

8 | 05

Justitiële verkenningen

Beleidsevaluatie

verschijnt 9 maal per jaar jaargang 31 december

Bju | Boom Juridische uitgevers



**Wetenschappelijk Onderzoek-
en Documentatiecentrum**

Colofon

Justitiële verkenningen is een gezamenlijke uitgave van het Wetenschappelijk Onderzoek- en Documentatiecentrum van het Ministerie van Justitie en Boom Juridische uitgevers. Het tijdschrift verschijnt negen keer per jaar.

Redactieraad

drs. A.C. Berghuis
dr. M. Croes
prof. dr. F. L. Leeuw
mr. dr. M. Malsch
mr. drs. M.B. Schuilenburg
mr. drs. P.J.J. van Voorst

Redactie

drs. M.P.C. Scheepmaker
mr. drs. P.B.A. ter Veer
drs. M.A.V. Klein-Meijer

Redactieadres

Ministerie van Justitie, WODC
Redactie Justitiële verkenningen
Postbus 20301
2500 EH 's-Gravenhage
tel.: 070-370 71 47
fax: 070-370 79 48
e-mail: p.ter.veer@minjus.nl

WODC-documentatie

Voor inlichtingen:
Infodesk WODC, tel.: 070-370 65 53
(09.00 uur-13.00 uur). E-mail:
wodc-informatiedesk@minjus.nl
Internet-adres: www.wodc.nl

Abonnementen

Justitiële verkenningen wordt gratis verspreid onder personen en instellingen die beleidsmatig werkzaam zijn ten behoeve van het Ministerie van Justitie. Degenen die in-aanmerking denken te komen voor een gratis abonnement, kunnen zich uitsluitend schriftelijk wenden tot-bovenstaand redactieadres. Andere belangstellenden kunnen zich tegen betaling abonneren. Zij dienen zich te wenden tot Boom Juridische uitgevers.

De abonnementenadministratie voor betalende abonnees wordt verzorgd door Boom Juridische uitgevers, Postbus 85576, 2508 CG Den Haag, tel.: 070-330 70 33, fax: 070-330 70 30, e-mail: info@bju.nl, internet: www.bju.nl. De abonnementsprijs bedraagt € 78,- (inclusief BTW en verzendkosten). Prijs losse aflevering € 15,-.

Abonnementen kunnen op elk gewenst tijdstip ingaan. Valt de aanvang van een abonnement niet samen met het kalenderjaar, dan wordt over het resterende gedeelte van het jaar een evenredig deel van de-abonnementsprijs in rekening gebracht. Het abonnement kan alleen schriftelijk tot uiterlijk 1 december van het lopende abonnementsjaar worden opgezegd. Bij niet-tijdige opzegging wordt het abonnement automatisch voor een jaar verlengd. Gratis abonnementen kunnen desgevraagd te allen tijde beëindigd worden.

Administratie

De abonnementenadministratie wordt verzorgd door Boom Juridische uitgevers, Postbus 85576, 2508 CG Den Haag, tel.: 070-33 070 33, fax: 070-33 070 30, e-mail: info@bju.nl, internet: www.bju.nl

Ontwerp

Tappan, Den Haag

Opmaak

Textcetera, Den Haag

ISSN: 0167-5850

Opname van een artikel in dit tijdschrift betekent niet dat de inhoud ervan het standpunt van de minister van justitie weergeeft.


Nederlands
uitgeversverbond
Groep uitgevers voor
vak en wetenschap

Inhoud

Voorwoord	5
<i>D.P. Farrington en B.C. Welsh</i> Het belang van experimentele evaluaties in de criminologie	11
<i>R. Pawson en C. Klein Haarhuis</i> Evaluatie van complexe programma's; een theoriegestuurde aanpak	42
<i>C. Klein Haarhuis, M. van Ooyen-Houben, E. Kleemans en F. Leeuw</i> Rechtshandhaving geëvalueerd; een synthese van 31 onderzoeken	54
<i>H.D. Webbink</i> Causale effecten van beleid; over gecontroleerde en natuurlijke experimenten	72
<i>R. Bijl</i> Evaluatieonderzoek en het integratiebeleid	85
<i>A.H.M. de Jong en D.L. Kabel</i> De rol van evaluaties in het begrotingsproces	100
Internetsites	110
Summaries	112
Jaarboek	115
De WODC-rapporten	122
Register Justitiële verkenningen 2005	127

Voorwoord

Dit nummer van *Justitiële verkenningen* gaat over beleidsevaluatie. We hebben het dan over een uit de twintigste eeuw stammend fenomeen, maar vanzelfsprekend evalueren mensen al eeuwenlang hun activiteiten. Zo zouden al in 2200 v. Chr. evaluaties van persoonsselectie hebben plaatsgevonden in China (Shadish, Cook e.a., 1995, p. 21). Op zoek naar een effectief middel tegen scheurbuik ontwikkelde de Britse scheepschirurgijn James Lind in 1747 iets wat leek op een zogeheten experimenteel design. Op het schip waar hij diende, selecteerde hij twaalf mannen die aan scheurbuik leden en verdeelde hen in zes paren. Elk paar kreeg dagelijks een andere aanvulling op het basisdieet als mogelijke 'remedie': cider, een elixer, zeewater, een mix van knoflook, mosterd en radijs, azijn, en het laatste paar kreeg dagelijks twee sinaasappelen en een citroen. De twee mannen die het citrusfruit aten, herstelden opmerkelijk snel (Lind, 1753). Hoewel er naar huidige maatstaven wel wat aan te merken valt op zijn onderzoeksopzet, had Lind zo de superioriteit van citrusvruchten als remedie aangetoond.

