87, p=0.00*	0.55
Ernstige problemen (N=138)	18.38 (1.10)	15.30 (4.65)	T=7.78, p=0.00*	0.91
Zeer ernstige problemen (N=85)	23.39 (2.32)	16.47 (5.76)	T=11.65, p=0.00*	1.58

Noot: Beginklassen gebaseerd op ruwe score. Indeling indicatief op basis van percentielscore voor Amerikaanse Jeugdigen www.sdqinfo.org <11: geen problemen; 12-13=matige problemen; 14-16= aanzienlijke problemen; 17-20 = ernstige problemen; >21= zeer ernstige problemen; SD = Standaarddeviatie; ES = Effect Size.

Bij de laagste klasse is er sprake van een hogere probleemscore aan het einde, wellicht door een verbeterd diagnostisch beeld of door een zekere mate van onbetrouwbaarheid in de scoring. Bij ernstige problemen bij aanvang is er sprake van een grote tot zeer grote verandering. Ook hier geldt weer: dat kan komen doordat er meer behandel-effect te halen is, en door de statistische kans om bij een hoge score aan het begin een lagere score aan het einde te krijgen. Ook hier laat de groep met hoge scores aan het begin tevens een hoge score aan het einde zien. De verandering in die groep is dus wel groot, maar niet groot genoeg om uit de problemen te zijn.

Overigens blijkt er geen verschil te zijn in de mate van verandering in de totale scores en de verschillende hulpvormen ($F(2, 785)=.96, p=0.38$).

3.7 CBCL: verandering problematiek jeugdige volgens moeders

3.7.1 Verandering gemeten met CBCL 1½ - 5

In Tabel 10 staan de gemiddelde normatieve deviatiescores op de CBCL volgens de moeders bij het begin en einde van de zorgperiode weergegeven.

Tabel 10. Gemiddelde normatieve deviatiescores (en SD) op de CBCL 1½ - 5 bij begin en einde hulp

	Begin	Einde	t-toets	ES
Internaliseren (N=93)	1.34 (1.48)	.59 (1.56)	T=7.26, p=0.00*	0.49
Externaliseren (N=93)	1.43 (1.28)	.49 (1.27)	T=8.62, p=0.00*	0.67
Totale problemen (N=93)	1.46 (1.32)	.49 (1.37)	T=9.25, p=0.00*	0.72

Noot: Normatieve deviatiescore <1.00: geen problemen; 1.00-1.28 = matige problemen; 1.29-1.64 = aanzienlijke problemen; 1.65-1.96 = ernstige problemen; >1.96 = zeer ernstige problemen (Veerman, 2008, p. 313). SD = Standaarddeviatie; ES = Effect Size.

Uit de gepaarde t-toets komt naar voren dat het internaliserende en het externaliserende probleemgedrag en de totale problemen van de jeugdigen volgens de moeders een positieve verandering over de tijd laat zien. Er is daarbij sprake van een kleine verandering voor internaliserend probleemgedrag en een middelgrote verandering voor externaliserend probleemgedrag en totale problemen.

Gezien de relatief kleine groepsgrootte (N=93) hebben we aanvullende analyses achterwege gelaten.

3.7.2 Verandering gemeten met CBCL 6 - 18

In Tabel 11 staan de gemiddelde normatieve deviatiescores op de CBCL volgens de moeders bij het begin en einde van de zorgperiode weergegeven.

Tabel 11. Gemiddelde normatieve deviatiescores (en SD) op de CBCL 6 - 18 bij begin en einde hulp

	Begin	Einde	t-toets	ES
Internaliseren (N=2435)	2.10 (1.86)	1.22 (1.71)	T=30.22, p=0.00*	0.49
Externaliseren (N=2435)	3.24 (1.99)	2.08 (1.95)	T=36.07, p=0.00*	0.59
Totale problemen (N=2438)	3.02 (1.87)	1.88 (1.85)	T=38.11, p=0.00*	0.61

Noot: Normatieve deviatiescore <1.00: geen problemen; 1.00-1.28=matige problemen; 1.29-1.64 = aanzienlijke problemen; 1.65-1.96 = ernstige problemen; >1.96 = zeer ernstige problemen (Veerman, 2008, p. 313). SD = Standaarddeviatie; ES = Effect Size.

Voor de CBCL 6 - 18 is eveneens een gepaarde t-toets uitgevoerd. We zien hier eenzelfde uitkomst: het internaliserende en het externaliserende probleemgedrag en de totale problemen van de jeugdigen laten volgens de moeders een positieve verandering over de tijd zien. Er is daarbij sprake van een kleine verandering bij internaliseren en een middelgrote verandering bij externaliseren en totale problemen.

Ook voor de CBCL 6 - 18⁸ is berekend bij hoeveel jeugdigen er sprake is van 'betrouwbare verandering' van gedragsproblemen op basis van de Reliable Change Index (RCI). Tabel 12 toont het percentage jeugdigen waarbij volgens moeders sprake is van een negatieve, geen of een positieve verandering van probleemgedrag op basis van de RCI.

Tabel 12. Aantal jeugdigen bij wie volgens de moeders sprake is van een verandering van probleemgedrag op de CBCL 6 - 18

	Negatieve verandering	Geen verandering	Positieve verandering
Internaliseren (N=2358)	12%	34%	55%
Externaliseren (N=2361)	8%	35%	57%
Totale problemen (N=2361)	2%	68%	30%

Tabel 12 laat zien dat bij een derde van de jeugdigen volgens de moeders het internaliserende en externaliserende probleemgedrag niet verandert en bij meer dan de helft het gedrag positief verandert. Op de totale problemen ziet driekwart van de opvoeders geen verandering in probleemgedrag.

Overigens geldt ook hier weer het algemene fenomeen dat er weinig of zelfs een negatieve verandering is te verwachten bij een lage beginscore en veel (positieve) verandering bij een hoge beginscore. Eerder illustreerden we dit getalsmatig met de scores op de schaal Functioneren Jeugdige van de STEP. De vraag is nu of deze verwachting voor de CBCL uitkomt. Tabel 13 toont op de volgende pagina het resultaat. We hebben daarbij gekeken naar de schaal Totale problemen van de CBCL. Omdat een zeer groot aantal cases in de categorie 'Zeer ernstige problemen' valt (N=1677), is deze verder opgesplitst in een aantal subcategorieën. We zien ook hier dat naarmate de beginscore op de CBCL hoger is, de grootte van de verandering toeneemt. Dat komt ook hier weer doordat bij ernstiger problemen meer verandering te halen valt en er een statistisch artefact in de effectscores sluipst. De relatie tussen de hoogte van de scores en de effect sizes is echter hier minder sterk dan we eerder bij de STEP zagen.

⁸ Voor de CBCL 1½ - 5 hadden we geen normscores tot onze beschikking.

Met name de rubrieken ‘Matige problemen’ en ‘Zeer ernstige problemen 2’ doorbreken de trend. Dat komt voor een deel door de fijnmazige indeling in de rubrieken ‘Matige problemen’ t/m ‘Ernstige problemen’ (de gemiddelde effect size van deze drie rubrieken is 0,83). In de discussieparagraaf gaan we verder op de betekenis van de cijfers in.

Tabel 13. Gemiddelde scores (en SD) Totale problemen CBCL 6-18 binnen ernstklassen (N=2438)

Ernstklasse begin	Begin	Einde	t-toets	ES
Geen problemen (N=380)	0.29 (0.55)	0.10 (1.01)	T=4.17; p=0.000	0.23
Matige problemen (N=104)	1.16 (0.09)	0.51 (0.99)	T=6.68; p=0.000	0.92
Aanzienlijke problemen (N=135)	1.49 (0.10)	0.88 (1.17)	T=6.12; p=0.000	0.73
Ernstige problemen (N=142)	1.80 (0.09)	1.14 (1.09)	T=7.16; p=0.000	0.85
Zeer ernstige problemen 1 (N=467)	2.50 (0.30)	1.54 (1.26)	T=16.73; p=0.000	1.05
Zeer ernstige problemen 2 (N=849)	3.91 (0.20)	2.43 (0.51)	T=29.69; p=0.000	3.82
Zeer ernstige problemen 3 (N=307)	5.78 (0.52)	3.74 (1.86)	T=20.11; p=0.000	1.49
Zeer ernstige problemen 4 (N=54)	7.61 (0.50)	5.33 (1.89)	T=9.04; p=0.000	1.65

Gem.
0.83

Noot: Beginklassen gebaseerd op Veerman (2008, p. 313): <1.00: geen problemen; 1.00-1.28=matige problemen; 1.29-1.64 = aanzienlijke problemen; 1.65-1.96 = ernstige problemen; >1.96 = zeer ernstige problemen (hier verder opgesplitst in 1.97-3.00 = Zeer ernstige problemen 1; 3.01-5.00 = Zeer ernstige problemen 2; 5.01-7.00 = Zeer ernstige problemen 3; > 7.00 = Zeer ernstige problemen 4). SD = Standaarddeviatie; ES = Effect Size.

Kijken we naar de groep met zeer ernstige problemen aan het begin, dan zien we net als bij de eerder besproken vragenlijsten dat de eindscores ook hoog zijn, variërend van aanzienlijke tot zeer ernstige problemen. Een grote mate van verandering betekent dus ook hier dat de cliënten nog niet uit de problemen zijn. Integendeel, de problemen zijn – gemiddeld gesproken – nog steeds zeer ernstig.

Er is overigens een verschil in de mate van verandering van probleemgedrag bij de jeugdigen tussen de verschillende hulpvormen. Dit geldt echter alleen voor internaliserend probleemgedrag ($F(3, 2354)=5.57, p<0.001$). Uit een post-hoc vergelijking volgens de Scheffé-methode blijkt dat er een significant verschil is tussen pedagogische thuishulp en daghulp ($p<0.02$). De verschillscore is pedagogische thuishulp groter in vergelijking met degene die daghulp ontvangen.

3.8 SDQ: verandering problematiek jeugdige volgens ouders

Vanwege het grote aantal onbekende informanten bij de SDQ-ouders zijn de analyses uitgevoerd met de informant ‘onbekend’, in de veronderstelling dat het hier meestal zal gaan om de ouders.

Net als bij de SDQ-jeugdige presenteren we hier de gemiddelde ruwe score, omdat de normatieve deviatiescores ontbreken. In Tabel 14 op de volgende pagina staan de gemiddelde ruwe scores op de SDQ bij begin en einde van de zorgperiode weergegeven. Eveneens is de ES berekend. We gaan ervan uit dat de waarde van de ES niet verandert door het gebruik van de ruwe scores. Bij de resultaten moet worden aangetekend dat de subschalen van de SDQ uit een gering aantal items bestaan. Dat maakt de veranderscores gevoelig voor toevallige fluctuaties en daarom minder betrouwbaar. Omwille van de vergelijkbaarheid met de CBCL zijn de subschalen hier wel gemeld, maar feitelijk is alleen de totaalscore als een belangrijke indicator voor de mate van verandering te gebruiken.

Tabel 14. Gemiddelde ruwe scores (en SD) op de SDQ-ouders bij begin en einde hulp (N=1473)

	Begin	Einde	t-toets	ES
Emotionele symptomen (N=1472)	3.89 (2.57)	3.15 (2.46)	T=12.32, p=0.00*	0.29
Gedragsproblemen (N=1472)	3.56 (2.28)	2.74 (2.16)	T=16.05, p=0.00*	0.37
Hyperactiviteit–Aandachtstekort (N=1472)	5.44 (2.82)	4.86 (2.78)	T=11.74, p=0.00*	0.21
Problemen met leeftijdsgenoten (N=1472)	2.93 (2.20)	2.53 (2.10)	T=8.00, p=0.00*	0.19
Prosociaal gedrag (N=1468)	7.15 (2.25)	7.58 (2.18)	T=-9.07, p=0.00*	-0.19
Totale problemen (N=1473)	15.92 (6.57)	13.28 (6.83)	T=18.08, p=0.00*	0.39

Noot: SD = Standaarddeviatie; ES = Effect Size.

Uit de gepaarde t-toets komt naar voren dat de totale problemen een positieve verandering over de tijd laten zien. Er is sprake van een kleine verandering. Op subschaalniveau is de grootste verandering te zien bij Gedragsproblemen en de kleinste bij Problemen met leeftijdsgenoten.

In Tabel 15 laten we zien welke relatie er is tussen de hoogte van de beginscore en de effectgrootte. We hebben daarvoor de ruwe score aan de hand van percentielscores in klassen ingedeeld, om een vergelijkbare ordening te krijgen als bij de ernstindeling van de CBCL-scores.

Tabel 15. Gemiddelde scores (en SD) Totale problemen SDQ-ouders binnen ernstklassen (N=1473)

Ernstklasse begin	Begin	Einde	t-toets	ES
Geen probleem (N=392)	7.67 (2.74)	7.90 (4.74)	T=-1.03, p=0.30	-0.06
Matige problemen (N=142)	12.51 (.50)	11.06 (5.06)	T=3.46, p=0.001*	0.40
Aanzienlijke problemen (N=238)	15.01 (.80)	12.40 (5.05)	T=7.93, p=0.00*	0.72
Ernstige problemen (N=325)	18.45 (1.11)	14.41 (5.49)	T=13.52, p=0.00*	1.02
Zeer ernstige problemen (N=376)	24.22 (3.03)	19.30 (6.16)	T=15.70, p=0.00*	1.01

Noot: Beginklassen gebaseerd op ruwe score. Indeling indicatief op basis van percentielscore voor Amerikaanse Jeugdigen www.sdqinfo.org <11: geen problemen; 12-13=matige problemen; 14-16= aanzienlijke problemen; 17-20 = ernstige problemen; >21= zeer ernstige problemen; SD = Standaarddeviatie; ES = Effect Size.

Net zoals bij de SDQ-jeugdige zien we ook bij de SDQ-ouders dat bij de laagste klasse er sprake is van een iets hogere probleemscore aan het einde, wellicht door een verbeterd diagnostisch beeld of door een zekere mate van onbetrouwbaarheid in de scoring. Bij ernstige problemen bij aanvang is er sprake van een grote tot zeer grote verandering. Ook hier geldt weer: dat kan komen doordat er meer behandel-effect te halen is, en door de statistische kans om bij een hoge score aan het begin een lagere score aan het einde te krijgen. Ook hier laat de groep met hoge scores aan het begin tevens een hoge score aan het einde zien. De verandering in die groep is dus wel groot, maar niet groot genoeg om uit de problemen te zijn.

Er blijkt geen verschil te zijn in de mate van verandering in de totale scores en de verschillende hulpvormen ($F(2, 1454)=3.78, p=0.023$; N.B. we hanteren hier een strengere grens voor de significantie ($p < 0.001$) omdat de verschillen niet met de RCI konden worden gecorrigeerd en daarmee een stapeling van meetfouten kan optreden).

3.9 SDQ: verandering problematiek jeugdige volgens leerkracht

Ook bij de SDQ-leerkracht is gebruik gemaakt van de informant 'onbekend', in de veronderstelling dat het hier meestal zal gaan om de leerkracht.

Ook hier geldt dat de normatieve deviatiescores ontbreken in de aangeleverde bestanden en we daarom de gemiddelde ruwe scores presenteren. In Tabel 16 staan de gemiddelde ruwe scores op de SDQ bij begin en einde van de zorgperiode weergegeven. Eveneens is de ES berekend. We gaan ervan uit dat de waarde van de ES niet verandert door het gebruik van de ruwe scores.

Tabel 16. Gemiddelde ruwe scores (en SD) op de SDQ-leerkracht bij begin en einde hulp (N=451)

	Begin	Einde	t-toets	ES
Emotionele symptomen (N=451)	2.92 (2.42)	2.83 (2.34)	T=-.80, p=0.42	0.04
Gedragsproblemen (N=451)	2.55 (2.39)	2.37 (2.34)	T=1.85, p=0.06	0.08
Hyperactiviteit-Aandachtstekort (N=451)	5.44 (3.08)	5.14 (3.07)	T=2.61, p=0.009*	0.10
Problemen met leeftijdsgenoten (N=451)	2.75 (2.29)	2.43 (2.17)	T=3.30, p=0.001*	0.14
Prosociaal gedrag (N=450)	5.77 (2.71)	6.23 (2.60)	T=-4.17, p=0.00*	-0.17
Totale problemen (N=451)	13.65 (6.61)	12.77 (6.55)	T=3.25, p=0.001*	0.13

Noot: SD = Standaarddeviatie; ES = Effect Size.