De aanleiding voor dit themanummer is niet alleen de explosieve toename van het aantal evaluaties in Nederland en andere landen. Ook de recente discussies over de methoden van evaluatieonderzoek en het maken van syntheses van bestaande evaluaties bieden meer dan genoeg stof. Een interessante ontwikkeling is de opleving van het experiment als evaluatiemethode. Amerikaanse sociaal wetenschappers werkten al experimenteel in de jaren twintig en dertig, de Britten deden er in de jaren vijftig ervaring mee op. De samenleving mag dan geen laboratorium zijn, het voorbeeld van de medische wetenschap met haar uitgebreide gecontroleerde testen van medicijnen en andere behandelmethoden is toch voor menig sociaal wetenschapper een inspiratiebron, zoals bijvoorbeeld blijkt uit het inmiddels klassieke werk van Donald Campbell en Julian Stanley, *Experimental and Quasi-Experimental Designs for Research* (1966). Dat het experiment als onderzoeksmethode weer terug is van weggeweest, werd afgelopen november duidelijk op de conferentie '(Quasi) Experimental Evaluations and Dutch Society; trends and developments in criminal justice and policies, social integration and education'. De artikelen in dit nummer zijn bijna allemaal

gebaseerd op presentaties die daar zijn gehouden. De conferentie was georganiseerd door het WODC in samenwerking met het ministerie van Financiën en de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Een onderwerp dat tijdens de discussies op de conferentie wél aan de orde kwam, maar in deze aflevering van JV nauwelijks, is het tegenstrijdige fenomeen dat evaluaties een bescheiden rol spelen in het beleidsproces. Tegenstrijdig, omdat er wél keer op keer opdrachten voor evaluatie worden verstrekt en er dus veel geld aan wordt besteed. Afgezien van de soms gebrekkige kwaliteit en ongelukkige timing van evaluaties speelt hier iets heel anders: de kloof tussen onderzoekers enerzijds en beleidsambtenaren en politici anderzijds, botsende denkwerelden, logica's, belangen en loyaliteiten. Misschien komen we daar nog eens op terug in een ander thema-nummer, want het is een onderwerp op zich.

In Nederland begon de evaluatie van overheidsbeleid rond 1980. Sindsdien is het aantal evaluaties gestaag toegenomen. Op het terrein van rechtshandhaving bijvoorbeeld werden in 1995 door het ministerie van Justitie achttien evaluaties uitgevoerd. In 2005 zijn dat er 37.

Evaluatie-‘societies’ van onderzoekers en beleidsmakers kwamen aanvankelijk alleen voor in de VS en Canada, maar figureren nu in vrijwel elk Europees land en in verscheidene Afrikaanse en ‘Australaziatische’ landen. Ook supranationale instituties zoals de Europese Unie en de Wereldbank evalueren hun beleid. Alleen al op het terrein van ontwikkelingshulp lopen de aantallen evaluaties in de duizenden binnen een tijdsbestek van enkele jaren.

Een mogelijke verklaring voor deze groei is dat overheden, maar ook bedrijven en maatschappelijke organisaties tegenwoordig transparant moeten zijn en verantwoording moeten kunnen afleggen over hun beleid en de wijze waarop geld wordt besteed. In de jaren negentig van de vorige eeuw was dit idee sterk gelieerd aan de New Public Management-filosofie die populair was in veel westerse samenlevingen.

Deze ontwikkelingen ontmoeten ook veel kritiek. Er wordt wel gesproken van een ‘evaluatie-industrie’, een ‘audit society’ of ‘controlemaatschappij’. Meer evaluatie, monitoring en controle leiden niet automatisch tot succesvoller beleid. Het kan zelfs de efficiency en effectiviteit van beleid en organisaties schaden, omdat er teveel

op output-indicatoren wordt gestuurd. Dit wordt ook wel de 'performance paradox' genoemd (Van Thiel en Leeuw, 2003).

Een ander punt van kritiek is dat de methodische kwaliteit van veel evaluaties tekortschiet. Zo wordt – althans in Nederland – zelden een experimenteel onderzoeksdesign bij beleidsevaluaties gebruikt. Op basis van zo'n onderzoeksdesign kan worden vastgesteld of waargenomen 'effecten' ook daadwerkelijk kunnen worden toegeschreven aan beleid, en niet aan bijkomstige factoren. Daarnaast is er bij evaluaties nauwelijks aandacht voor de wijze waarop beleid de gevonden 'effecten' veroorzaakt: de 'theorie' over hoe beleid de gewenste effecten zou moeten sorteren.

In het antwoord op deze methodologische problemen zijn twee stromingen te onderscheiden. De 'experimentele' benadering legt de nadruk op onderzoeksdesigns die het mogelijk maken met een grote mate van zekerheid te bepalen of beleid effect sorteert en of geconstateerde veranderingen aan het beleid kunnen worden toegeschreven. De belangrijkste representant van deze stroming is de Campbell Collaboration. Deze internationale non-profitorganisatie van sociaal wetenschappers, statistici en beleidsmakers beheert een database van (experimentele) onderzoeksresultaten over de effectiviteit van uiteenlopende sociale interventies. De benadering van de 'theoriegestuurde' evaluaties legt de nadruk op de vooronderstellingen die ten grondslag liggen aan de (beleids)programma's. Beleid en beleidsprogramma's zijn theorieën: stelsels van aannames en feiten op basis waarvan beleid wordt geacht te werken. Evaluatoren dienen deze theorieën te reconstrueren zodat inzicht wordt verkregen in de onderliggende mechanismen van een interventie. Dit kan aan de hand van beleidsdocumenten, uitspraken, doel- en missierapporten, debatten, diepte-interviews met beleidsmakers enzovoort. De laatste tijd groeit het inzicht dat deze stromingen elkaar eerder aanvullen dan uitsluiten. Het ontwikkelen van een aanpak waarin zowel de óf-vraag als de hóe-vraag wordt betrokken, kan worden gezien als een mooie uitdaging voor (beleids)evaluatieonderzoekers. In het openingsartikel geven de 'experimentalisten' Farrington en Welsh een overzicht van experimentele evaluatieonderzoeken op het terrein van de criminologie in de periodes 1957-1981 en van 1982 tot 2004. Vooral de onderzoeken van later datum wijzen uit dat sommige interventies effectief zijn en andere duidelijk niet. De meeste experimenten vonden plaats in de Verenigde Staten. De auteurs

hopen dat dit type onderzoek ook in andere landen, waaronder Nederland, opgang zal maken.

Pawson en Klein Haarhuis beschrijven vervolgens de theoriegestuurde aanpak van evaluatieonderzoek. Zij laten zien hoe de theorie van een programma kan worden achterhaald. Zo wordt duidelijk of de veronderstelde onderliggende mechanismen die het beoogde effect moeten bewerkstelligen, daadwerkelijk optreden. Zij illustreren het nut van deze aanpak met een aansprekend voorbeeld: de evaluatie van 'Megan's law', een regeling in de VS waarbij buurten worden ingelicht als een uit de gevangenis of inrichting ontslagen seksueel delinquent zich daar vestigt.

Experimentele evaluaties komen in Nederland nog nauwelijks voor. Hoewel er methodologisch het nodige is aan te merken op bestaande evaluaties, kunnen deze toch bruikbaar zijn, zo betogen Klein Haarhuis, Van Ooyen-Houben, Kleemans en Leeuw in hun bijdrage. Hierin wordt een synthese van 31 onderzoeken op het terrein van rechtshandhaving opgesteld. Van belang is dat onderzoeken een gedegen beschrijving bevatten van de onderzoeksopzet, de meetmethode en de eventuele selectiemethode. Ook het gebruik van meerdere informatiebronnen draagt bij aan de bruikbaarheid. Voorts moet duidelijk zijn of beleid daadwerkelijk is geïmplementeerd en moet aandacht zijn besteed aan de vooronderstellingen die ten grondslag liggen aan de interventie. De inhoudelijke synthese wijst onder meer uit dat projecten in de rechtshandhaving vaak bestaan uit verhoogde input van mankracht, expertise en middelen, en uit intensivering van samenwerking. Daarnaast noemen de auteurs een aantal systematische knelpunten.

Webbink borduurt in zijn bijdrage voort op de moeizame relatie tussen experiment en sociale wetenschappen. Het gerandomiseerd toewijzen aan een interventie- of een controlegroep wordt vaak bemoeilijkt door ongewilde doch onvermijdelijke selectieprocessen. Verschillende recente onderzoeken maken tegenwoordig gebruik van natuurlijke experimenten om de causale effecten van beleidsinterventies te bepalen. Het gaat dan om situaties die toevallig zijn ontstaan door natuurlijke processen of door overheidsbeleid en die lijken op een omgeving waarbinnen een gecontroleerd experiment zou kunnen worden uitgevoerd. De auteur geeft verscheidene voorbeelden.

Bijl bespreekt hoe evaluatieonderzoek een rol zou kunnen spelen in het integratiebeleid. Een complicerende factor daarin is dat meerdere en naar hun aard verschillende doelstellingen worden nagestreefd. Besproken wordt wat men verstaat onder het begrip integratie alsmede hoe het beleid op zijn kwaliteiten zou kunnen worden beoordeeld. Positief is de grotere beschikbaarheid van kwantitatieve gegevens en de ruimere mogelijkheden om databestanden van uiteenlopende organisaties aan elkaar te koppelen, meent de auteur. Dat biedt nieuwe mogelijkheden voor goede beleidsevaluaties. De Jong en Kabel, ten slotte, stellen dat evaluatie onontbeerlijk is als men antwoord wil geven op de vraag of de belastingbetaler waar krijgt voor zijn geld. Het aantal evaluaties in Nederland is explosief gestegen, maar de kwaliteit ervan is niet navenant toegenomen. Een verbetering kan worden bereikt door meer effectevaluaties, duidelijker beleidsdoelstellingen, meer garanties voor de onafhankelijkheid van evaluaties en betere verslaglegging. Hoewel consensus lijkt te bestaan over de potentieel grote waarde van evaluaties, is de invloed ervan op het beleidsproces doorgaans marginaal. De auteurs, werkzaam bij het ministerie van Financiën, kondigen aan dat een strategie wordt ontwikkeld ter bevordering van ex ante beleidsevaluaties en effectevaluaties. Ook zou de focus van evaluatieonderzoek moeten verschuiven van instrumenten naar concrete beleidsterreinen.

C. Klein Haarhuis
F.L. Leeuw
M.P.C. Scheepmaker

Literatuur

Campbell, D.T., J. Stanley

Experimental and quasi-experimental designs for research
Chicago, Rand McNally, 1966

Lind, J.

A treatise of the scurvy in three parts; containing an inquiry into the nature, causes and cure of that disease, together with a critical and chronological view of what has been published on the subject

London, A. Millar, 1753

Shadish, W.R., T.D. Cook e.a.