Uit de gepaarde t-toets komt naar voren dat de totale problemen een positieve verandering over de tijd laten zien. Er is sprake van een kleine verandering. Op subschaalniveau is de grootste positieve verandering te zien bij Problemen met leeftijdsgenoten en de kleinste bij Emotionele problemen. Echter de grootste verandering zien we bij Prosociaal gedrag. Dit betreft een negatieve verandering.

In Tabel 17 laten we zien welke relatie er is tussen de hoogte van de beginscore en de effectgrootte. We hebben daarvoor de ruwe score aan de hand van percentielscores in klassen ingedeeld, om een vergelijkbare ordening te krijgen als bij de ernstindeling van de CBCL-scores.

Tabel 17. Gemiddelde scores (en SD) Totale problemen SDQ-leerkracht binnen ernstklassen (N=451)

Ernstklasse begin	Begin	Einde	t-toets	ES
Geen probleem (N=179)	7.16 (2.94)	8.65 (5.02)	T=-4.36, p=0.00*	-0.36
Matige problemen (N=51)	12.47 (.50)	12.76 (5.38)	T=-.39, p=0.70	-0.08
Aanzienlijke problemen (N=69)	14.99 (.87)	13.46 (4.83)	T=2.62, p=0.011*	0.44
Ernstige problemen (N=75)	18.12 (1.05)	15.91 (5.31)	T=3.61, p=0.001*	0.58
Zeer ernstige problemen (N=77)	23.97 (2.68)	18.69 (6.47)	T=6.99, p=0.00*	1.07

Noot: Beginklassen gebaseerd op ruwe score. Indeling indicatief op basis van percentielscore voor Amerikaanse Jeugdigen www.sdqinfo.org <11: geen problemen; 12-13=matige problemen; 14-16= aanzienlijke problemen; 17-20 = ernstige problemen; >21= zeer ernstige problemen; SD = Standaarddeviatie; ES = Effect Size.

Bij de twee laagste klassen is sprake van een hogere probleemscore aan het einde, wellicht door een verbeterd diagnostisch beeld of door een zekere mate van onbetrouwbaarheid in de scoring. Bij ernstige problemen bij aanvang is er sprake van een grote tot zeer grote verandering. Ook hier geldt weer: dat kan komen doordat er meer behandel-effect te halen is, en door de statistische kans om bij een hoge score aan het begin een lagere score aan het einde te krijgen. Ook hier laat de groep met hoge scores aan het begin tevens een hoge score aan het einde zien. De verandering in die groep is dus wel groot, maar niet groot genoeg om uit de problemen te zijn.

Er blijkt geen verschil te zijn in de mate van verandering in de totale scores en de verschillende hulpvormen ($F(2, 448)=1.18, p=0.31$).

3.10 NOSI: verandering opvoedingstress volgens moeders

Tabel 18 laat de gemiddelde normatieve deviatiescores zien op de NOSI volgens de moeders bij het begin en het einde van de hulp.

Tabel 18. Gemiddelde normatieve deviatiescores (en SD) op de NOSI bij begin en einde hulp

	Begin	Einde	t-toets	ES
Kinddomein (N=676)	1.51 (1.55)	.73 (1.57)	T=16.22, $p=0.00^*$	0.50
Ouderdomein (N=676)	2.28 (1.37)	1.50 (1.53)	T=16.53, $p=0.00^*$	0.54
Opvoedingsstress totaal (N=676)	2.07 (1.45)	1.22 (1.58)	t=17.71, $p=0.00^*$	0.56

Noot: Normatieve deviatiescore <1.00: geen problemen; 1.00-1.28 = matige problemen; 1.29-1.64 = aanzienlijke problemen; 1.65-1.96 = ernstige problemen; >1.96 = zeer ernstige problemen (Veerman, 2008, p. 313). SD = Standaarddeviatie; ES = Effect Size.

Uit Tabel 18 komt naar voren dat moeders gemiddeld genomen zeer ernstige stress ondervinden in de opvoeding, vooral ten aanzien van hun eigen opvoedingstaak en ook in totaal. Daarbij ervaren ouders aanzienlijke stress ten aanzien van de karaktereigenschappen van hun kind. De stress neemt significant af gedurende de hulp. De effectgrootte wijst op een middelgrote verandering.

Ook voor de NOSI is berekend bij hoeveel ouders er sprake is van 'betrouwbare verandering' in de opvoedingsstress op basis van de Reliable Change Index (RCI). Tabel 19 toont het percentage ouders dat aangeeft dat de opvoedingsstress negatief, positief of niet is veranderd op basis van de RCI.

Tabel 19. Aantal ouders dat aangeeft dat de opvoedingsstress (NOSI) is veranderd

	Negatieve verandering	Geen verandering	Positieve verandering
Kinddomein (N= 555)	16%	26%	58%
Ouderdomein (N=556)	16%	27%	57%
Opvoedingsstress totaal (N= 555)	16%	21%	63%

Als we kijken naar de percentages in Tabel 19, dan is te zien dat meer dan de helft van de ouders een positieve verandering in de stress ervaart. Dat wil zeggen, zij geven aan bij einde hulp minder stress te ervaren. Ongeveer een derde zegt geen verandering in de stress te ervaren.

Tabel 20 op de volgende pagina laat de effectgroottes zien voor de verschillende beginscores. Omdat een relatief groot aantal cases in de categorie 'Zeer ernstige problemen' valt ($N = 362$), is deze verder opgesplitst in een aantal subcategorieën.

Tabel 20. Gemiddelde scores (en SD) Totale problemen NOSI binnen ernstklassen (N=676)

Ernstklasse begin	Begin	Einde	t-toets	ES	
Geen problemen (N=162)	0.12 (0.62)	-0.21 (0.99)	T=4.46; p=0.00	0.40] Gem. 0.84
Matige problemen (N=36)	1.13 (0.78)	0.54 (1.04)	T=3.46; p=0.001	0.64	
Aanzienlijke problemen (N=55)	1.45 (0.09)	0.67 (1.07)	T=5.48; p=0.00	1.03	
Ernstige problemen (N=51)	1.79 (0.09)	1.06 (1.21)	T=3.04; p=0.01	0.85	
Zeer ernstige problemen 1 (N=191)	2.52 (0.31)	1.62 (1.11)	T=11.30; p=0.00	1.10	
Zeer ernstige problemen 2 (N=170)	3.70 (0.51)	2.39 (1.50)	T=11.95; p=0.00	1.17	
Zeer ernstige problemen 3 (N=11)	5.52 (0.37)	3.37 (2.45)	T=3.04; p=0.012	1.23	
Zeer ernstige problemen 4 (N=0)	-	-	-	-	

Noot: Beginklassen gebaseerd op Veerman (2008, p. 313): <1.00: geen problemen; 1.00-1.28=matige problemen; 1.29-1.64 = aanzienlijke problemen; 1.65-1.96 = ernstige problemen; >1.96 = zeer ernstige problemen (hier verder opgesplitst in 1.97-3.00 = Zeer ernstige problemen 1; 3.01-5.00 = Zeer ernstige problemen 2; 5.01-7.00 = Zeer ernstige problemen 3; > 7.00 = Zeer ernstige problemen 4). SD = Standaarddeviatie; ES = Effect Size.

Ook hier zien we dat naarmate de beginscore hoger is, de grootte van de verandering toeneemt. De relatie is echter – net als bij de CBCL – hier minder sterk dan we eerder bij de STEP (en de SDQ) zagen. Dat komt voor een deel door de fijnmazige indeling in de rubrieken ‘Matige problemen’ t/m ‘Ernstige problemen’ (de gemiddelde effect size van deze drie rubrieken is 0.84). In de discussieparagraaf gaan we verder op de betekenis van de cijfers in.

Kijken we naar de groep met zeer ernstige problemen aan het begin, dan zien we net als bij de eerder besproken vragenlijsten dat de eindscores ook hoog zijn, variërend van aanzienlijke tot zeer ernstige problemen. Een grote mate van verandering betekent dus ook hier dat de cliënten nog niet uit de problemen zijn. Integendeel, de problemen zijn – gemiddeld gesproken – nog steeds ernstig.

Er is een verschil in de mate van verandering van opvoedingsstress bij de ouders tussen de verschillende hulpvormen. Dit geldt voor het kinddomein ($F(3, 540)=3.88, p<0.009$), het ouderdomein ($F(3, 540)=2.77, p<0.04$), en de totale opvoedingsstress ($F(3, 540)=3.43, p<0.017$). Uit een post-hoc vergelijking volgens de Scheffé-methode blijkt dat er bij het kinddomein een significant verschil is tussen pedagogische thuishulp en hulp bij een uithuisplaatsing (kinddomein, $p<0.023$; totale stress $p<0.04$). De verschillscore is bij degene die pedagogische thuishulp ontvangen groter in vergelijking met degenen die hulp ontvangen bij een uithuisplaatsing.

3.11 Beëindiging en duur van de hulp

De wijze waarop de afsluiting van de hulp gebeurt kan worden onderscheiden in ‘beëindiging volgens overeenstemming’ en ‘hulp die voortijdig en eenzijdig door de cliënt of hulpverlener wordt afgebroken’. Hulp die voortijdig en eenzijdig door de cliënt wordt afgebroken is een signaal dat het aanbod wellicht niet afdoende op de hulpvraag is ingespeeld. Bij een eenzijdig afbreken van de hulp is er sprake van een beëindiging van de hulpverleningsrelatie op initiatief van de cliënt en/of hulpverlening zonder expliciete, wederzijdse instemming met de beëindiging van de relatie. Als we spreken over ‘voortijdig’, dan wordt verondersteld dat bureau jeugdzorg, de instelling voor J&O en de cliënt een periode hebben afgesproken waarin de hulp zal plaatsvinden. Een voortijdig einde wordt gemarkeerd door een afbreken van de hulp voordat de duur van de hulp – zoals vastgelegd in het laatste indicatiebesluit – is bereikt.

In het bestand is voor 3225 cases de wijze van de beëindiging bekend (55% van het totale bestand van 5906 cases). In Tabel 21 worden de wijze beëindiging van de hulp weergegeven voor de ‘unieke casussen’.

Tabel 21. Wijze beëindiging van de hulp voor de unieke casussen (N=3225)

Reden beëindiging hulp	N	%
Hulp in overeenstemming met BJZ, J&O, en cliënt beëindigd	2844	88%
Hulp eenzijdig beëindigd door cliënt	355	11%
Hulp eenzijdig beëindigd door hulpverlener	15	<1%
Hulp beëindigd door externe omstandigheden*	7	<1 %
Overig	4	<1 %

* Hulp beëindigd door externe omstandigheden kan bijvoorbeeld zijn: zwangerschap, verhuizing cliënt.

De overgrote meerderheid van de cliënten waarvan de wijze van beëindiging bekend is, heeft op de geplande wijze de hulp beëindigd⁹. Slechts bij een klein deel van de cliënten is er sprake van voortijdige en eenzijdige beëindiging van de hulp.

Er is een verschil in de wijze van beëindiging van de hulp bij de cliënten tussen de verschillende soorten zorg ($\chi^2(6)=25.74, p=0.00$). Cliënten die pedagogische thuishulp ontvangen en cliënten die hulp ontvangen in verband met een uithuisplaatsing beëindigen vaker dan de cliënten met ambulante hulp en daghulp de hulp door externe omstandigheden. Cliënten die ambulante hulp ontvangen beëindigen vaker dan cliënten met pedagogische thuishulp, daghulp of residentiële hulp de hulp in overeenstemming. Cliënten met daghulp beëindigen vaker dan cliënten met pedagogische thuishulp, ambulante hulp of hulp in verband met uithuisplaatsing voortijdig de hulp.

In het bestand is voor 3442 cases de duur van de hulp te berekenen (58% van het totale bestand van 5906 cases). In Tabel 22 wordt de gemiddelde duur van de verschillende soorten hulp gepresenteerd. Bij deze analyses is eveneens uitgegaan van de ‘unieke casussen’.

Tabel 22. Gemiddelde duur van de hulp in dagen (N=3442)

	Pedagogische thuishulp (N=250)	Ambulant (N=1970)	Daghulp (N=602)	Residentieel (N=478)	Totaal (N=3442)
Gemiddeld	230.4	239.4	322.3	378.4	270.3
SD	236.26	154.61	194.58	296.51	200.42
Range	14-2150	12-3005	24-2121	29-4859	12-4859

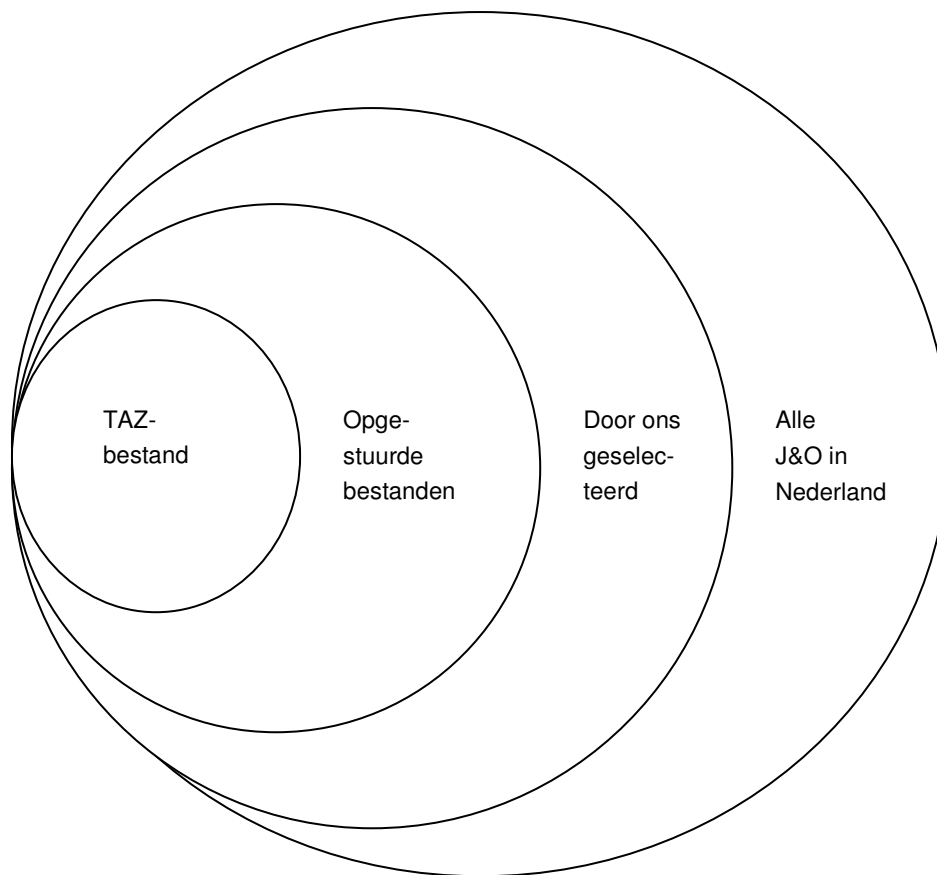
De cliënten ontvangen gemiddeld genomen 270 dagen hulp. Dit betekent dat cliënten ongeveer 8,5 maand hulp ontvangen. Er is een verschil in gemiddelde duur van de hulp en de verschillende hulpvormen ($F(3,3296)=82.64, p<0.00$). Uit een post-hoc vergelijking volgens de Scheffé-methode blijkt dat er een verschil is in duur van de hulp tussen pedagogische thuishulp en daghulp ($p=.00$), pedagogische thuishulp en hulp bij uithuisplaatsing ($p=.00$), ambulante hulp en dagbehandeling

⁹ De wijze van beëindiging is alleen berekend voor die cliënten waarvan voor- en nameting van minimaal één instrument bekend was. Mogelijk kan hierdoor een oververtegenwoordiging van cliënten aanwezig zijn die volgens plan de hulp hebben beëindigd.

($p=.00$), en ambulante hulp en hulp bij uithuisplaatsing ($p=.00$). De gemiddelde duur van de hulp is voor cliënten met residentiële hulp het langst en voor cliënten met pedagogische thuishulp het kortst.

3.12 Discussie

In het vorige hoofdstuk is beschreven dat het TAZ-bestand uit ruwweg vijftig procent van de cases bestaat die werden aangeleverd door de instellingen voor J&O en onderzoekers. Die reductie komt voort uit de selectieve deelname van instellingen en onderzoekers aan het project, moeilijkheden met het koppelen van bestanden, selectie op cliënten waarvan een voor- en een nameting voorhanden is van eenzelfde informant et cetera (zie ook Figuur 1).