Foundations of program evaluation, theories of practice
London, Sage, 1995

Thiel, S. van, F.L. Leeuw

The performance paradox in the public sector

Public performance & management review, 25e jrg., nr. 3, 2002, p. 267-281

Het belang van experimentele evaluaties in de criminologie

*D.P. Farrington en B.C. Welsh**

De term 'experiment' wordt vaak gebruikt als algemene verwijzing naar een sociale activiteit (met name innovatieve activiteiten) waarvan het uiteindelijke effect onzeker is. Volgens deze definitie zijn de methodes voor de aanpak van misdaad dus meestal experimenteel van aard. In dit artikel heeft het woord 'experiment' echter een enger gedefinieerde en meer technische betekenis. Het woord verwijst naar een systematische aanpak bij het onderzoek naar het effect die prospectieve geplande variaties in één factor (de onafhankelijke variabele) op een tweede factor (de afhankelijke variabele) hebben.

De aard van experimenten kan het beste worden uitgelegd aan de hand van een specifiek voorbeeld. Sherman en Berk (1984) wilden voor het terugdringen van recidive de relatieve effectiviteit onderzoeken van drie door de politie gehanteerde methodes bij de aanpak van huiselijk geweld: arrestatie van de dader, separatie van de dader en het slachtoffer, en bemiddeling tussen de dader en het slachtoffer. Alle drie de methodes waren in gebruik ten tijde van het onderzoek (in 1981-1982). Het bleek echter onmogelijk om door het bestuderen van politiestatistieken voor recidive precies na te gaan welke methode het meest effectief was in het terugdringen van recidive omdat de daders die voor een bepaalde methode waren geselecteerd op allerlei punten verschilden van de daders die na de selectie via een andere methode werden aangepakt. Het bleek onmogelijk om een onderscheid te maken tussen de invloed van de politieaanpak en de invloed van bestaande verschillen tussen de daders.

Met de hulp van de burgemeester en de hoofdcommissaris van politie van Minneapolis voerden Sherman en Berk een experiment uit waarin de daders van huiselijk geweld willekeurig werden toegewezen aan één van de volgende condities: arrestatie, separatie of

* Prof. David Farrington is hoogleraar aan het Institute of Criminology, University of Cambridge. Dr. Brandon Welsh is als assistent-hoogleraar verbonden aan het Department of Criminal Justice, University of Massachusetts-Lowell.

bemiddeling. De beste manier om te zorgen dat de groep mensen die een behandeling ondergaan vergelijkbaar is met de groep die een andere behandeling krijgt, is door ze willekeurig aan de verschillende behandelingen toe te wijzen. Worden er voldoende mensen willekeurig toegewezen, dan zijn, binnen de beperkingen van statistische schommelingen, degenen die de ene behandeling ontvangen gelijk aan degenen die de andere behandeling ontvangen. Het is dan mogelijk om de effecten van de behandelingen los te zien van de effecten van externe variabelen (ongecontroleerde verschillen tussen de groepen). Dit is de reden dat medische behandelingen standaard worden geëvalueerd middels gerandomiseerde en gecontroleerde steekproeven. Sherman en Berk constateerden in de follow-up studie van zes maanden dat de gearresteerde daders aanzienlijk minder kans liepen om te recidiveren (56%) dan degenen die de separatie of bemiddeling was aangeboden. Dit duidde erop dat arrestatie de meest effectieve manier was om dergelijke zaken aan te pakken.

In dit voorbeeld was de onafhankelijke variabele de manier waarop daders werden aangepakt, en de grote afhankelijke variabele de recidive (gemeten aan de hand van strafbladen en interviews met slachtoffers). Dankzij de willekeurige toewijzing waren de daders in de drie condities vergelijkbaar bij eventueel aan de recidive gerelateerde externe factoren (zoals leeftijd, ras, strafblad, werkloosheid) die in de ongecontroleerde (natuurlijke) situatie als uitgangspunt genomen hadden kunnen worden in de aanpak van de daders. De verschillen die vervolgens werden aangetroffen, waren dus het gevolg van de verschillende methodes die door de politie werden gebruikt in haar aanpak van de daders (zie ook Sherman, 1992, hoofdstuk 3).

Een experiment wordt gekenmerkt door de controle van de onafhankelijke variabele. Sherman en Berk wezen alle daders aan één van de drie condities toe. De onderzoekers in het experiment controleerden dus welke dader aan welke conditie werd toegewezen. In dit artikel behandelen wij gerandomiseerde experimenten. In het Minneapolis-experiment werden de daders toegewezen aan de condities via verwijzing naar een tabel willekeurige getallen. De externe variabelen waren dus gecontroleerd (dus onder alle condities gelijkwaardig). In tegenstelling tot matching of statistische controle van variabelen, controleert randomisatie voor zowel onbekende en ongemeten variabelen, als voor bekende en gemeten variabelen (Weisburd, Lum e.a., 2001). In gerandomiseerde experimenten is er

controle van zowel de onafhankelijke als de externe variabelen. De controle van externe variabelen door randomisatie is tot op zekere hoogte te vergelijken met de controle van externe variabelen in de natuurwetenschappen waar fysieke condities constant gehouden worden (bijvoorbeeld temperatuur, druk). Door voor de behandeling te randomiseren, zorgt men ervoor dat de gemiddelde eenheid in de ene behandelgroep ongeveer hetzelfde is als de gemiddelde eenheid in een andere groep. En waar het constant houden van de fysieke condities een grotere gelijkwaardigheid tussen de experimentele units garandeert, is de achterliggende theorie voor experimentele controle in deze gevallen gelijk.