Figuur 1. Selectie van het TAZ-bestand

De figuur laat zien dat van de populatie (alle J&O-instellingen in Nederland) er feitelijk in drie stappen selecties zijn gemaakt: de door ons geselecteerde instellingen (1^e selectie), de door de instellingen geselecteerde bestanden (2^e selectie) en de selectie van cases in het uiteindelijke TAZ-bestand op basis van respons voor- en nameting (3^e selectie).

In dit hoofdstuk is getoond dat van de overgebleven cases de leeftijd, sekse en hulpvorm over het algemeen wel bekend is, maar de etnische achtergrond, gezinssituatie, leefsituatie bij aanvang, de dagbesteding en de classificatie van de aard van de problematiek per variabele voor weer ongeveer de helft van de gevallen bekend is. Welke helft dat is, wisselt per variabele. Eerder stelden we al dat het opmerkelijk is dat veel relatief eenvoudig te registreren gegevens, die ook voor de behandeling niet

onbelangrijk zijn, ontbreken. Bovendien is de hoeveelheid cases met bekende kenmerken en voor- en nametingen op instrumenten nogal verschillend. Al met al betekent dit dat het TAZ-bestand niet geschikt is om een goed beeld te geven van *de cliëntèle* en *de effectiviteit* van de jeugd- en opvoedhulp. We hebben immers geen zicht op de selectie-effecten die zijn opgetreden.

Kijken we naar deze selecte TAZ-groep, dan is het volgende te zien. Ten eerste blijkt de overgrote meerderheid van de cliënten in het TAZ-bestand op de geplande wijze de hulp te hebben beëindigd. Ten tweede blijken in het bestand de hulpverleners en de jeugdigen zelf over het algemeen een kleine verandering te rapporteren en de moeders over het algemeen een middelgrote verandering (zie ook Tabel 23).

Tabel 23. Samenvatting gerapporteerde veranderingen problematiek in TAZ-selectie

Instrument	Wat?	Volgens wie?	Hoe groot is de verandering?*	
Geïnternaliseerde problematiek				
jeugdige				
YSR-INT	Geïnternaliseerde problematiek jeugdige	Jeugdige	Klein	(ES=0.30)
CBCL-INT (6-18)	Geïnternaliseerde problematiek jeugdige	Moeders	Klein	(ES=0.49)
		<i>Gemiddeld</i>	<i>Klein</i>	<i>(ES=0.40)</i>
Geëxternaliseerde problematiek				
jeugdige				
YSR-EXT	Geëxternaliseerde problematiek jeugdige	Jeugdige	Klein	(ES=0.24)
CBCL-EXT (6-18)	Geëxternaliseerde problematiek jeugdige	Moeders	Middelgroot	(ES=0.59)
		<i>Gemiddeld</i>	<i>Klein</i>	<i>(ES=0.42)</i>
Totale problematiek jeugdige				
STEP-FJ	Ernst problematiek functioneren jeugdige	Hulpverlener	Klein	(ES=0.38)
YSR-TOT	Totale problematiek jeugdige	Jeugdige	Klein	(ES=0.30)
SDQ –J - TOT	Totale problematiek jeugdige - Jongerenversie	Onbekend ^A	Klein	(ES=0.27)
CBCL-TOT (6-18)	Totale problematiek jeugdige	Moeders	Middelgroot	(ES=0.61)
SDQ- Ou -TOT	Totale problematiek jeugdige – Ouderversie	Onbekend ^B	Klein	(ES=0.39)
SDQ- L - TOT	Totale problematiek jeugdige - leerkrachtversie	Onbekend ^C	Klein	(ES=0.13)
		<i>Gemiddeld</i>	<i>Klein</i>	<i>(ES=0.35)</i>
Totale problematiek (opvoeders)omgeving				
STEP-KO	Ernst problematiek kwaliteit omgeving	Hulpverlener	Klein	(ES=0.27)
NOSI-TOT	Totale opvoedingsstress ouder	Moeders	Middelgroot	(ES=0.56)
		<i>Gemiddeld</i>	<i>Klein</i>	<i>(ES=0.42)</i>
		<i>Gemiddeld</i>	<i>Klein</i>	<i>(ES=0.38)</i>

* Verwaarloosbaar = < 0.20; Klein = ES 0.20 - 0.49; Middelgroot = ES 0.50 - 0.79; Groot = ES >0.80; ^A Vermoedelijk de jeugdige

^B Vermoedelijk een van de ouders of beide ouders; ^C Vermoedelijk de leerkracht.

De effectgroottes die we voor de geselecteerde groep hebben gevonden, blijken redelijk overeen te komen met bevindingen in ander onderzoek als de ouders de informant zijn. Zo rapporteren Veerman e.a. (2005) in een overzichtstudie naar de effectiviteit van Intensieve Pedagogische Thuishulp op de Externaliseren schaal van de CBCL een middelgrote verandering ($ES = 0.52$) en ten aanzien van de opvoedingsstress (gemeten met de NOSI of de Vragenlijst Gezins Problemen / VGP) eveneens een middelgrote verandering ($ES = 0.55$). Daarbij moeten we aantekenen dat we niet kunnen uitsluiten dat cases uit deze studie voor een deel ook in het TAZ-bestand zijn terechtgekomen. In een ouder follow-up onderzoek (Veerman & Ten Brink, 2001) met betrekking tot jeugdigen die Families First, een behandeltraject van een Paedologisch instituut, een Boddaert-opname of een vorm van Intensieve Ambulante Gezinsbehandeling hebben gehad, liet de totaalscore op de CBCL eveneens een middelgrote verandering zien ($ES = 0.73$). De kans is klein dat cases uit die studie ook in het TAZ-bestand terecht zijn gekomen, omdat in het TAZ-project alleen gegevens van de afgelopen tien jaar zijn betrokken. Harder en Knorth (2009) vonden in een meta-analyse naar de uitkomsten van de residentiële jeugdzorg een effectgrootte van .45 voor internaliserend probleemgedrag tot .60 voor externaliserend probleemgedrag en algemeen probleemgedrag. Van der Ploeg en Scholte (2003) vonden in een studie naar de effecten van residentiële behandelprogramma's op de CBCL een effectgrootte die veel lager ligt dan in ons onderzoek: 0.15. De gegevens uit die studie konden niet in het TAZ-bestand worden meegenomen.

Opvallend is ook dat de perceptie van de informant ertoe doet: de ouders (in ons geval de moeders) zien meer veranderingen dan de jeugdigen zelf en de hulpverleners. Ook dat is een bekend fenomeen (zie o.a. Kroes, 2006).

Bij de effectgroottes die we voor de geselecteerde groep hebben berekend zijn we uitgegaan van een strenge variant, zodat we een indicatie krijgen van waar we minimaal van uit kunnen gaan (zie paragraaf 3.3 en bijlage 4). Nadere inspectie van de gegevens leert dat de mate van verandering samenhangt met de ernst van de problematiek bij het begin van de hulp. Hoe hoger de probleemscore aan het begin, hoe groter over het algemeen de verandering. Bij de ernstige problematiek zien we dan ook vaak effectgroottes die groot tot zeer groot zijn. Daar hoort dan wel een aantal kanttekeningen bij.

- Eerder is al genoemd dat er bij ernstiger problemen gewoonweg meer behandel-effect te halen is en dat er een statistisch artefact in het spel is (hoe extremer de beginscore, hoe groter de statistische kans om bij het einde lager te scoren).
- Daarnaast is niet duidelijk waarom er per instrument een aanzienlijk aantal cliënten scoort in de klasse van 'geen problemen'. Het kan zijn dat er op een ander terrein dan het instrument meet een probleem is. Er zijn echter maar weinig casus waarbij dat beeld compleet is. Het kan ook zijn dat in de praktijk de selectie van de cliënten niet goed verloopt. Hoe dan ook is de vraag of deze groep in de berekening van de algemene effect size meegenomen moet worden, omdat deze groep de score op de index 'naar beneden trekt'. Zolang echter niet helder is voor welke probleemgroep (qua ernst en aard) een hulpvorm bedoeld is en wat de doelen waren van de hulp, is het niet gerechtvaardigd om de mate van verandering alleen bij de groep met de hogere probleemscores in het begin van de hulp te berekenen.
- Hetzelfde punt gaat feitelijk op voor de groep die extreem hoog scoort. Bijvoorbeeld, bij de CBCL vonden we 54 cases met een normatieve deviatiescore van meer dan 7,0. Statistisch gezien is dat mogelijk, maar in de praktijk roept een dergelijke score de vraag op wat hier gemeten is. Bijvoorbeeld: is de problematiek werkelijk zo ernstig of is de lijst door de informant gebruikt om een noodkreet te uiten? De vraag is ook hier of dit soort extreme scores meegenomen moeten worden in de berekening van de algemene effect size, nu echter omdat ze

de score op de algemene effect size 'omhoog trekt'. Zonder een heldere onderbouwing van keuzes die daarin te maken zijn, mag deze groep niet zomaar buiten de berekeningen worden gehouden.

- Kijken we wat verder naar de groep die aan het begin een hoge probleemscores heeft, dan valt op dat die aan het einde doorgaans nog steeds – gemiddeld gesproken – een hoge score heeft. Hoewel in deze groep dus grote veranderingen (en dus hoge effect sizes) te observeren zijn, lijken die veranderingen nog niet voldoende om cliënten uit de probleemzone te halen. Dat is overigens een gegeven dat ook uit ander onderzoek naar vormen komt (zie bijvoorbeeld Veerman & Ten Brink, 2001).
- De doelen van de hulp en de hulpvormen zijn in het TAZ-bestand niet helder te identificeren. Dat maakt het moeilijk vast te stellen waarop de hulp was gericht en in welke mate er sprake is van gestapelde hulp. Zou dat wel mogelijk zijn, dan is een betere relatie te onderzoeken tussen de ernst van de problematiek aan het begin van de hulp, de doelen van de hulp, de zwaarte en duur van de hulp en de mate van verandering.

Hoe dan ook leert dit fenomeen ons dat effectgroottes van instellingen of behandelvormen nooit voor zich spreken; ze moeten altijd in relatie worden gebracht met de kenmerken van de doelgroep. Eigenlijk zou er per cliënt een probleemprofiel moeten worden gemaakt over alle relevante terreinen (gedrag jeugdige, vaardigheden en ontwikkeling, lichamelijke problemen, problemen in gezin of opvoeding, problemen in of met de bredere omgeving). Alleen dan wordt zichtbaar waar problematiek aan de orde is. Voorts zouden de gegevens moeten worden gerelateerd aan de doelen van de hulp en de zwaarte en duur van de hulp.

Overigens geldt dat, hoewel we hier steeds spreken van effectgroottes, het in dit onderzoek niet mogelijk is te spreken van de *effecten van de hulp*. Immers, er is geen vergelijkingsgroep die laat zien wat de mate van verandering is zonder jeugd- en opvoedhulp. Er wordt *verandering* gemeten. Waardoor die verandering veroorzaakt wordt (hulp, spontane verbetering, statistische artefacten) blijft onduidelijk. Wil daar meer zicht op komen, dan is het nodig cliënten met thans aangeboden zorg te vergelijken met cliënten die geen zorg krijgen (bijvoorbeeld omdat ze op een wachtlijst staan). Grote veranderingen bij cliënten met bij aanvang ernstige problematiek, kunnen dan in vergelijking met een controlegroep maar voor een klein deel aan de effecten van de behandeling toe te schrijven zijn (of juist niet en blijkt er sprake van een groot effect). Ook is het niet mogelijk uitspraken te doen of de thans aangeboden zorg effectief genoeg is. Om daar iets over te kunnen zeggen, is een vergelijking nodig met cliënten die nieuwe, experimentele behandelvormen krijgen.

Al met al vormt het TAZ-bestand een selectie van cliënten die het moeilijk maakt uitspraken te doen over de mate van verandering bij *de* jeugd- en opvoedhulp. Beperken we ons tot deze selectie, dan zijn ten aanzien van de problematiek van de cliënten in het TAZ-bestand over het algemeen kleine tot middelgrote veranderingen te zien. Bij ernstige beginproblematiek is vaak een grote verandering te noteren. In welke mate die verandering wordt veroorzaakt door de hulp, door spontane verbetering of door statistische artefacten is nog onduidelijk. Voorst blijkt dat ondanks die grote verandering in deze groep de problematiek gemiddeld nog steeds ernstig is.

De gemiddelde duur van de hulp die we in ons onderzoek hebben gevonden, blijkt overeen te komen met bevindingen in branche rapporten (MOgroep Jeugdzorg, 2009). Zo rapporteren zij dat ambulante hulp gemiddelde 8.2 maanden duurde in 2008. Daghulp 10.3 maanden, pleegzorg 21.2 maanden, en residentiële hulp 10.9 maanden duurden.

4. Gebruik TAZ-bestand als bron controlegroepen

4.1 Inleiding

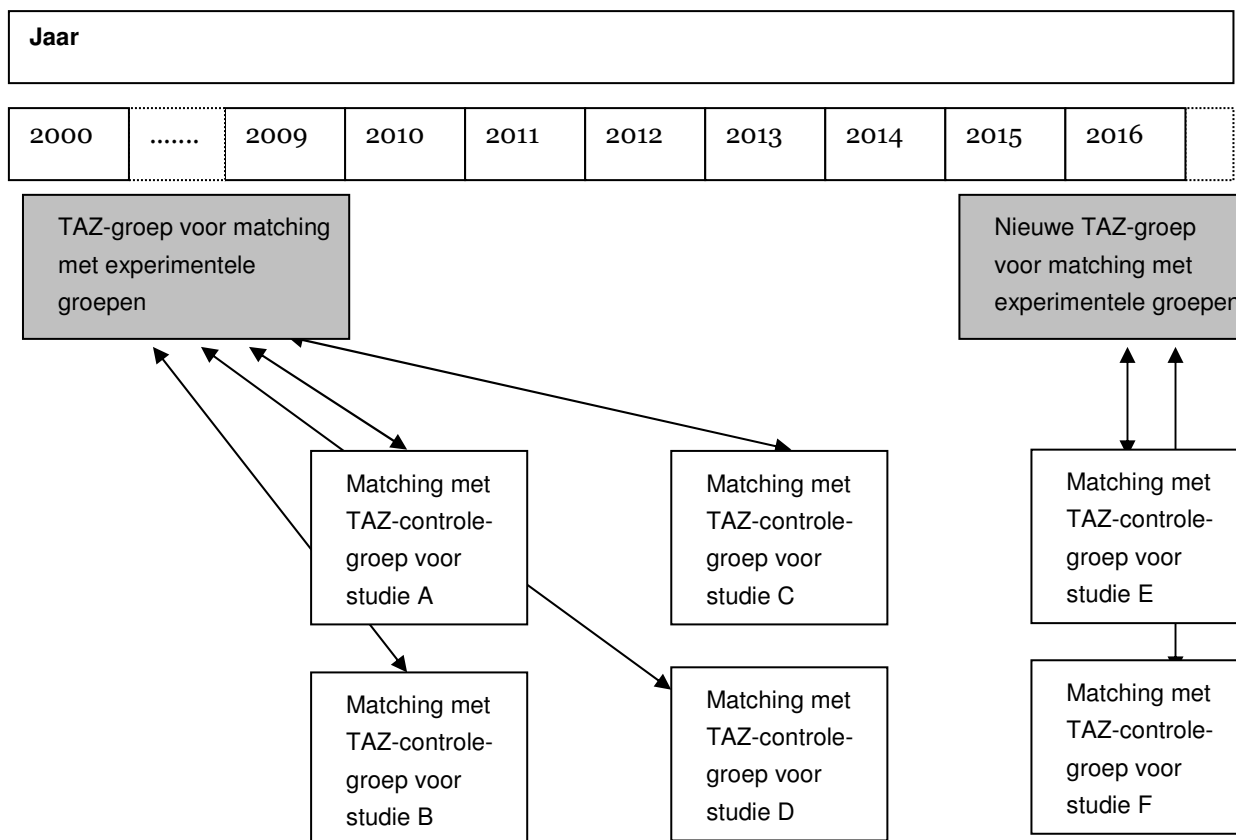
Eén van de doelen van het project is dat het aangelegde bestand een verzameling van demografische en veranderingsgegevens over de alledaagse jeugdzorg bij een groot aantal cliënten in diverse hulpvormen in de sector biedt. Hieruit zijn voor wetenschappelijke onderzoeksdoeleinden controlegroepen samen te stellen, die vergelijkingsmateriaal kunnen bieden bij de effectbepaling van experimentele interventies. In dit hoofdstuk doen we verslag van hoe het opgebouwde TAZ-bestand daarbij te gebruiken is. Eerst bespreken we kort de principes die daarbij gehanteerd worden. Dan volgt aan de hand van een simulatie een illustratie van hoe controlegroepen zich uit het TAZ-bestand laten samenstellen. Tot slot komen de kanttekeningen die bij de getoonde aanpak aan de orde zijn aan bod.

4.2 De principes

Zoals in hoofdstuk 1 is opgemerkt, is het bij wetenschappelijk effectonderzoek gebruikelijk om een experimentele groep te vergelijken met een controlegroep. In de experimentele groep krijgen cliënten de interventie waarvan men het effect wil vaststellen. De controlegroep krijgt dan meestal de in de dagelijkse praktijk uitgevoerde zorg (hier Thans Aangeboden Zorg/ TAZ genoemd). In dit project gaan we uit van een quasi-experimentele onderzoeksopzet. Anders dan bij een RCT vindt er geen random verdeling plaats, maar wordt bijvoorbeeld de nieuwe interventie die in instelling A is ingevoerd, vergeleken met de TAZ van een andere afdeling van instelling A, of met een afdeling van instelling B.

TAZ-controlegroepen zijn op verschillende manieren samen te stellen: gelijktijdig of volgtijdelijk. In een gelijktijdige opzet wordt voor het bepalen van de effectiviteit van een nieuwe interventie in het bestand gezocht naar een vergelijkbare controlegroep die in dezelfde periode de TAZ krijgt. In een volgtijdelijke opzet wordt met cohorten gewerkt. In het ene jaar wordt dan het werken met de controlegroep in beeld gebracht, in een jaar daarop het werken met de experimentele groep. Aan de hand van kenmerken van de cliënten zijn er dan paren samen te stellen (één cliënt uit het eerste cohort, één uit het tweede) die aan het begin van de behandeling maximaal op elkaar lijken voor wat betreft demografische gegevens en aard en ernst van de problematiek. Dit wordt in de literatuur 'Matching through cohort controls' genoemd (Shadish, Cook & Campbell, 2002).

Beide varianten zijn interessant als het gaat om benutting van ons TAZ-bestand als bron waaruit verschillende controlegroepen zijn samen te stellen. Het is immers te beschouwen als een grote verzameling van cliënten waarvan in een bepaald tijdsbestek relevante demografische kenmerken, aard en ernst van de problematiek bij aanvang en de effectvariabelen in beeld zijn gebracht. Heeft men een experimentele groep waarvoor dezelfde gegevens bekend zijn, dan is bij elke casus uit die experimentele groep een zoveel mogelijk gelijkende casus uit het TAZ-bestand te zoeken. Figuur 2 laat zien hoe dit in principe werkt. In studie A wordt bijvoorbeeld gekeken naar de effecten van een experimentele, kortdurende behandeling van kinderen met gedragsproblemen. Studies B t/m D worden in ongeveer dezelfde periode gedaan voor andere experimentele groepen.



Figuur 2. Gebruik van TAZ-groep als bron voor gematchte controlegroepen in specifieke studies

Na een aantal jaren is het TAZ-bestand verouderd, omdat de gebruikelijke zorg is verbeterd door de inbreng van kennis uit de afgeronde studies. Door opnieuw gegevens uit lopende studies en registraties te bundelen ontstaat een geactualiseerd TAZ-bestand, waaruit weer nieuwe controlegroepen zijn te halen (in het voorbeeld: voor studies E en F). De matching in de studies A t/m D is volgtijdelijk (de controlegroepen zijn ten opzichte van de experimentele groepen eerder in de tijd behandeld), de matching in de studies E en F is gelijktijdig. In dit voorbeeld is ervan uit gegaan dat het TAZ-bestand na zo'n jaar of zes ververs is. Als de procedures voldoende zijn gestandaardiseerd en geautomatiseerd kan dit zo vaak geschieden als men wenst. Het streven zou dan kunnen zijn het TAZ-bestand eenmaal per jaar te actualiseren.

Een dergelijke opzet levert onder bepaalde voorwaarden een acceptabel, voorlopig alternatief op voor de samenstelling van een controlegroep via een RCT. Deze voorwaarden zijn dat (1) het aannemelijk moet zijn dat de TAZ-groep voor het aangegeven tijdsbestek een geldige vergelijkingsgroep is en dat (2) de juiste variabelen in beeld zijn gebracht om zowel de matching op relevante kenmerken mogelijk te maken als de effecten van uiteenlopende interventies in verschillende werksoorten in beeld te brengen. Bovendien moet er (3) een goede methode beschikbaar zijn om bij elke casus uit de experimentele groep een zoveel mogelijk gelijkende casus uit het TAZ-bestand te halen. Voor dit laatste doen we in dit hoofdstuk verslag van een simulatie-onderzoek naar de bruikbaarheid van verschillende soorten matchingprocedures. Doel is antwoord te krijgen op twee vragen:

- Welke procedure lijkt het meest hanteerbaar?
- Welke trefkans levert het TAZ-bestand op als geprobeerd wordt een willekeurig samengestelde experimentele groep te matchen met casussen in het TAZ-bestand?

4.3 Opzet van het simulatie-onderzoek

In deze studie zijn drie procedures op bruikbaarheid onderzocht (voor uitgebreide informatie over de procedures zie bijlage 5):

- *Simple Profile Matching (SPM)*. Deze procedure maakt gebruik van een profiel van scores op de matchingsvariabelen dat voor elke casus is geconstrueerd. Vervolgens worden subjecten in groepen met een zelfde profiel ingedeeld. Binnen elke groep worden handmatig en op basis van visuele inspectie casussen uit de experimentele groep aan casussen uit het TAZ-bestand gekoppeld. Lukt dat niet, dan zoekt men koppels met een zo goed mogelijke gelijkens.
- *Cluster Analysis Matching (CAM)*. Deze procedure gaat uit van een berekening van afstand van casussen op basis van de scores op de matchingsvariabelen. Hoe kleiner de afstand tussen twee casussen, hoe meer ze op elkaar lijken.
- *Propensity Score Matching (PSM)*. Deze variant werkt met een index die met behulp van een logistische regressie-analyse voor elke casus wordt berekend: de zogeheten propensity score. De score geeft de waarschijnlijkheid weer waarmee een casus op basis van de matchingsvariabelen te rekenen is tot de experimentele groep of het TAZ-bestand. Hiervoor zijn op internet tegenwoordig goede beschrijvingen voorhanden die aangeven hoe met behulp van SPSS – een bekend statistisch programma dat we in ons project gebruikten – de matching is uit te voeren. Zo is aan PSM een speciale spss-macro geplakt (de zogeheten ‘procedure van Levesque’), die geheel geautomatiseerd bij elke casus uit de experimentele groep een casus uit het TAZ bestand haalt met een zo klein mogelijk verschil in propensity score.

De simulatie vond in twee rondes plaats.

- In de eerste ronde is een beperkt databestand gebruikt met STEP-gegevens van vier instellingen, om de matchingprocedures aan een eerste test te onderwerpen (totale N=2580; het totale bestand was op het moment van deze test nog niet klaar).
- De tweede ronde is uitgevoerd op een van de laatste versies van het totale bestand, met alle cases die een voor- en nameting op de CBCL 6-18 jaar hebben (totale N = 2856).¹⁰

Bij elke ronde is uit het TAZ-bestand eerst willekeurig een groep van 40 casussen gehaald die is bestempeld als ‘de gesimuleerde experimentele groep’. In het gehele CBCL-bestand missen veel achtergrondgegevens. Missende gegevens over de leeftijd bij opname zijn aangevuld met de leeftijd bij de eerste afname van de CBCL (die meestal hetzelfde blijkt te zijn als leeftijd bij opname). Voor alle casussen in de gesimuleerde experimentele CBCL-groep hebben we de etniciteit ingevuld; de matchingprocedure gaat in het TAZ-bestand dan zoveel mogelijk op zoek naar de spaarzame, soortgelijke casussen waarvan gegevens over etniciteit aanwezig zijn.

¹⁰ Het definitieve TAZ-bestand is iets groter dan het hier gebruikte bestand. Door het grotere aantal casussen heeft de matching in het definitieve bestand een grotere trefkans. Voor de hier uitgevoerde simulatie is dat niet erg; het betekent dat de simulatie onder strengere omstandigheden is uitgevoerd dan bij het definitieve bestand.

4.4 Eerste simulatie-onderzoek: matching STEP

4.4.1 Variabelen in de matching

Gesimuleerd is dat de experimentele groep bestond uit een groep gezinnen met kinderen van 6-17 jaar die een vorm van oplossingsgerichte begeleiding kregen aangeboden in combinatie met verschillende hulpvormen. Doel van de begeleiding was uiteindelijk om de ernst van de problematiek van de jeugdige te verminderen. De kenmerken van de gesimuleerde experimentele groep worden in Tabel 24 gegeven. Matching vond plaats op de variabelen:

- Sekse (jongen, meisje);
- Leeftijd bij opname, ingedeeld in de categorieën 0-4 jaar, 5-7 jaar, 8-12 jaar, 13-16 jaar, 17 jaar of meer (geselecteerd op leeftijd 6-17 jaar);
- Etniciteit, ingedeeld in de categorieën wel of niet allochtoon;
- Beginscore op de Functioneren Jeugdige-schaal van de Standaard Taxatie Ernst Problematiek (STEP-FJ, inclusief item duur van de problematiek) met de scores ingedeeld in de categorieën: Goed functioneren (score 6-9), Redelijk goed functioneren (10-14), Matige problemen in functioneren (15-20), Aanzienlijke problemen in functioneren (21-25), Zware tot extreme problemen in functioneren (26-30).

4.4.2 Resultaten Simple Profile Matching (SPM)

Middels deze procedure kon voor alle veertig casussen in de gesimuleerde experimentele groep een match gevonden worden in het TAZ-bestand.

Tabel 24. Resultaten Simple Profile Matching (SPM) in STEP-bestand*

Kenmerk	Gesimuleerde Experimentele groep (N=40)		Controlegroep (N=40; gematched uit TAZ)		Sign.	ES nameting
Aantal jongens	28	(70%)	28	(70%)	n.s.	
Aantal allochtoon	23	(57%)	23	(57%)	n.s.	n.v.t.
Gemiddelde leeftijd in jaren	11.2	(sd=3.1)	10.7	(sd=3.3)	n.s.	
Gemiddelde score voormeting STEP-FJ	17.5	(sd= 4.0)	17.2	(sd= 4.0)	n.s.	
Gemiddelde score nameting STEP-FJ						
Inclusief item duur problematiek	11.8	(sd=4.7)	15.5	(sd=4.7)	.001	0.79
Exclusief item duur problematiek	8.6	(sd=3.7)	11.2	(sd=3.8)	.002	0.69
Gemiddelde score voormeting STEP-KO	12.0	(sd= 4.0)	12.6	(sd= 4.1)	n.s.	
Gemiddelde score nameting STEP-KO						
Inclusief item duur problematiek	8.9	(sd=3.7)	11.9	(sd=4.4)	.002	0.74
Exclusief item duur problematiek	6.3	(sd=3.1)	8.4	(sd=3.2)	.004	0.67

* sd = standaarddeviatie; significantie = significantie verschil experimentele – controlegroep; ns = geen significant verschil; ES = effect size

Tabel 24 geeft een beeld van de kenmerken van de groepen op de achtergrondvariabelen waarop gematcht is, de STEP-FJ-schaal bij aanvang en afloop van de hulp en de scores op de STEP-schaal voor de Kwaliteit van de Omgeving (KO).

Te zien is dat de matchingprocedure heeft geleid tot de samenstelling van een controlegroep die op de matchingsvariabelen niet significant verschilt. Opmerkelijk is dat ook bij de score op een variabele die niet in de procedure is meegenomen (voormeting STEP-KO) er geen significant verschil is te zien. Op de nameting zien we een middelgroot effect op de FJ- en de KO-schaal. Bij de gesimuleerde experimentele groep is de ernst van de problematiek dus in vergelijking met de controlegroep in redelijke mate verminderd. Dat is puur toeval, omdat we daar bij de samenstelling van de groepen niet mee rekening hebben gehouden.

Nadere inspectie van de gegevens laat overigens zien dat de gesimuleerde experimentele groep en controle groep verschillen als het gaat om de aangeboden hulp. Zo bleek in de controlegroep iets minder ambulante hulp geboden (72%) in vergelijking met de gesimuleerde experimentele groep (79%), terwijl er in de gesimuleerde experimentele groep voor 3 casussen gegevens over de geboden hulp ontbraken. Hier is op twee manieren de matching beter te maken. Ten eerste kan het soort geboden hulp in de matchingprocedure worden meegenomen. Ten tweede bleek dat er per experimentele casus meerdere matches in de controlegroep beschikbaar waren; men kan dan op achtergrondvariabelen die oorspronkelijk niet in de procedure zijn meegenomen alsnog met behulp van gegevens over de aangeboden hulp verder *finetunen*.

4.4.3 Resultaten Cluster Analysis Matching (CAM)

De CAM-procedure bleek bij toepassing op het relatief grote bestand niet goed uitvoerbaar. SPSS produceerde geen scorelijsten en figuren die gemakkelijk hanteerbaar waren. Om dit op te lossen, zouden we bijvoorbeeld aparte clusteranalyses per leeftijdsgroep moeten draaien. Aangezien dit veel werk is en er twee andere procedures beschikbaar zijn die zich beter laten hanteren bij grote bestanden, is de CAM per subgroep niet uitgevoerd.

4.4.4 Resultaten Propensity Score Matching (PSM)

Net als bij de eerste procedure kon ook bij de PSM voor alle veertig casussen in de gesimuleerde experimentele groep een match gevonden worden in het TAZ-bestand. Tabel 25 op de volgende pagina geeft weer een beeld van de kenmerken van de groepen op de achtergrondvariabelen waarop gematcht is, de CBCL-schalen bij aanvang en afloop van de hulp en de scores op de STEP-schaal voor de Kwaliteit van de Omgeving (KO).

Wederom heeft de matchingprocedure geleid tot de samenstelling van een controlegroep die op de matchingsvariabelen niet significant verschilt. De nadere inspectie van de gegevens laat overigens ook hier zien dat ten aanzien van de verdeling van de casussen op de verschillende hulpvormen er een (niet-significant) verschil is tussen de twee groepen. Door dezelfde maatregelen te nemen als bij de SPM-methode is beschreven, is hiervoor te corrigeren.

Tabel 25. Resultaten Propensity Score Matching (PSM) in STEP-bestand *

Kenmerk	Gesimuleerde Experimentele groep (N=40)		Controlegroep (N=40; gematched uit TAZ)		Sign.	ES nameting
Aantal jongens	28	(70%)	28	(70%)	n.s.	
Aantal allochtoon	23	(57%)	23	(57%)	n.s.	n.v.t.
Gemiddelde leeftijd in jaren	11.2	(sd=3.1)	12.2	(sd=3.6)	n.s.	
Gemiddelde score voormeting STEP-FJ	17.5	(sd= 4.0)	17.6	(sd= 4.0)	n.s.	
Gemiddelde score nameting STEP-FJ						
Inclusief item duur problematiek	11.8	(sd=4.7)	15.5	(sd=5.0)	.001	0.76
Exclusief item duur problematiek	8.6	(sd=3.7)	11.2	(sd=4.2)	.005	0.66
Gemiddelde score voormeting STEP-KO	12.0	(sd= 4.0)	13.3	(sd= 3.9)	n.s.	
Gemiddelde score nameting STEP-KO						
Inclusief item duur problematiek	8.9	(sd=3.7)	11.9	(sd=4.3)	.002	0.75
Exclusief item duur problematiek	6.3	(sd=3.1)	8.2	(sd=3.2)	.009	0.60

* STEP-FJ = score op STEP Functioneren Jeugdige; STEP-KO = score op STEP Kwaliteit Omgeving; sd = standaarddeviatie; significantie = significantie verschil experimentele – controlegroep; ns = geen significant verschil ; ES=effect size

4.5 Tweede simulatie-onderzoek: matching CBCL

4.5.1 Variabelen in de matching

Gesimuleerd is dat de gesimuleerde experimentele groep bestond uit een groep cases van 4-18 jaar¹¹ die een nieuwe, sterk geprotocolleerde vorm van ambulante ouderbegeleiding specifiek voor de aanpak van gedragsproblemen van de kinderen kregen aangeboden, als alternatief voor de gebruikelijke ambulante hulp. Hier vond matching vond plaats op de variabelen:

- Hulpvorm (alleen ambulante hulp);
- Informant (alleen moeder);
- Sekse (jongen, meisje);
- Leeftijd bij opname, ingedeeld in de categorieën 0-4 jaar, 5-7 jaar, 8-12 jaar, 13-16 jaar, 17 jaar of meer;
- Etniciteit, ingedeeld in de categorieën wel of niet allochtoon;
- Normatieve deviatiescore op de CBCL 6-18 jaar, schaal externaliseren, bij het begin van de behandeling, met de scores ingedeeld in de categorieën: Geen tot matige problemen externaliseren (NDS t/m 1.28), Aanzienlijke problemen externaliseren (1.29 - 1.64), Ernstige problemen externaliseren (1.65 - 1.95), Zeer ernstige problemen externaliseren 1 (1.96 – 3.00), Zeer ernstige problemen externaliseren 2 (3.01 – 5.00), Zeer ernstige problemen

¹¹ De CBCL 6-18 wordt doorgaans gebruikt in de leeftijdsgroep 4-18 jaar.

externaliseren 3 (5.01 – 7.00), Zeer ernstige problemen externaliseren 4 (7.01 of meer). Geselecteerd is op kinderen met een NDS hoger dan 1.28.