Voordelen van gerandomiseerde experimenten

Experimenten worden opgezet om de causale hypothese over het effect van variaties in de ene variabele op variaties in een andere variabele te testen. Een hypothese kan alleen experimenteel worden getest als deze in dergelijke termen kan worden geformuleerd. In het klassieke model van gemeenschappelijke vooruitgang volgt er op iedere theorie een serie testbare causale hypothesen. Als iedere hypothese in een experiment wordt getest, kan het resultatenpatroon worden vergeleken met het patroon van theoretische voorspellingen. Op basis van deze vergelijkingen, en met inachtneming van andere overwegingen, zoals de complexiteit van een theorie, moet het mogelijk zijn om een conclusie te trekken waarbij de ene theorie boven de andere wordt verkozen. Idealiter dient ieder experiment een schakel te vormen in een keten van cumulatieve kennis, aangestuurd door de theorie. In de praktijk zijn goed ontwikkelde, expliciet gespecificeerde en falsifieerbare theorieën in de criminologie echter zeldzaam, en staan de hypothesen die middels experimenten worden getest meestal op zich, en zijn het geen systematische toetsen van een bredere theorie. (Voor systematische vergelijkingen van ontwikkelings- en levenslooptheorieën bij delictpleging, zie Farrington, 2005, 2006.)

Volgens Campbell en Stanley (1966), Cook en Campbell (1979), en Shadish, Cook e.a. (2002) kan de methodologische geschiktheid van een causalehypotesetest op de volgende vier belangrijke criteria worden getoetst:

- statistische conclusievaliditeit is het bepalen of er inderdaad een verband bestaat tussen de twee variabelen in kwestie (bijvoorbeeld politieaanpak en recidive);
- interne validiteit bepaalt of een verandering in de ene variabele daadwerkelijk tot een verandering in de andere variabele leidt;
- constructvaliditeit bepaalt wat de theoretische constructen zijn waarop de gemeten variabelen zijn gebaseerd; en
- externe validiteit bepaalt in hoeverre de resultaten gegeneraliseerd kunnen worden naar verschillende personen, instellingen en tijden.

De unieke voordelen van gerandomiseerde experimenten ten opzicht van andere methodes zijn de grote interne validiteit en het feit dat er geen onduidelijkheid bestaat over de richting van de causale invloed. In gerandomiseerde experimenten worden veel van de risico's voor de interne validiteit die in niet-experimenteel onderzoek grote gevolgen kunnen hebben, geëlimineerd.

De belangrijkste risico's voor de interne validiteit zijn bekend (zie bijvoorbeeld Farrington, 2003c):

- Selectie: het effect is een weerslag van reeds bestaande verschillen tussen experimentele en controlecondities (bijvoorbeeld verschillen tussen daders die voor verschillende behandelingen zijn geselecteerd).
- Historie: het effect wordt veroorzaakt door een andere gebeurtenis die zich op hetzelfde tijdstip voordeed als de interventie.
- Maturatie: het effect is een weerslag van een continuering van reeds bestaande trends.
- Instrumentatie: het effect is veroorzaakt door een verandering in de manier waarop de uitkomst wordt gemeten (bijvoorbeeld recidive).
- Testen: de pre-testmeting leidt tot een verandering in de post-testmeting.
- Regressiegemiddelde: waar een interventie op eenheden met uitzonderlijk hoge scores wordt toegepast (bijvoorbeeld gebieden met veel misdaad), kunnen natuurlijke fluctuaties tot lagere scores in de post-test leiden, die ten onrechte geïnterpreteerd kunnen worden als het effect van een interventie.
- Differentiële uitval: het effect wordt veroorzaakt door de gedifferentieerde uitval van eenheden (bijvoorbeeld mensen) uit de experimentele condities, en niet uit de controlecondities, ten tijde van de follow-up.

- Causaal verband: het is niet duidelijk of de interventie geschiedde voor de verandering in de uitkomst.

In principe heeft een gerandomiseerd experiment waarin eenheden willekeurig worden toegewezen aan experimentele of controlecondities de hoogst mogelijke interne validiteit, omdat hiermee alle risico's uitgesloten kunnen worden (of alternatieve verklaringen voor de bevindingen). Mits er voldoende eenheden zijn toegewezen, zijn na randomisatie de eenheden in de ene conditie op alle mogelijke (gemeten en ongemeten) externe variabelen die van invloed kunnen zijn op de afhankelijke variabele gelijk aan de eenheden in de andere conditie. Het grootste voordeel van een gerandomiseerd experiment is dat de invloed van de onafhankelijke variabele losgekoppeld kan worden van alle andere mogelijke invloeden op de afhankelijke variabele.

Andere methoden

Zoals hierboven reeds genoemd, is een experiment vooral geschikt voor het testen van een causale hypothese over de invloed die de ene factor op de andere heeft. Sherman en Berk (1984) waren immers geïnteresseerd in het testen van de hypothese dat arrestatie, separatie en bemiddeling verschillende effecten hadden op de recidive bij plegers van huiselijk geweld. De experimentele methode is bij afwezigheid van dergelijke causale hypothesen niet geschikt. Onderzoek waarbij hypothesen worden getest, dienen idealiter vooraf te worden gegaan door hypothese-genererend onderzoek. Dit hoeft geen experimenteel onderzoek te zijn. Zo kunnen ongestructureerde interviews of deelnemerobservaties bijvoorbeeld effectiever zijn dan experimenten in het genereren van hypothesen, hoewel experimenten effectiever zijn voor het testen van hypothesen.