4.5.2 Resultaten Simple Profile Matching (SPM)

Omdat in het TAZ-bestand veel casussen geen gegevens over etniciteit hebben, bleek het – zoals verwacht - moeilijk om goede matches te vinden met de gesimuleerde experimentele groep, waar deze gegevens wel compleet zijn. Op basis van de profielen konden koppels worden gezocht met soortgelijke etniciteit, maar met een klein verschil in leeftijdsklasse of beginscore op de CBCL. Soms moest echter een koppel worden gemaakt waarvan de etniciteit in de controlegroep onbekend was, omdat zo op het oog het verschil in leeftijd of beginscore op de CBCL binnen de perken kon worden gehouden. Uiteindelijk kon voor alle veertig casussen in de gesimuleerde experimentele groep een match gevonden worden in het TAZ-bestand. Het resultaat staat in Tabel 26.

Tabel 26. Resultaten Simple Profile Matching (SPM) in CBCL-bestand *

Kenmerk	Gesimuleerde Experimentele groep (N=40)		Controlegroep (N=40; gematched uit TAZ)		Sign.	ES nameting
Aantal jongens	23	(57.5%)	23	(57.5%)	n.s.	
Aantal allochtoon	12	(30.0%)	8	(20.0%)	n.s.	n.v.t.
Aantal etniciteit onbekend	0	(0.0%)	4	(10.0%)		
Gemiddelde leeftijd in jaren	11.4	(sd=3.8)	11.3	(sd=3.8)	n.s.	
Gemiddelde NDS voormeting CBCL-EXT	4.1	(sd= 1.5)	3.9	(sd= 1.3)	n.s.	
Gemiddelde NDS nameting CBCL-EXT	2.4	(sd= 2.2)	2.6	(sd= 1.7)	n.s.	0.10
Gemiddelde NDS voormeting CBCL-INT	2.5	(sd= 1.8)	2.1	(sd= 1.6)	n.s.	
Gemiddelde NDS nameting CBCL-INT	1.1	(sd= 1.6)	1.4	(sd= 1.6)	n.s.	0.19
Gemiddelde NDS voormeting CBCL-TOT	3.7	(sd= 1.6)	3.5	(sd= 1.2)	n.s.	
Gemiddelde NDS nameting CBCL-TOT	2.0	(sd= 2.0)	2.4	(sd= 1.6)	n.s.	0.22

* CBCL-EXT = Normatieve Deviatie Score (NDS) op CBCL schaal Externaliseren; CBCL-INT = Normatieve Deviatie Score (NDS) op CBCL schaal Internaliseren; sd = standaarddeviatie; significantie = significantie verschil experimentele – controlegroep; ns = geen significant verschil ; ES=effect size

Ondanks de verschillen die bij de selectie van de koppels zijn toegestaan, blijkt het verschil op groepsniveau erg mee te vallen. Te zien is dat de matchingprocedure heeft geleid tot de samenstelling van een controlegroep die op de matchingsvariabelen niet significant verschilt. Opmerkelijk is dat ook de score op een variabele die niet in de procedure is meegenomen (voormeting CBCL-INT) er geen significant verschil is te zien. Overigens zijn de effect sizes van de scores bij nameting op de Externaliseren- en de Internaliseren-schaal verwaarloosbaar. De gesimuleerde experimentele groep is er dus niet beter op geworden dan de controlegroep. De effect size voor de totale scores op de CBCL is klein. Dat is niet verbazingwekkend, want beide groepen zijn ten behoeve van de simulatie uit het TAZ-

bestand gehaald. De hoop is natuurlijk dat bij een *echte* experimentele groep er *wél* klinkende effectsizes uitkomen.

4.5.3 Resultaten Cluster Analysis Matching (CAM)

Ook hier blijkt deze procedure op het relatief grote bestand niet goed uitvoerbaar. Twee computers van verschillende netwerken lopen er zelfs op vast. SPSS produceerde geen scorelijsten en figuren die gemakkelijk hanteerbaar waren. De procedure zou wellicht kunnen werken als het bestand in subgroepen wordt opgesplitst, maar gezien het vele werk dat dit kost en de beschikbaarheid van de twee andere, meer hanteerbare procedures, is de CAM per subgroep niet uitgevoerd.

4.5.4 Resultaten Propensity Score Matching (PSM)

Net als bij de eerste procedure kon ook bij de PSM voor alle veertig casussen in de gesimuleerde experimentele groep een match gevonden worden in het TAZ-bestand. Tabel 27 geeft een beeld van de kenmerken van de groepen op de achtergrondvariabelen waarop gematcht is, de CBCL Externaliseren-schaal bij aanvang en afloop van de hulp en de scores op de Internaliseren-schaal.

Tabel 27. Resultaten Propensity Score Matching (SPM) in CBCL-bestand *

Kenmerk	Gesimuleerde Experimentele groep (N=40)		Controlegroep (N=40; gematched uit TAZ)		Sign.	ES nameting
Aantal jongens	23	(57.5%)	23	(57.5%)	n.s.	
Aantal allochtoon	12	(30.0%)	12	(30.0%)	n.s.	n.v.t.
Aantal etniciteit onbekend	0	(0.0%)	0	(0.0%)		
Gemiddelde leeftijd in jaren	11.4	(sd=3.8)	12.1	(sd=4.0)	n.s.	
Gemiddelde NDS voormeting CBCL-EXT	4.1	(sd= 1.5)	4.0	(sd= 1.4)	n.s.	
Gemiddelde NDS nameting CBCL-EXT	2.4	(sd= 2.3)	2.4	(sd= 1.7)	n.s.	0.00
Gemiddelde NDS voormeting CBCL-INT	2.5	(sd= 1.8)	2.0	(sd= 1.5)	n.s.	
Gemiddelde NDS nameting CBCL-INT	1.1	(sd= 1.6)	1.2	(sd= 1.6)	n.s.	0.06
Gemiddelde NDS voormeting CBCL-TOT	3.7	(sd= 1.6)	3.6	(sd= 1.3)	n.s.	
Gemiddelde NDS nameting CBCL-TOT	2.0	(sd= 2.0)	2.2	(sd= 1.5)	n.s.	0.11

* CBCL-EXT = Normatieve Deviatie Score (NDS) op CBCL schaal Externaliseren; CBCL-INT = Normatieve Deviatie Score (NDS) op CBCL schaal Internaliseren; sd = standaarddeviatie; significantie = significantie verschil experimentele – controlegroep; ns = geen significant verschil; ES=effect size

Opvallend is dat de procedure een exacte match heeft gemaakt op etniciteit, door iets meer verschil in de leeftijd toe te laten dan wij bij de SPM-methode geneigd waren te doen. Dit leeftijdsverschil is niet significant, net als overigens de verschillen op de andere variabelen. Het eindresultaat is per saldo

beter dan de SPM-procedure. Dat laat een interessant verschil zien tussen de twee methoden. Bij SPM moeten we zelf de koppels bij elkaar zoeken, hetgeen veel controle op de matching geeft, maar niet altijd tot goede beslissingen leidt. PSM zoekt geheel automatisch de beste koppels, waardoor subjectieve – en kennelijk soms ongunstige - beslissingen zijn uitgesloten. Alle gevonden effect sizes geven overigens aan dat de gesimuleerde experimentele groep ten opzichte van de controlegroep bij de nameting geen verschil in effecten laat zien. Ook hier geldt weer dat dit niet verbazingwekkend is omdat beide groepen uit het TAZ-bestand gehaald; bij een *echte* experimentele groep hoopt men *wél* klinkende effectsizes te zien.

4.6 Discussie

Onze conclusie is dat het TAZ-bestand zich redelijk goed lijkt te lenen voor een matching van experimentele groepen. Qua omvang en samenstelling lijkt het bestand groot en gevarieerd genoeg. Groot nadeel is dat van belangrijke matchingsvariabelen zoals etniciteit en reden beëindiging hulp er veel gegevens missen. Dat beperkt de trefkans bij matchingprocedures sterk, zeker als er wordt voorgeselecteerd op bijvoorbeeld hulpvorm. Bij onze simulatie hebben we de hulpvorm buiten beschouwing gelaten of we hebben ons tot de meest voorkomende soort (ambulante hulp) beperkt. Matching binnen bijvoorbeeld de residentiële hulp zal moeilijker zijn.

Twee van de drie matchingprocedures bleken goed hanteerbaar en leidden tot een 100% matching van de experimentele groep (N=40) aan casussen uit het TAZ-bestand.

- De Statification Profile Matching (SPM) is daarbij eenvoudig te begrijpen en zonder veel high-tech spss-procedures uit te voeren. De methode is bewerkelijk omdat het zoeken van matches gebeurt aan de hand van visuele inspectie van de profielscores. Eventueel is dit door inzet van een speciale spss-procedure (analoog aan de eerder genoemde procedure van Levesque) te automatiseren.
- De Propensity Score Matching (PSM) is geavanceerder en door de beschikbare SPSS-procedures bijna helemaal geautomatiseerd uit te voeren. De procedure kan in principe ook casussen matchen die niet exact dezelfde propensity score hebben. In onze simulatie blijkt dit in feite beter uit te pakken dan de SPM-methode, omdat er geen subjectieve beslissingen meer aan te pas komen.

Hoe dan ook is controle achteraf op de gematchte groepen altijd van belang. Eventuele ongewenste verschillen tussen de groepen op belangrijke kenmerken zijn te verkleinen door – bij meerdere matches per casus in de experimentele groep – de meest gelijkende casus uit het TAZ-bestand te selecteren, door meer variabelen in te matching te betrekken en/of door fijnere categorie-indelingen te kiezen.

Overigens blijft ook bij het gebruik van het TAZ-bestand voor matchingdoeleinden de vraag of de aan de experimentele groep gematchte casussen een representatieve weerspiegeling geeft van de thans aangeboden zorg. In onze simulaties zat wat dat aangaat een systematische vertekening, omdat de experimentele groep uit het totale bestand getrokken werd, om deze vervolgens te matchen met de overige casussen in het bestand. Vertekeningen in het totale bestand zijn daardoor meegenomen in de trekking van de experimentele groep. De echte test zal dan ook uit de praktijk moeten komen.

Bij een matching met een experimentele groep uit de praktijk is het selectieprobleem van het TAZ-bestand overigens wat kleiner. In die context gaat het er immers vooral om dat de experimentele groep een goede weerspiegeling is van de populatie waarover de onderzoeker een uitspraak wil doen. De kenmerken van de controlegroep worden vervolgens via de matching op belangrijke variabelen zoveel mogelijk met de experimentele groep vergelijkbaar gemaakt. Toch is ook dan het selectieprobleem niet helemaal verdwenen, omdat er bij de samenstelling van het TAZ-bestand selecties een rol kunnen hebben gespeeld die niet zichtbaar zijn (zie ook Shadish, Cook & Campbell, 2008). Wij raden dan ook aan om bij een matching met een 'echte' experimentele groep niet te volstaan met één gematchte controlegroep uit het TAZ-bestand, maar zo mogelijk er een aantal controlegroepen uit te trekken. Daarmee scheidt men de mogelijkheid een eventuele vertekening in een gematchte controlegroep te onderkennen. Ook dit sluit het selectieprobleem niet geheel uit, maar verkleint het verder.

Door per experimentele groep meer controlegroepen uit het TAZ-bestand te trekken, voorkomt men ook een afhankelijkheid van toevalstreffers. In onze simulatie met de STEP bleken bijvoorbeeld de experimentele en controlegroep op de eindmeting significant te verschillen. De effect size laat bovendien een middelgroot effect zien, terwijl in werkelijkheid beide groepen uit hetzelfde TAZ-bestand afkomstig waren. Dit betekent dat bij een matching tussen een experimentele groep die niet uit het TAZ-bestand komt en een controle groep die daar wel uit is gehaald een zelfde toevalstreffer mogelijk is. Dit is beter te onderkennen als er meerdere controlegroepen worden gebruikt.

Tot slot geldt dat men in het oog moet houden dat het hier gaat om een quasi-experimentele procedure waarvan de bewijskracht minder sterk is dan bijvoorbeeld een RCT of een gedegen herhaalde case-study (Van Yperen, Bijl & Veerman, 2008). Met andere woorden, wil men meer zekerheid over de vraag of de experimentele interventie werkt, dan zijn andere onderzoeksopzetten nodig. Met deze studie hebben we een aanpak willen onderzoeken die weliswaar minder zekerheid biedt dan een RCT, maar meer informatie biedt dan alleen een voor- en een nameting bij een experimentele groep zonder enig vergelijkingsmateriaal.

5. Conclusies

5.1 Inleiding

Dit project is erop gericht een gegevensbestand bijeen te brengen uit diverse studies en registraties die zicht bieden op de effectiviteit van de Thans Aangeboden Zorg in de provinciaal gefinancierde jeugdzorg (ook wel jeugd- en opvoedhulp genoemd). Dat gebeurt door de opbouw van een bestand met gegevens over de Thans Aangeboden Zorg. Dit bestand noemen we hier het TAZ-bestand. De gegevens zijn afkomstig uit reeds uitgevoerde en lopende studies bij niet-experimentele groepen en uit lopende cliëntregistraties. In het oorspronkelijke projectplan zijn de volgende drie doelen geformuleerd:

- Ten eerste geeft het TAZ-bestand een beeld van de effectiviteit van de dagelijkse praktijk bij een aantal hulpvormen in de jeugd- en opvoedhulp. Het biedt een soort nulpunt voor de te behalen winst met (nieuwe) verbeterlagen in de sector.
- Ten tweede biedt het TAZ-bestand een verzameling van demografische en veranderingsgegevens over een groot aantal cliënten in diverse hulpvormen in de sector. Hieruit zijn voor wetenschappelijke onderzoeksdoeleinden TAZ-controlegroepen samen te stellen om de effectiviteit van experimentele interventies te bepalen.
- Ten derde geeft het TAZ-bestand zicht op ‘gaten’. Het bestand laat de ‘gaten’ zien en toont welke vervolgstudies nodig zijn om deze te dichten.

In dit hoofdstuk staan we stil bij de vraag in hoeverre het in het project opgebouwde bestand aan de doelen beantwoordt.

5.2 Beeld van effectiviteit

Het TAZ-bestand bestaat uit ruwweg vijftig procent van de cases die werden aangeleverd door de instellingen voor J&O en onderzoekers. Die reductie komt voort uit de moeilijkheden met het koppelen van bestanden, selectie op cliënten waarvan een voor- en een nameting voorhanden is van een zelfde informant et cetera. Voorts is in het project gebleken dat van de overgebleven cases de leeftijd, sekse en hulpvorm over het algemeen wel bekend is, maar de etnische achtergrond, gezinssituatie, leefsituatie bij aanvang, de dagbesteding en de classificatie van de aard van de problematiek per variabele voor weer ongeveer de helft van de gevallen bekend is. Welke helft dat is, wisselt per variabele. Bovendien is de hoeveelheid cases waarvan de kenmerken bekend zijn en voor- en nametingen op instrumenten beschikbaar zijn nogal verschillend. Daarnaast is niet bekend of de door de instellingen aangeleverde gegevens representatief zijn voor de betrokken instellingen en of de instellingen die deelnamen aan dit onderzoek representatief zijn voor de Nederlandse Jeugd & Opvoedhulp (J&O-hulp). Dat maakt het TAZ-bestand niet geschikt om een goed beeld te geven van *de* verandering van problematiek bij *de* cliëntèle van de J&O-hulp. We hebben immers geen zicht op de selectie-effecten die zijn opgetreden.

Kijken we naar deze selecte TAZ-groep, dan is het volgende te zien.

- Ten eerste blijkt de overgrote meerderheid van de cliënten in het TAZ-bestand op de geplande wijze de hulp te hebben beëindigd.

- Ten tweede blijken in het bestand de hulpverleners en de jeugdigen zelf over het algemeen een kleine verandering te rapporteren (de effect size ligt tussen de 0.24 en 0.30) en de moeders over het algemeen een (bijna) middelgrote verandering (de effect size ligt tussen de 0.49 en 0.61). Over alle vragenlijsten en informanten gemiddeld is de mate van verandering klein (effect size 0.38). De verandering heeft betrekking op het gedrag en de (opvoeders)omgeving.
- Ten derde blijkt er per instrument een aanzienlijk aantal cliënten in de klasse van ‘geen of matige problemen’ te scoren. Algemeen beeld is dat bij deze ‘normale groep’ er weinig verandering te zien is. Dat is logisch, want er valt ook weinig aan te verbeteren. De vraag is waarom er zoveel casus zijn die aan het begin zo laag scoren op een instrument. Het kan zijn dat er op een ander terrein dan het instrument meet een probleem is, of een andere informant ernstiger problemen ziet. Er zijn echter maar weinig casus waarbij dat beeld compleet is. Het kan ook zijn dat in de praktijk de selectie van de cliënten niet goed verloopt.
- Samenhangend met het vorige punt: het blijkt dat de mate van verandering samenhangt met de ernst van de problematiek bij het begin van de hulp. Hoe hoger de probleemscore aan het begin, hoe groter over het algemeen de verandering. Voor een deel is hier een statistisch artefact in het spel, maar afgezien daarvan roept de gevonden relatie veel vragen op. Hoe dan ook maakt dit duidelijk dat effectgroottes van instellingen of van behandelvormen nooit voor zich spreken; ze moeten altijd in relatie worden gebracht met de kenmerken van de doelgroep. Zoals eerder opgemerkt zou er eigenlijk per cliënt een probleemprofiel moeten worden gemaakt over alle relevante terreinen (gedrag jeugdige, vaardigheden en ontwikkeling, lichamelijke problemen, problemen in gezin of opvoeding, problemen in of met de bredere omgeving). Alleen dan wordt zichtbaar waar problematiek aan de orde is en waar er verandering beoogd is. In dat licht zouden de gegevens ook moeten worden gerelateerd aan de doelen van de hulp en de zwaarte en duur van de hulp.
- Ten vijfde blijkt de groep met ernstige problematiek bij aanvang ook bij de afsluiting gemiddeld nog steeds aanzienlijke tot ernstige problematiek te hebben. De verandering is bij deze groep dus wel groot, maar niet voldoende om uit de ‘gevaarzone’ te komen.

Vergelijken we de gevonden effectgroottes met ander onderzoek waarin over meerdere instellingen of interventies werd gerapporteerd, dan passen deze goed in het algemene beeld dat hieruit oprijst: er worden in de thans aangeboden zorg van J&O-hulp significante veranderingen gevonden, met kleine tot middelgrote effectgroottes (zie de discussie in Hoofdstuk 3 voor een meer gedetailleerde bespreking). Duidelijk is ook dat de perceptie van de informant ertoe doet: ouders zien meer veranderingen dan jeugdigen zelf en hulpverleners. Ook dat is een bekend fenomeen.

Algemeen is van belang in het oog te houden dat in het TAZ-bestand *verandering* wordt gemeten, het gaat om veranderingsonderzoek (Van Yperen & Veerman, 2008b). Waardoor die verandering veroorzaakt wordt (hulp, spontane verbetering, statistische artefacten) blijft onduidelijk. Om oorzaken te kunnen vaststellen is (quasi-) experimenteel onderzoek nodig. In dat licht is het TAZ-bestand als ‘datapool’ interessant.

5.3 Datapool matching met experimentele groepen

Op basis van simulatie-onderzoek blijkt het TAZ-bestand zich redelijk goed te lenen als een datapool om via matching controlegroepen samen te stellen die vergeleken kunnen worden met experimentele groepen. Groot nadeel is dat van belangrijke matchingsvariabelen (zoals etniciteit, reden beëindiging

hulp) er veel gegevens missen. Dat beperkt de trefkans bij matchingprocedures sterk, zeker als er wordt voorgeselecteerd op bijvoorbeeld hulpvorm.

Twee van de drie uitgeteste matchingprocedures bleken goed hanteerbaar en leidden tot een 100% matching van de experimentele groep (N=40) aan casussen uit het TAZ-bestand.

- *Simple Profile Matching (SPM)* is eenvoudig te begrijpen en zonder veel high-tech SPSS-procedures uit te voeren. De methode is bewerkelijk omdat het zoeken van matches gebeurt aan de hand van visuele inspectie van de profielscores. Dat levert veel controle op de matching op, maar leidt tot subjectieve en niet altijd goede beslissingen.
- *Cluster Analysis Matching (CAM)* blijkt moeilijk hanteerbaar. Mogelijk werkt deze procedure wel goed als er binnen kleinere subgroepen wordt gematcht; gezien de beschikbaarheid van andere, goed-hanteerbare procedures is in de studie dit niet verder uitprobeerde.
- *Propensity Score Matching (PSM)* is geavanceerd en door beschikbare SPSS-procedures bijna helemaal geautomatiseerd uit te voeren. In onze simulatie blijkt dit beter uit te pakken dan de SPM-methode, omdat er geen subjectieve beslissingen meer aan te pas komen.

Het probleem van het selectieve karakter van het TAZ-bestand is bij een matching met een experimentele groep uit de praktijk wat kleiner dan bij gebruik van het bestand voor de bepaling van de effectiviteit van de thans aangeboden zorg. Niettemin raden wij aan om bij een matching met een 'echte' experimentele groep niet te volstaan met één gematchte controlegroep uit het TAZ-bestand, maar er een aantal controlegroepen uit te trekken. Daarmee scheidt men de mogelijkheid een eventuele vertekening in een gematchte controlegroep en toevalstreffers in de gemeten verschillen te onderkennen.

5.4 De gaten zichtbaar

Gezien het selectieve karakter van het TAZ-bestand is niet te zeggen welke gaten er in het zicht op de effectiviteit van de jeugd- en opvoedhulp bestaan. Wel hebben we met het project een goede indruk gekregen van de lacunes die er optreden, zodra men resultaatgegevens van instellingen bijeen brengt.

- Instellingen moeten zich committeren aan het getrouw vastleggen van de belangrijkste cliëntkenmerken en kenmerken van de geboden zorg om de beschrijving en vergelijking van cliëntgroepen beter mogelijk te maken. Ook moet dit ervoor zorgen dat de gegevens representatief zijn voor bijvoorbeeld alle afgesloten behandelingen in een bepaalde periode.
- Het verkrijgen van metingen is nog een knelpunt. Het ontbreken van deze metingen maakte een groot aantal cases in het TAZ-bestand onbruikbaar voor verdere analyse. Voor de toekomst ligt voor instelling hier een belangrijk verbeterpunt.
- Nodig is de wijze van gegevensverzameling met betrekking tot belangrijke achtergrondgegevens en de registratie van de geboden zorg vergaand te standaardiseren.
- Voor alle opgenomen jeugdigen zou het probleemprofiel over alle relevante terreinen (gedrag jeugdige, vaardigheden en ontwikkeling, lichamelijke problemen, problemen in gezin of opvoeding, problemen in of met de bredere omgeving) en de doelen van de hulp in beeld moeten zijn, zodat duidelijker is op welke van de terreinen men verandering beoogt.
- De uitwisselbaarheid en koppelmogelijkheden van bestanden en systemen verdient veel aandacht. Hier zou binnen de instellingen al meer rekening mee kunnen worden

gehouden, zodat vergelijking tussen instellingen en het samenstellen van controlegroepen wordt vergemakkelijkt.

- De invoering van het Burger Service Nummer als unieke identifieerder van alle vastgelegde gegevens is een onmisbaar element.
- De wijze van berekening van de effectgrootte (ES) in de situatie dat er ‘slechts’ voor- en nametingen zijn en geen controlegroep verdient aandacht.

Voorts merken we op dat instellingen voor jeugd- en opvoedhulp nog erg huiverig zijn om gegevens voor onderzoek te leveren. We zijn bijzonder trots op de instellingen die in het kader van dit project bereid waren hun bestanden beschikbaar te stellen. Het project heeft aangetoond dat het de moeite waard is om alle gegevens uit de kast te trekken die er zijn, al is het alleen maar om op basis van de verzamelde gegevens de juiste vragen te kunnen stellen of om het gebruik van gematchte controlegroepen beter mogelijk te maken.

Het is ons bekend dat een aantal instellingen voor J&O-hulp al bezig is een aantal van bovengenoemde punten te verbeteren. Deze instellingen beseffen dat de verzameling van effectgegevens en achtergrondgegevens ingebed moet zijn in de werkprocessen. Hierdoor wordt monitoring en controle een vanzelfsprekende zaak die ook onder de verantwoordelijkheid van het management valt. In deze instellingen is ook vaak de benutting van de effectgegevens door inhoudelijk deskundigen (behandelcoördinatoren, gedragswetenschappers) goed op gang gekomen. Hierdoor ontstaat er een intrinsieke motivatie om de gegevensverzameling zo compleet mogelijk te hebben. Het gaat dan immers om het verbeteren en verder professionaliseren van het werk (interne kwaliteitszorg) alsook om het beter verantwoorden van het werk naar overheden en subsidiegevers. De interne kwaliteitszorg en de verantwoordingsplicht kan men zien als twee kanten van de effectiviteitsmedaille. Deze staan soms op gespannen voet met elkaar, maar aan beide valt niet te ontkomen. De uitdaging voor instellingen voor J&O-hulp is dan ook de komende jaren met beide kanten ervaring op te doen en deze ervaringen uit te wisselen. Een belangrijke taak van het Samenwerkingsverband Effectieve Jeugdzorg Nederland (SEJN) is hierbij ondersteuning te geven en een platform voor de uitwisseling van ervaringen te bieden.

5.5 Tot slot

Het project ‘Nulmeting’ was een eerste poging om bestanden bijeen te brengen voor de doelen die in de vorige paragrafen zijn beschreven. Op basis van de wetenschappelijke rapportage die in dit rapport is geboden, zullen er afgeleide publicaties en presentaties worden gemaakt voor praktijkonderzoekers, managers en beleidsmakers. Wat ons betreft blijft het daar niet bij.

In de eerste plaats is het de bedoeling het TAZ-bestand onder te brengen bij een instantie die de gegevens beschikbaar stelt voor onderzoekers. Op verzoek van de aan het project deelnemende instellingen wordt overwogen om het bestand niet in zijn geheel voor onderzoekers beschikbaar te stellen. Geopperd is dat het Nederlands Jeugdinstituut de gegevens beheert en voor de onderzoekers – tegen kostprijs – de matching uitvoert. Daarmee wil men voorkomen dat de instellingen die de gegevens aanleveren het zicht kwijtraken op wat er met de gegevens gebeurt. De exacte opzet van de beschikbaarstelling van de gegevens zal met de opdrachtgever van dit project, de begeleidingscommissie en de deelnemende instellingen verder afgestemd worden.

De in het project opgezette matchingprocedures zullen binnen SEJN-verband op korte termijn beschikbaar komen. Ze laten zich immers goed binnen instellingen gebruiken. Ze bieden de mogelijkheid om ‘interne’ experimentele groepen te vergelijken met gegevens die zijn vastgelegd in ‘interne’ resultaatmonitors. Ze vormen daarmee een belangrijke tussenstap tussen het monitoren van de resultaten van de hulp met ‘slechts’ een voor- en nameting zonder controlegroep en de experimentele studie naar de effectiviteit van nieuwe behandelvormen. De bedoeling is dat de uitvoerders van dit project nader over de matchingprocedures publiceren. Direct daarop volgend komen ze via internet voor een breder publiek beschikbaar.

Tevens is het de bedoeling het TAZ-bestand verder te ontwikkelen. Dat gebeurt voor een belangrijk deel door een combinatie te maken van twee nieuwe ontwikkelingen. De eerste is dat alle aanbieders van jeugd- en opvoedhulp hebben afgesproken gegevens vast te leggen over de resultaten van de hulp. Het Inter Provinciaal Overleg en de Mogroep Jeugdzorg hebben het SEJN opdracht gegeven om op proef de landelijke gegevensverzameling en rapportage te verzorgen (zie Van Yperen & Veerman, 2009). Veel instellingen zullen als aanvulling op deze afspraken meer specifieke informatie vastleggen over de resultaten van de hulp middels instrumenten als de CBCL, SDQ en NOSI. Dat brengt de tweede ontwikkeling in beeld: veel bureaus jeugdzorg en aanbieders van jeugd- en opvoedhulp maken gebruik van BergOp, een programma om vragenlijstgegevens vast te leggen en te verwerken. Dat biedt de mogelijkheid om met de instellingen afspraken te maken over gebruik van de gegevens voor onderzoeksdoeleinden.

Voorts wordt momenteel een studie naar de berekening van de effectgrootte bij een voor- en nameting zonder controlegroep uitgevoerd. De uitkomsten daarvan moeten statistische artefacten beter inculceren en een goed interpretatiekader bieden van de resulterende cijfers. Dat moet de praktijk helpen een antwoord te vinden op wat de geconstateerde veranderingen bij groepen cliënten eigenlijk voorstellen.

Tot slot blijven we instellingen oproepen gegevens voor onderzoek beschikbaar te stellen. Hoewel het onderhavige project eindig is, hopen we nieuwe manieren te vinden om het zicht op de effectiviteit van de jeugdzorg te vergroten. Uitbreiding van het TAZ-bestand is daarbij een van de aandachtspunten.

6. Literatuurverwijzingen

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (nr. 1), p. 155-159.
- Harder, A.T. & Knorth, E.J. (2009). 2.345 Jeugdigen (z)onder dak. Een meta-analyse naar uitkomsten van residentiële jeugdzorg. *Kind en adolescent*, 30, 210-230.
- Kroes, G. (2006). *The perception of child problem behavior. The role of informant personality and context*. Academisch proefschrift. Nijmegen: Radboud Universiteit Nijmegen.
- Lekkerkerker, L., Geurts, E., Yperen, T.A. van & Veerman, J.W. (2010). *Project nulmeting jeugdzorg. Rapportage fase 2*. Utrecht / Nijmegen: Nederlands Jeugdinstituut / Praktikon.
- MOgroep Jeugdzorg (2009). *Brancherapportage Jeugdzorg 2008*. Utrecht: MOgroep.
- Ploeg, J.D. van der & Scholte, E.M. (2003). *Effecten van behandelingsprogramma's voor jeugdigen met ernstige gedragsproblemen in residentiële settings*. Amsterdam: Nippo.
- Reith, W. & Stals, K. (persoonlijke communicatie 2010). Relatie ernstscores begin van de hulp en gerapporteerde effect sizes. Rijswijk/Utrecht: Jeugdformaat / Universiteit Utrecht.
- Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton.
- Veerman, J.W. (2008). Methoden voor het kwantificeren en toetsen van effecten. In: In: T.A. van Yperen & J.W. Veerman (Red.). *Zicht op effectiviteit. Handboek voor praktijkgestuurd effectonderzoek in de jeugdzorg* (p. 307-330). Delft: Eburon.
- Veerman, J.W., & Brink, L.T. ten (2001). Lessen uit follow-up onderzoek. In H. van Leeuwen, W. Slot & M. Uijterwijk (Red.), *Antisociaal gedrag bij jeugdigen. Determinanten en interventies* (pp. 207-224). Lisse: Swets & Zeitlinger.
- Veerman, J.W., Janssens, J.M.A.M., & Delicat, J.W. (2005). Effectiviteit van intensieve pedagogische thuishulp. Een meta-analyse. *Pedagogiek*, 25, 176-196.
- Veerman, J.W. & Yperen, T.A. van (2007). Degrees of freedom and degrees of certainty. A developmental model for the establishment of evidence-based youth care. *Evaluation and Program Planning*, 30 (2), 212-221.
- Weisz, J.R., Donenberg, G.R., Han, S.S. & Kauneckis, D. (1995). Child and adolescent psychotherapy outcomes in experiments versus clinics: why the disparity? *Journal of Abnormal Child Psychology*, 23, 1995.
- Yperen, T.A. van (2005). Hoe effectief is de jeugdzorgpraktijk? *Kind en Adolescent*, 26, p. 255-257.
- Yperen, T.A. van, Bijl, B. & Veerman, J.W. (2008). Opzet van praktijkgestuurd onderzoek. In: T.A. van Yperen & J.W. Veerman (Red.). *Zicht op effectiviteit. Handboek voor praktijkgestuurd effectonderzoek in de jeugdzorg* (p. 209-244). Delft: Eburon.
- Yperen, T.A. van, Schouten, R., Lekkerkerker, L. & Veerman, J.W. (2009). *Project nulmeting jeugdzorg. Rapportage fase 1*. Utrecht / Nijmegen: Nederlands Jeugdinstituut / Praktikon.
- Yperen, T.A. van & Veerman, J.W. (2008a). *Project Nulmeting Jeugdzorg. Thans aangeboden zorg als startpunt kwaliteitsverbetering*. Projectvoorstel Samenwerkingsverband Effectieve Jeugdzorg Nederland / SEJN. Utrecht/Nijmegen: Nederlands Jeugdinstituut / Praktikon.
- Yperen, T.A. van & Veerman, J.W. (2008; Red.). *Zicht op effectiviteit. Handboek voor praktijkgestuurd effectonderzoek in de jeugdzorg*. Delft: Eburon.
- Yperen, T.A. van & Veerman, J.W. (2009). *Rapportage prestatie-indicatoren jeugdzorg. Werken aan toegankelijke cijfers*. Projectvoorstel Samenwerkingsverband Effectieve Jeugdzorg Nederland / SEJN. Utrecht/Nijmegen: Nederlands Jeugdinstituut / Praktikon.