Experimenten lenen zich beter voor het bepalen of veranderingen in een onafhankelijke variabele veranderingen in een afhankelijke variabele tot gevolg hebben dan het bepalen van het relatieve belang van, of interactie tussen, een groot aantal onafhankelijke variabelen. Sherman en Berk toonden aan dat arrestaties leidden tot een afname in recidive (vergeleken met andere methodes), maar constateerden geen relatieve invloed van de politieaanpak en andere variabelen zoals leeftijd, ras, strafblad en werkloosheid op

recidive. Vele factoren, zoals bijvoorbeeld leeftijd en ras, konden alleen in simulatie-experimenten als onafhankelijke variabelen worden gemanipuleerd. Correlatie- of regressieanalyses zijn wellicht beter geschikt voor het vaststellen van het relatieve belang. Een interessant punt werd gemaakt door Sherman (1992). Hij ontdekte een interactie-effect, omdat arrestaties voor huiselijk geweld het effectiefst waren in het reduceren van recidive bij daders met een 'hoge mate van conformiteit' (bijvoorbeeld omdat zij een baan hadden of getrouwd waren).

Quasi-experimentele analyses zijn geschikt voor het testen van causale hypothesen op basis van naturalistische data, waar de onderzoeker niet prospectievelijk kan bepalen welke persoon welke behandeling krijgt. Farrington (1977) testte bijvoorbeeld de hypothese dat officiële labelling door het crimineel justitieel apparaat leidt tot een toename van afwijkend gedrag. De theoretische construct van officiële labelling werd operationeel gedefinieerd en gemeten aan de hand van een eerdere veroordeling, terwijl het afwijkend gedrag werd gemeten aan de hand van self-reports voor delictpleging. De resultaten toonden aan dat self-reported delictpleging na een eerste veroordeling voor mannen toenam, vergeleken met mannen die eerder (op 14-jarige leeftijd) op self-reported delictpleging waren gematcht maar die niet veroordeeld waren. Bovendien concludeerde Farrington (1977) dat deze resultaten standhielden na ze te hebben getoetst op risico's voor de interne validiteit zoals hierboven uiteengezet, en dat de meer vijandige houding tegenover de politie een indirecte factor zou kunnen zijn in het verband tussen de eerste veroordeling en de toename in delictpleging.

Hoewel een bepaald experiment op overtuigende wijze kan aantonen of een verandering in A tot een verandering in B leidt, is er meestal geen verklaring voor de indirecte variabelen of processen die tussen A en B interveniëren. Zo verklaren Sherman en Berk (1984) bijvoorbeeld niet waarom arrestatie effectiever was dan separatie of bemiddeling, hoewel het afschrik-effect meer voor de hand lijkt te liggen dan het effect van labelling of versterking van afwijkend gedrag. Het zijn losse eindjes zoals die bij de meeste experimenten worden gevonden en die vaak leiden tot nog meer testbare hypothesen die in andere experimenten getoetst kunnen worden.

Eerdere experimenten in de criminologie

In het licht van de methodologische voordelen van gerandomiseerde experimenten is het verassend om te zien dat er in de criminologie relatief weinig van deze experimenten zijn uitgevoerd. De reden hiervoor zijn de vele ethische, juridische en praktische obstakels waar onderzoekers die een gerandomiseerd experiment willen opzetten, tegenaan lopen. Met dit artikel willen wij onderzoekers die meer willen weten over gerandomiseerde experimenten die in de criminologie zijn uitgevoerd, een bruikbaar middel aanreiken. Wij hopen dat de informatie over succesvol afgeronde experimenten onderzoekers zal helpen bij het in de toekomst opzetten van gerandomiseerde experimenten.

Farrington (1983) bestudeerde de gerandomiseerde experimenten op het gebied van criminaliteit en justitie die tussen 1957 en 1981 in het Engels waren gepubliceerd. Hierbij werd alleen gekeken naar experimenten die waren uitgevoerd in 'real life' settings, dus zonder laboratoriumexperimenten en simulatieoefeningen. Ook waren methodologische experimenten, bijvoorbeeld het vergelijken van verschillende methodes voor het maken van een 'self report' vragenlijst (Hindelang, Hirschi e.a., 1981), uitgesloten van het onderzoek. Ook werden binnensubject-onderzoeken, waarbij dezelfde mensen alle behandelingen in willekeurige volgorde ondergingen, niet meegenomen. Tot slot werden alleen die experimenten meegenomen waarin ten minste vijftig personen aan iedere conditie werden toegewezen, omdat er werd geredeneerd dat het bij kleinere aantallen onwaarschijnlijk was dat het belangrijkste voordeel van randomisatie (zorgen voor gelijkwaardigheid van personen op alle gemeten en niet-gemeten externe variabelen) zou worden gerealiseerd. Farrington (1983) presenteerde overzichtstabellen met daarin de belangrijkste kenmerken voor 22 sociale interventie-experimenten op het gebied van preventie, voorwaardelijke vrijlating voor jongeren, proeftijd en behandeling in de gemeenschap; en voor twintig experimenten waarbij de politie, rechtbanken en opvoedkundige behandelingen waren betrokken. Opvallend is het relatief kleine aantal criminologische experimenten dat tot dan was gepubliceerd. De meeste zijn uitgevoerd in de Verenigde Staten (31) of Groot-Brittannië (8), twee in Canada en één in Denemarken.