Bijlage 1. Verslag proefbestanden

Doel proefbestanden

Voorafgaand aan de dataverzameling hebben we enkele proefbestanden verzameld en bekeken. Doel hiervan is op kleine schaal na te gaan in hoeverre en op welk niveau harmonisatie van gegevens mogelijk is en welke stappen daarvoor nodig zijn. Op basis van de bevindingen kunnen we iets zeggen over de haalbaarheid en de kwaliteit van het uiteindelijke TAZ-bestand.

Aanpak

Voor de pilot hebben we een aantal instellingen Jeugd& Opvoedhulp en onderzoekers benaderd om mee te doen. Vier databestanden met gegevens over Thans Aangeboden Zorg zijn vervolgens geselecteerd:

1. Bestand Jeugdformaat
2. Bestand STEP-onderzoek van het Nederlands Jeugdinstituut
3. Bestand promotie-onderzoek Esther Geurts bij Stek Jeugdzorg en een vergelijkingsgroep
4. Bestand Tender

De bestanden zijn eerst met een SPSS-procedure zodanig bewerkt, dat de gegevens niet meer herleidbaar zijn naar cliënten (bijv. sekse en leeftijd veranderd). Voor de pilot is dit geen probleem, omdat het ons immers niet om de gegevens zelf gaat, maar om het soort gegevens die de bestanden bevatten en de manier waarop de gegevens in het bestand zitten.

Vervolgens is per bestand bekeken welke variabelen het bevat die we in het TAZ-bestand willen opnemen. Alle variabelen die niet in het TAZ-bestand zullen worden opgenomen, zijn verwijderd. Indien het bestand gegevens bevatte van cliënten die een experimentele interventie hebben ontvangen, zijn deze gegevens verwijderd. Aan het einde van deze bijlage is een overzicht opgenomen van de variabelen die de proefbestanden bevatten.

In de volgende stap zijn per bestand de variabelen bewerkt om, voor zover mogelijk, de variabelen uit de verschillende proefbestanden te harmoniseren. Het gaat hier om het aanpassen van variabelennamen, het hercoderen van variabelen en het construeren van nieuwe variabelen op basis van variabelen die het bestand al bevat. In de laatste stap zijn de verschillende proefbestanden aan elkaar geplakt.

Bevindingen

Kenmerken van cliënten

Alle proefgegevens bevatten een aantal gegevens over achtergrondkenmerken van de jeugdige. De meeste achtergrondvariabelen komen zelfs in alle vier de bestanden voor. Variabelen die niet in elk bestand voorkomen, zijn: gezinssamenstelling, juridisch kader en aard van de problematiek. Niet alleen bevatten de proefbestanden vrijwel alle gegevens over alle geselecteerde achtergrondkenmerken, ook zijn deze variabelen praktisch op dezelfde manier gecodeerd of met een

simpele procedure om te zetten naar dezelfde variabelen. Hierbij moet wel worden opgemerkt dat drie van de vier bestanden voor het registreren van achtergrondkenmerken het formulier van de STEP hebben gebruikt. Het is dus mogelijk dat de harmonisatie voor het TAZ-bestand iets minder makkelijk te bewerkstelligen is. Het zicht krijgen op hoe het bestand is opgebouwd en welke variabelen bruikbaar zijn, zal redelijk wat tijd in beslag nemen. Het vergt daarnaast tijd om alle variabelen van dezelfde namen en codes te voorzien.

Wat betreft ernst van de problematiek en kwaliteit van de omgeving: alle proefbestanden bevatten hier gegevens over. Het proefbestand van Tender bevat geen gegevens over de kwaliteit van de omgeving¹². Drie proefbestanden bevatten gegevens van de STEP. Om de gegevens van de STEP te harmoniseren is alleen het aanpassen van de variabelennamen noodzakelijk. Hetzelfde geldt voor de CBCL-gegevens die in twee proefbestanden zitten. Voor de overige instrumenten die gericht zijn op de ernst van de problematiek en de kwaliteit van de omgeving (SDQ, YSR, NOSI, GKS) geldt dat de gegevens slechts in één van de proefbestanden zijn opgenomen. Deze variabelen hoefden dus voor de pilot niet aangepast te worden.

In het uiteindelijke TAZ-bestand zullen echter niet alleen de gegevens *binnen* één instrument geharmoniseerd moeten worden, maar ook de gegevens *tussen* instrumenten, zodat de gegevens die met verschillende instrumenten verkregen zijn, met elkaar vergeleken kunnen worden. Het vergelijken van gegevens van verschillende instrumenten kan door het berekenen van normatieve deviatiescores. Voorwaarde daarbij is dat het instrument genormeerd is. Voor veel gebruikte instrumenten als de SDQ, CBCL, YSR, TRF en NOSI is dit het geval. Voor de STEP is dit (nog) niet het geval.

Kenmerken uitgevoerde hulp

De meeste proefbestanden bevatten gegevens over verschillende hulpvormen. Eén bestand bevat alleen gegevens over residentiële hulp, de andere bestanden bevatten gegevens over meerdere hulpvormen. In alle bestanden is informatie opgenomen over de hulpvorm, vaak op moduleniveau. Meestal kan uit deze gegevens informatie worden getrokken over de inhoud van de hulp (bijv. ambulante of residentiële) en de zorgaanspraak (conform het advies van Kaiser). Dit vergt wel enig puzzelwerk.

In elk bestand kan de duur van de hulp uit de variabelen worden afgeleid. Intensiteit van de hulp is in geen enkel bestand opgenomen.

Kenmerken van uitvoerende organisatie en hulpverleners

Sommige proefbestanden bevatten gegevens over de hulp die door één instelling is uitgevoerd, andere proefbestanden bevatten gegevens over verschillende hulpvormen uitgevoerd door verschillende instellingen. Alle bestanden bevatten informatie over de instelling en afdeling/unit die de hulp uitvoert. Geen van de bestanden bevat gegevens over kenmerken van de hulpverleners (bijv. jaren werkervaring) en kenmerken van de instelling (bijv. aantal medewerkers in dienst).

Kenmerken van resultaat van de hulp

Voor zover de proefbestanden gegevens bevatten van instrumenten die de ernst van de problematiek en de kwaliteit van de omgeving in kaart brengen, zijn deze gegevens zowel bij aanvang als na afloop gemeten en in de bestanden opgenomen. Uiteraard is er bij sommige cliënten sprake van non-respons. De reden voor beëindiging is in alle bestanden opgenomen. Deze gegevens zijn te harmoniseren door de variabelen te hercoderen. In alle bestanden is onderscheid te maken tussen 'hulp in overeenstemming', 'hulp eenzijdig beëindigd', 'hulp beëindigd extern', 'hulp beëindigd voordat

¹² Voor de pilot is bewust gekozen om vanuit Tender alleen CBCL-gegevens aan te leveren.

hulpverleningsplan tot stand is gekomen' en 'reden onbekend'. Indien de hulp eenzijdig is beëindigd is in veel, maar niet alle gevallen onderscheid te maken tussen 'eenzijdige beëindiging door cliënt' en 'eenzijdige beëindiging door zorgaanbieder'.

Twee bestanden bevatten gegevens over doelrealisatie. Het is echter lastig om deze gegevens te harmoniseren, omdat in het ene bestand drie variabelen zijn opgenomen voor drie einddoelen van de hulpverlening en in het andere bestand van een reeks standaarddoelen is aangegeven of het doel van toepassing is. Harmonisatie van de gegevens is alleen mogelijk als er op basis van de variabelen in het bestand één score wordt geconstrueerd die globaal aangeeft in hoeverre de doelen zijn behaald. Tenslotte bevatten twee bestanden informatie over de cliënttevredenheid. De cliënttevredenheid is op twee verschillende manieren gemeten. Harmonisatie is voor deze bestanden niet mogelijk, omdat er niet, zoals bij doelrealisatie, één rapportcijfer uit af te leiden is.

Betekenis voor TAZ-bestand

Uit de pilot met de vier proefbestanden bleek dat de meeste gegevens die we in het TAZ-bestand willen opnemen in de bestanden voorkomen. Een uitzondering hierop vormen de kenmerken van de uitvoerende organisatie en hulpverleners.

Ook lijkt voor de meeste gegevens harmonisatie mogelijk. Om verschillende instrumenten die hetzelfde concept meten met elkaar te kunnen vergelijken, is het wel van belang dat er een normatieve deviatiescore berekend kan worden. Voor de meeste instrumenten die in het TAZ-bestand opgenomen gaan worden, is dit geen probleem, met uitzondering van de STEP.

Binnen de pilot was harmonisatie voor doelrealisatie en cliënttevredenheid een stuk lastiger. Om gegevens van verschillende bestanden te kunnen vergelijken, zal harmonisatie op een abstract niveau het hoogst haalbare zijn.

Voor het TAZ-bestand betekenen de bevindingen vanuit de pilot grofweg drie dingen:

1. het is mogelijk om de bestanden op veel punten vergelijkbaar met elkaar te maken;
2. harmonisatie van gegevens is vaak mogelijk op een laag abstractieniveau;
3. het harmoniseren van de bestanden vergt veel tijd: niet alleen moeten variabelen gehercodeerd worden en van dezelfde namen worden voorzien (helaas is hier geen eenvoudige procedure voor te schrijven in SPSS omdat elk bestand andere variabelen(namen) hanteert), ook kost het tijd om het bestand te 'ontrafelen' en op basis van bestaande variabelen nieuwe variabelen te construeren zodat gegevens vergelijkbaar worden.

Overzicht variabelen in proefbestanden

1. *Bestand Jeugdformaat (N=2242)*
 - Leeftijd, geslacht, gezinssamenstelling, culturele achtergrond, onderwijs/dagbesteding
 - Aard problematiek aanvang en einde (volgens indeling STEP-formulier)
 - Duur en inhoud hulp
 - STEP voor- en eindmeting
 - SDQ voor- en eindmeting
 - Reden beëindiging
 - Doelrealisatie van 3 einddoelen
 - Cliënttevredenheid

2. *Bestand STEP-onderzoek (N=620)*

- Leeftijd, geslacht, gezinssamenstelling, culturele achtergrond, onderwijs/dagbesteding, juridisch kader
- Aard problematiek aanvang en einde (volgens indeling STEP-formulier)
- Duur en inhoud hulp
- STEP voor- en eindmeting

3. *Bestand promotieonderzoek Esther Geurts (N=72)*

- Leeftijd, geslacht, gezinssamenstelling, culturele achtergrond, onderwijs/dagbesteding, juridisch kader
- Aard problematiek aanvang (volgens indeling STEP-formulier)
- Duur en inhoud hulp
- STEP voor- en eindmeting
- CBCL voor- en eindmeting
- YSR voor- en eindmeting
- NOSIK voor- en eindmeting
- GKS voor- en eindmeting
- Reden beëindiging
- Doelrealisatie voor alle afzonderlijke doelen
- BESTE (cliënttevredenheid)

4. *Bestand Tender (N=55)**

- Leeftijd, geslacht, culturele achtergrond, onderwijs/dagbesteding
 - Duur en inhoud hulp
 - CBCL voor- en eindmeting
- Voor de pilot is gebruik gemaakt van een bestand van Tender met een beperkt aantal variabelen; voor het uiteindelijke TAZ-bestand zullen ook gegevens van andere instrumenten opgenomen worden.

Bijlage 2. Respons

Overzicht deelnemende onderzoekers

Naam	Titel onderzoek	N	Overeenkomst	Data levering
H. Damen, J. Janssen, R. de Meyer, J.W. Veerman	- Meer indien nodig, minder als het kan: Een evaluatieonderzoek naar de Integrale Pleegzorg Gelderland (door Damen, H. & Veerman, J. W., 2005)	88	Opgesteld en retour ontvangen	Data geleverd
Instelling: Praktikon	- Evaluatieonderzoek Innovatieprogramma Jeugdzorg Limburg. Deel I: Achtergronden, opzet en resultaten. (door Damen, H. R., Veerman, J. W., & Janssen, J. , 2002). - Het IOG-FF-bestand evaluatieonderzoek intensieve pedagogische thuishulp, hulp aan huis en families first in de periode 1999-2008. (door Praktikon)	169		
E.M.W. Geurts	Ouders betrekken in de residentiële jeugdzorg. Een onderzoek naar inhoud en uitkomsten van contextgerichte hulpverlening (door Geurts, E.M.W., 2010).	173	Opgesteld en retour ontvangen	Data geleverd
Instelling: STEK, NJI				
A.A.M. Kemper, J. Janssens	Intensieve ambulante gezinsbehandeling: een alternatief voor uithuisplaatsing? (door Janssens, J. M. A. M., & Kemper, A. A. M. , 2003; Kemper, A. ,2004).	59	Nee, wegens kwijtraken van onderzoeksgegevens	X
Instelling: Univ. Nijm.				
E.M. Scholte, J. van der Ploeg	- ADHD-kinderen in ontwikkeling. Diagnostiek en effectieve pedagogische aanpak in de jeugdzorg. (door Ploeg, J. D. van der, & Scholte, E. M. , 2001). - Effecten van behandelingsprogramma's voor jeugdigen met ernstige gedragsproblemen in residentiële settings (door Ploeg, J. D. van der, & Scholte, E. M. , 2003c).	71 115	Nee, delen visie en aanpak project niet	X
Instelling: Univ. Leiden	- Effectieve hulpverlening aan jeugdigen met meervoudige psychosociale problemen. Deel II Effecten van de hulp en werkzame bestanddelen in de hulpverlening (door Scholte, E. M., & Ploeg, J. D. van der, 2002).	137		
J. Strijker, Tj. Zandberg	- Breakdown in foster care (door Strijker, J., & Zandberg, Tj., 2005).	127	Deelname gestopt ivm persoonlijke omstandigheden	X
Instelling: Univ. Groningen				
J. van Zeijl	Externalizing problems in 1- to 3-year-old children. Screening, intervention, and the role of child temperament (door Zeijl, J. van , 2006).	237	Adresgegevens onbekend	X
Instelling: Onbekend				
Van Yperen	Onderzoek evaluatieve waarde STEP, 2009	481	Opgesteld en retour ontvangen	Data geleverd
Instelling: NJI				

Deelnemende instellingen voor Jeugd & Opvoedhulp

Instelling	Overeenkomst	Data levering
De Hoender-loogroep	Opgesteld en retour ontvangen	Data geleverd
Entréa	Opgesteld en retour ontvangen	Data geleverd
Horizon	Wil niet deelnemen, gezien hoeveelheid inf. gebrek aan tijd en mankracht om gegevens te leveren	
Jarabee	Opgesteld en retour ontvangen	Data geleverd
Stichting Jeugdformaat Pactum	Opgesteld en retour ontvangen Wil niet deelnemen, gezien hoeveelheid inf. gebrek aan tijd en mankracht om gegevens te leveren	Data geleverd
Parlan	Opgesteld en retour ontvangen	Data geleverd
Stek Jeugdzorg	Opgesteld en retour ontvangen	Geen deelname wegens gebrek aan tijd en mankracht
Vitree (voorheen SJF & De Reeve)	Opgesteld en retour ontvangen	Data geleverd
Tender	Opgesteld en retour ontvangen	Data geleverd
Yorneo	Opgesteld en retour ontvangen	Data geleverd

Bijlage 3. Inhoud TAZ-bestand

N=5906 (unieke casussen)

Achtergrondgegevens/gegevens hulp:	N
Sekse	5893
Geboortedatum en/of leeftijdopname	5730
Etnische achtergrond	3161
Gezinssituatie	3206
Leefsituatie	3119
Dagbesteding	3192
Juridisch kader	766
Aard van de problematiek aanvang	2723
Hulpvorm	5609
Omschrijving hulp	5540
Datum opname	3731
Datum einde	3485
Duur van de hulp	3442
Reden beëindiging hulp	3225
Informant	2977
Vragenlijstgegevens (voor- en eindmeting) van:	
STEP	2877
SDQ jongere	795
SDQ ouder	1473
SDQ leraar	451
YSR	224
CBCL 1,5-5	93
CBCL 6-18	2438
TRF 6-18	4
NOSI	676

Bijlage 4. Verschillen in berekening effect sizes ES

Effect Size (ES) geeft aan hoe een verschil tussen voor- en nameting is te interpreteren. Er is veel discussie over de te gebruiken formule. Verschillende formules geven uiteenlopende uitkomsten, die elk om een apart interpretatiekader vragen.

Vijf varianten zetten we hier op een rij:

Voor de situatie met een experimentele en een controlegroep

- ES1=effect size met pooled sd in noemer (dit is Cohens d, zie Cohen, 1988, 1992)
- ES2=effect size met sd eerste groep in noemer

Voor de situatie met alleen een voor- en een nameting

- ES3=effect size met wortel 1-r in de noemer (r =correlatie voor- en nameting) en pooled Sd groepen (voor- en nameting) (zie Lipsey, 1990, p. 135)
- ES4=effect size met wortel 1-r in de noemer (r =correlatie voor- en nameting) en Sd eerste groep (voormeting) (zie Lipsey, 1990, p. 135)

De effect size die veel in meta-analyses wordt gebruikt:

- ES5=effect size als de gestandaardiseerde verschillscore tussen de twee metingen.

Het veelgebruikte interpretatiekader van Cohen (1992) gaat uit van ES1 en kent de volgende indeling.

<0.20 = een te verwaarlozen effect;

0.20 - 0.49 = een klein effect;

0.50-0.79 = middelgroot effect

>0.80 = groot effect.

Met name de varianten ES3 t/m ES5 leiden meestal tot een (veel) hogere effect size. Onderstaand voorbeeld toont dat aan de hand van CBCL-scores.

	Begin	Einde	ES1	ES2	ES3	ES4	ES5
Internaliseren	2.10 (1.86)	1.22 (1.71)	0.49	0.47	0.87	0.84	0.62
Externaliseren	3.24 (1.99)	2.08 (1.95)	0.59	0.58	1.04	1.03	0.73
Totale problemen	3.02 (1.87)	1.88 (1.85)	0.61	0.61	1.07	1.06	0.77

Gaat men zonder meer uit van ES3, ES4 of ES5 dan kan dat een inflatoir beeld geven van de mate van verandering, tenzij men het interpretatiekader aanpast. Voorlopig lijkt het beter uit te gaan van ES1, ervan uitgaand dat dit een streng beeld geeft (een effect waar men ten minste van uit mag lijken te gaan).

De discussie over de te gebruiken maat is daarmee nog niet beslecht. Onder meer is het noodzakelijk een bekend statistisch artefact (hoe extremer de beginscore, hoe groter de statistische kans om bij het einde lager te scoren) in te calculeren. Twee auteurs van dit rapport (TvY en JWV) werken daarom in een reeds gestarte vervolgstudie aan een advies hoe de ES het beste te berekenen is.

Bijlage 5. De matchingprocedure

Tom van Yperen, Nederlands Jeugdinstituut / Universiteit Utrecht

Inleiding

Het TAZ-bestand is bedoeld als datapool waaruit via een matchingprocedure een 'Care as usual'-controlegroep is te halen, waarvan de effectgegevens met een experimentele groep worden vergeleken. Shadish, Cook en Campbell (2002) duiden dit aan als een van de quasi-experimentele designs in effectonderzoek.

Het principe van matching gaat uit van het paarsgewijs koppelen van zoveel mogelijk subjecten uit de experimentele groep aan subjecten uit het TAZ-bestand die op belangrijk geachte variabelen zoveel mogelijk gelijk zijn, zoals

- Sekse
- Leeftijd
- Etnische afkomst (bijvoorbeeld wel of niet allochtoon)
- Beginscore op de effectvariabele (bijvoorbeeld de mate van gedragsproblemen).

Via de matching wordt ervoor gezorgd dat voor elke casus in de experimentele groep een casus in de controlegroep wordt gezocht die zoveel mogelijk hetzelfde scoort op de genoemde variabelen. Dat moet ertoe leiden dat uiteindelijk de experimentele groep goed vergelijkbaar is met de 'Care as usual'-controlegroep en verschillen in effectgegevens niet zijn toe te schrijven aan verschillen in de samenstelling van de twee groepen.

Omdat het TAZ-bestand zo groot mogelijk is gemaakt, is de kans groot dat bij elke casus in de experimentele groep een goede match is te vinden voor de samenstelling van de controlegroep. Als er voor een casus geen match is te maken, kan deze in principe niet in de vergelijking worden meegenomen.

In deze bijlage wordt beschreven hoe de matchingprocedure in drie varianten is uit te voeren. De varianten zijn ook te combineren.

Voorwaarden

Ten aanzien van het databestand met gegevens van de experimentele groep geldt dat de variabelen waarop gematched wordt in het databestand dezelfde namen moeten hebben als die in het TAZ-bestand.

Bij de hier beschreven voorbeelden is uitgegaan van SPSS versie 11 of hoger.

Voor alle hieronder beschreven varianten geldt dat de twee databestanden (TAZ en Experimentele groep) worden samengevoegd ('onder elkaar gezet'). Dat gebeurt met het SPSS-commando `ADD FILES`. Voorwaarde is dat de variabelen in de samen te voegen bestanden dezelfde namen hebben. Zorg daarvoor met het `RENAME VARIABLES` commando. De variabele met bijvoorbeeld de naam

'treatment' geeft aan of een casus hoort tot de experimentele groep (code 1) of tot het TAZ-bestand (code 0). Voorts moet elke casus een uniek nummer hebben.

Gebruik niet teveel matchingsvariabelen. Hoe meer variabelen, hoe complexer de procedure en hoe minder de kans op goede matches.

Variant 1. Simple Profile Matching (SPM)

Deze procedure maakt gebruik van een profiel dat is geconstrueerd, waarmee vervolgens subjecten in groepen worden ingedeeld. In de groepen worden koppels gezocht. Het is een gemakkelijke procedure die ook door niet-ingewijden in de statistiek goed uit te voeren is. De volgende stappen zijn daarbij aan de orde.

Stap 1. Zorg ervoor dat de scores op de matchingsvariabelen in categorieën zijn ingedeeld, bijvoorbeeld:

- Sekse: 1 = jongen; 2 = meisje;
- Leeftijd: 1 = 0-4 jaar; 2 = 5-8 jaar; 3 = 9-12 jaar; 4 = 13-16 jaar; 5 = 17-23 jaar;
- Etnische afkomst: 1 = autochtoon; 2 = allochtoon;
- Beginscore CBCL: 1 = normale range; 2 = borderline range; 3 = klinische range.

Als regel geldt: hoe fijner de categorie-indeling, hoe exacter de matching, maar ook hoe minder de trefkans.

Stap 2. Construeer met de matchingsvariabelen één nieuwe variabele die het profiel vormt.

Bijvoorbeeld:

```
COMPUTE profiel = (sekse      *1000)      +
                  (Leeftijd   *100)      +
                  (Afkomst    *10)       +
                  (CBCLbegin  *1).
```

Elke casus krijgt een waarde op de nieuw berekende variabelen die het scoreprofiel aangeeft op de oorspronkelijke variabelen. Sorteert de cases op het profiel (SORT CASES) en laat een lijst opstellen (LIST VARIABLES) van de casussen met relevante variabelen. Bijvoorbeeld:

<u>Casus</u>	<u>sleutel</u>	<u>treatment</u>
12	1243	1
102	1243	0
632	1243	0
24	2411	1
432	2411	0

Stap 3. Zoek bij elke casus uit de experimentele groep een casus uit het TAZ-bestand die een zelfde waarde laat zien op het profiel. Zoek koppels die zo min mogelijk verschillen. Vergroot daarvoor bijvoorbeeld in stap 1 het aantal CBCL-rubrieken of sorteert in stap 2 ook op de ruwe CBCL-score (het profiel als eerste en de CBCLscore als tweede variabele noemen) en neem deze mee in de lijst met getoonde variabelen.

Stap 4. Noteer de casusnummers van de gevonden koppels en selecteer deze casussen uit het grote bestand om er de verdere analyses mee uit te voeren.

Een optie bij deze variant is om in een eerste ronde de ‘indikking’ in categorieën nog niet te maken, zodat met de vervolgstappen 2 en 3 eerst de perfect matchende casussen zijn te identificeren. Zorg er dan wel voor dat er in stap 2 rekening mee wordt gehouden dat bijvoorbeeld leeftijd uit twee getallen bestaat. De variabele ‘seks’ moet dan – om het voorbeeld verder te volgen – met 10000 worden vermenigvuldigd, zodat de score op deze variabele altijd op dezelfde positie in de sleutel komt te staan.

Sommige SPSS-versies hebben de optie om zogeheten ‘duplicate cases’ op te sporen. Dit werkt hetzelfde als hierboven. Voordeel is dat de stappen hier vrijwel zonder handwerk worden uitgevoerd. Kijk voor de beschikbaarheid en meer informatie in de menubalk van SPSS onder Data > Identify Duplicate Cases.

Variant 2. Cluster Analysis Matching (CAM)

Deze procedure is onder meer beschreven en in Nederland toegepast door Scholte (1988). De stappen zijn als volgt.

Stap 1. Deel het bestand op de categorische variabelen in subgroepen. Gebruik daarvoor een soortgelijke procedure als in variant 1. Pas stappen 2 en 3 afzonderlijk toe voor elke subgroep.

Stap 2. Pas per subgroep een clusteranalyse toe via de CLUSTER procedure. Zorg ervoor dat de overige variabelen op eenzelfde wijze zijn geschaald (bijvoorbeeld door leeftijd en beginscore te transformeren naar Z-scores). De clusterprocedure bepaalt op basis van de opgegeven variabelen de nabijheid van elke casus ten opzichte van alle andere casussen; het meest gebruikelijk is de zogeheten euclidische afstand te laten berekenen. Draai de clusterinformatie, een dendrogram en de correspondentie tussen de clusterparen uit. Deze informatie laat zien welke casussen het dichtst bij elkaar staan en daarom het eerste in een cluster worden gezet. Bijvoorbeeld, de clusterinformatie toont het volgende:

Cluster stap	Clustervorming		Coëfficiënt (Euclid)
	Cluster 1	Cluster 2	
1	Casus 004	Casus 008	.020408
2	Casus 001	Casus 004	.072704
3	Casus 010	Casus 011	.122908
4	Casus 002	Casus 012	.184133
5	Casus 003	Casus 009	.246173
6	Casus 006	Casus 010	.386352
7	Casus 001	Casus 013	.469813
Etc.			

Kijkend naar de coëfficiënt, is te zien dat casus 4 en 8 dicht bij elkaar staan. Als dat casussen zijn uit de experimentele en de controlegroep, hebben we een match.

Stap 3. Zoek bij elke casus uit de experimentele groep de casus uit het TAZ-bestand die de kleinste afstand heeft. Deze casussen die als eerste bij elkaar worden gestopt in een cluster lijken over het

algemeen zeer veel op elkaar. Gebruik bij het matchen van de casussen een aantal vuistregels. Bijvoorbeeld: de leeftijd mag niet meer dan 3 jaren verschillen en de CBCL-beginscore (in termen van de niet Normatieve Deviatie Score) niet meer dan 0,5 punten.

Stap 4. Noteer de casusnummers van de gevonden koppels en selecteer deze casussen uit het grote bestand om er de verdere analyses mee uit te voeren.

Variant 3. Propensity Score Matching (PSM)

Deze variant, oorspronkelijk voorgesteld door Rosenbaum & Rubin (1983), werkt met een index die voor elke casus wordt berekend: de zogeheten propensity score. De procedure is de laatste jaren erg populair geworden. Daar zijn verschillende redenen voor. Ten eerste is de procedure vrij trefzeker in het vinden van goede matches. Ten tweede zijn in het zoekproces automatisch zowel categorische variabelen als variabelen met een hoger meetniveau te combineren. Ten derde zijn er uitstekende SPSS-macro's beschikbaar gekomen die koppels van casussen opzoeken aan de hand van indexen die in de procedure berekend worden. Er komt praktisch geen handwerk meer aan te pas. De uitgebreide beschrijving van de stappen en de SPSS-procedure is op een heldere manier beschreven door John Painter (z.d.). Op TAZ-pagina van de SEJN-website is daar aan het einde van 2010 meer informatie over te vinden. Kort weergegeven ziet de procedure er als volgt uit.

Stap 1. Deel het bestand op de categorische variabelen in subgroepen. Gebruik daarvoor een soortgelijke procedure als in variant 1. Pas stappen 2 en 3 afzonderlijk toe voor de verschillende subgroepen. Doe dat overigens alleen als er koppels gemaakt moeten worden die in ieder geval op deze categorische variabelen perfect moeten matchen.

Stap 2. Voer een logistische regressie-analyse uit met de groepsindeling (experimenteel – TAZ) als afhankelijke variabele en de matchingsvariabelen als onafhankelijke variabelen. Bewaar de resultaten volgens de instructies van John Painter. Dit leidt tot een bestand waarin voor alle cases een propensity-score is uitgerekend die aangeeft in welke mate het profiel van een casus op de matchingsvariabele voorspelt of deze tot de experimentele of de controlegroep hoort. Casussen met een zelfde profiel hebben een zelfde score. Bijvoorbeeld

Casus	Groep	sekse	leeftijd	Var1	Var2	Var3	Propensity score
1	Exp.	1	9	1	1	1	.06848541
5	Exp.	1	9	1	1	1	.06848541
8	Exp.	1	7	1	1	1	.07321851
12	Exp.	2	8	0	1	1	.09652939
321	Contr.	1	9	1	1	1	.06848541
465	Contr.	1	9	1	1	1	.06848541
721	Contr.	1	9	1	1	1	.06848541
1242	Contr.	2	8	0	1	1	.09652939
2245	Contr.	2	7	1	1	0	.08234752
Etc.							

Stap 3. Voer de door John Painter gespecificeerde SPSS-macro (de zogeheten ‘procedure van Levesque’) uit om koppels op basis van de propensityscores bij elkaar te zoeken.¹³ De procedure maakt dan automatisch een nieuw databestand, waarbij bijvoorbeeld de volgende koppels zijn gemaakt:

Casus Experimentele groep	Propensity-score	Casus Controlegroep	Propensity-score
1	.06848541	321	.06848541
5	.06848541	465	.06848541
8	.07321851		
12	.09652939	1242	.09652939
Etc.			

Te zien is dat voor casus 1, 5 en 12 matches zijn gevonden, voor 8 niet.

Stap 4. Gebruik de casusnummers van de gevonden koppels om de verdere analyses mee te doen.

Referenties

Painter, J.S. (z.d.). *SPSS Propensity Matching Program Description*. Chapel Hill: Jordan Institute for Families. Gedownload op 3 mei 2010 van <http://ssw.unc.edu/VRC/Lectures/index.htm>

Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, p. 41-55.

Scholte, E.M. (1988). *Jeugd. Politie en hulpverlening. Een onderzoek naar preventieve hulpverlening aan jongeren met psychosociale problemen*. Academisch proefschrift. Leiden: Rijksuniversiteit Leiden.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton.

¹³ Deze procedure is niet alleen toepasbaar bij PSM, maar ook bijvoorbeeld in te zetten om casussen bij elkaar te zoeken bij de eerste procedure (SPM). In het onderhavige project ontbrak echter de tijd om daarvoor de spss-macro aan te passen.