Slechts 35 experimenten betroffen delictplegingen, waarvan er 25 geen significant effect vonden na een interventie. Omdat de

meeste behandelingen waren gebaseerd op een soort van counseling, was de ondoeltreffendheid vaak het gevolg van het kleine of niet-bestaande effect dat counseling heeft op recidive. In sommige gevallen waren de resultaten echter niet significant omdat de steekproef te klein was en de statistische kans een effect te vinden dienovereenkomstig klein. Tegenwoordig ligt de nadruk veel meer op het meten van de omvang van het effect dan op het tellen van statistisch significante resultaten, en wordt er voor het schatten van de omvang van het effect vaak gebruikgemaakt van meta-analysetechnieken (Farrington en Welsh, 2001; Lipsey en Wilson, 2001; Welsh en Farrington, 2006).¹

Meer recent keken Farrington en Welsh (2005) naar gerandomiseerde experimenten in de criminologie gepubliceerd in het Engels tussen 1982 en 2004, en stelden Farrington en Welsh (2006) een gedetailleerd overzicht samen van gerandomiseerde experimenten in de beide tijdperiodes die samen bijna een halve eeuw onderzoek besloegen. In de tweede tijdperiode (85) vonden zij veel meer experimenten met recidiveresultaten dan in de eerste periode (in totaal werden er 37 experimenten onderzocht). Net als in de eerste periode werden er in de tweede periode maar weinig experimenten buiten de Verenigde Staten uitgevoerd (vier in Australië, drie in Canada, een in Duitsland, een in Zwitserland en een gedeeltelijk in Hongkong).

Er werd een aantal opvallende veranderingen in de aard van de experimenten tussen de eerste en tweede tijdperiodes gevonden. Zo was er een toename in het aantal grootschalige replicatie-experimenten op verschillende locaties. Op wetenschappelijke gronden zijn deze experimenten zeer geliefd. Er was ook een toename van het aantal experimenten waarin locaties willekeurig werden toegewezen, vooral om politie-initiatieven te kunnen bestuderen (bijvoorbeeld politieaanpak op zogeheten 'hot spots'; zie Sherman en Weisburd, 1995). Er was ook een groter aantal experimenten (met

1 Tegenwoordig vindt men het beter om de gevonden effectgrootte van elke afzonderlijk onderzoek te combineren in een gewogen gemiddelde effectgrootte (Weighted Mean Effect Size, WMES). Niet-significante effectgroottes in alle onderzoeken kunnen gezamenlijk een WMES, een gewogen gemiddelde effectgrootte produceren. De WMES is een belangrijker en zinvoller gegeven dan te weten dat X van Y studies een significant resultaat hebben opgeleverd. Een resultaat dat significant is bij $p = .05$ kan zowel wijzen op een groot effect in een klein sample als op een klein effect in een groot sample. De basisassumptie van meta-analyse is dat de omvang van het effect (en niet de statistische significantie) de belangrijkste parameter is bij het antwoord op de vraag: hoe goed werkte het?

name op het gebied van ontwikkelingspreventie; zie Farrington en Welsh, 2003) met langtermijn follow-ups, en een trend waarbij deelnemerinterviews werden gecombineerd met officiële verslagen.

Opmerkelijke criminologische experimenten, 1957-1981

Hieronder volgt een korte beschrijving van de meest opmerkelijke criminologische experimenten uit de twee tijdperiodes. Zie voor meer informatie over deze experimenten Farrington (1983), Farrington en Welsh (2005) en Farrington en Welsh (2006). Tabel 1 bevat een overzicht van de belangrijkste kenmerken van twaalf experimenten gepubliceerd in de eerste tijdperiode. Voor de duidelijkheid geven wij (waar mogelijk) het procentuele verschil in recidive in de experimentele conditie vergeleken met de controleconditie. Onze eenvoudige maatvoering voor het meten van het effect is de proportionele afname in recidive in de experimentele groep, uitgedrukt als procentuele verandering. Zo is de recidiveafname van 63% in de controleconditie naar 44% in de experimentele conditie zoals die door Palmer (1974) wordt gerapporteerd bijvoorbeeld een afname van 30% (want $44/63 = 0,70$).

De datum in tabel 1 is die van de meest significante publicatie. Dus hoewel de eerste publicatie van de hand van Reiner en Warren (1957) was, was het eerste criminologische experiment dat werd gepubliceerd de Cambridge-Somerville Youth Study, geëvalueerd door niet alleen Powers en Witmer (1951), maar, belangrijker nog, door McCord (1978; zie hieronder). In de eerste experimentele evaluatie van supervisie in de gemeenschap wezen Reiner en Warren een kleine 3800 voorwaardelijk vrijgelaten mannen willekeurig toe aan ofwel een reclasseringsambtenaar met slechts 15 cliënten, die intensieve supervisie gaf; ofwel aan een reguliere supervisie door een reclasseringsambtenaar met negentig cliënten. Een follow-up van 23 maanden vond geen significant verschil in recidive tussen de twee condities, gemeten aan de hand van arrestaties voor zware misdrijven.

In het eerste politie-experiment evalueerden Rose en Hamilton (1970) een politie-liaisonprogramma voor minderjarigen in Blackburn, Engeland. Jongeren konden aan dit programma meedoen als het hun eerste overtreding betrof, als ze schuld bekenden, als hun families bereid waren mee te werken en akkoord gingen met super-

Tabel 1: Opmerkelijke criminologische experimenten, 1957-1981

Publicatie Locatie	Eerste Steekproef	Conditie	Resultaten (n)	% Verandering
Reimer en Warren (1957) CA	3.793 voorwaardelijk vrijgelaten mannen	E = geringe patiëntenlast C = hoge patiëntenlast	23 maanden zware arrestaties E 14,2% (1.479) C 15,7% (2.314)	- 10%
Rose en Hamilton (1970) Blackburn, VK	394 gearresteerde jongens	E = officiële waarschuwing en supervisie C = officiële waarschuwing	30 maanden recidive E 26,8% (194) C 25,5% (200)	+ 5%
Kassebaum, Ward en Wilner (1971) CA	512 mannelijke gevangenen	E1 = counseling in kleine groep E2 = counseling in grote groep C = controle	arrestaties na 3 jaar E1 57% (171) E2 70% (68) C 58% (269)	E1 v C: - 2% E2 v C: + 21%
Empey en Lubeck (1971) CA	261 mannelijke recidivisten	E = gemeenschapsprogramma incl. groepsinteractie onder begeleiding C = instelling	1 jaar gemiddelde delicten EB 2,71 (140) EA 0,73 (140) CB 2,66 (121) CA 0,74 (121)	-3%
Empey en Erickson (1972) UT	150 mannelijke recidivisten	E = gemeenschapsprogramma incl. groepsinteractie onder begeleiding C = proeftijd	arrestaties na 4 jaar EM 3,31 (62) CM 2,83 (69)	+ 17%
Palmer (1974) CA	802 mannelijke delinquenten	E = behandeling in de gemeenschap C = instituut	intrekking na 2 jaar E 44% (n.a.) C 63% (n.a.)	-30% (sig, gesteld, maar niet gerapporteerd)

Shaw (1974) VK	176 mannelijke gevangenen	E = met counseling C = controle	nieuwe veroordeling na 2 jaar E 57,3% (75) C 78,1% (73)	- 27%*
Cornish en Clarke (1975) Bristol, VK	173 delinquenten in inrichting	E = therapeutische gemeenschap C = traditionele training	nieuwe veroordeling na 2 jaar E 69,8% (86) C 69,0% (87)	+ 1%
Folkard, Smith en Smith (1976) 4 locaties in VK	900 voorwaardelijk vrijgelaten mannen en vrouwen	E = intensieveproeftijd C = standaardproeftijd	nieuwe veroordeling na 1 jaar E 32,2% (475) C 29,6% (425)	+ 9%
Berg e.a. (1978) Leeds, VK	96 jongens en meisjes spijbelaters	E = schorsing C = supervisie	6 maanden recidive E 13,3% (45) C 33,3% (51)	- 60%*
McCord (1978) Cambridge en Somerville, MA	650 schooljongens	E = met counseling C = geen behandeling	30 jaar volwassen zware veroordelingen E 19,4% (253) C 16,6% (253)	+ 17%
Rossi, Berk, en Lenihan (1980) LIFE, Baltimore, MD	432 mannelijke ex-gedetineerden	E1 = baan en financiële hulp E2 = financiële hulp E3 = baan C = geen behandeling	nieuwe arrestaties na 1 jaar (totaal) E1+E2 49,5% (216) E3+C 56,9% (216)	- 13%*

NB: * p<.05; E = experiment; C = controle; EB = voor experiment; CB = voor controle; EA = na experiment; CA = na controle; EM = experimentgemiddelde; CM = controlegemiddelde; n.be. niet beschikbaar.

visie door de politie, en als ook degene die de aangifte deed akkoord ging met het feit dat de politie niet zou vervolgen. De supervisie door de politie duurde zes maanden. Alle hiervoor in aanmerking komende gearresteerde jongens werden willekeurig toegewezen en kregen ofwel alleen een waarschuwing ofwel een waarschuwing en supervisie door de politie. De follow-up van dertig maanden liet zien dat de recidivepercentages van deze twee groepen grotendeels overeenkwamen, en dus was de conclusie dat de supervisie door de politie ineffectief was. Deze conclusie kwam de overheid toentertijd goed uit, omdat zij van mening was dat de supervisie van jongeren een taak was voor maatschappelijk werkers, niet de politie.

In het eerste experiment met volwassen gevangenen maakten Kassebaum, Ward e.a. (1971) gebruik van de opening van een nieuwe gevangenis om nieuwe gevangenen willekeurig toe te kunnen wijzen aan één van drie woonruimtes, elk met een ander regime: counseling in kleine groepen, counseling in grote groepen, of controle (traditioneel, geen counseling). Uit de follow-up van drie jaar bleek dat geen enkele vorm van counseling leidde tot een afname van de recidive (vergeleken met de controlegroep). Over het algemeen gesproken hebben ook andere experimenten aangetoond dat counseling niet effectief is bij het terugdringen van recidive. Een uur counseling per week is immers te verwaarlozen op de resterende 167 uur die de gevangene die week in de gevangenis doorbrengt. Drie klassieke criminologische experimenten waren het Silverlake-experiment (Empey en Lubeck, 1971), het Provo-experiment (Empey en Erickson, 1972) en het Californische 'Behandeling-in-de-gemeenschap-project' (Palmer, 1974). De eerste twee hiervan evalueerden het succes van een controleprogramma in een alternatieve woongemeenschap met daarin dagelijkse groepsbijeenkomsten met gestuurde groepsinteractie, scholing en beperkte werk- en lesactiviteiten. In het Silverlake-experiment werden daders willekeurig toegewezen aan hetzij dit programma, hetzij de institutionele behandeling, terwijl daders in het Provo-experiment willekeurig werden toegewezen aan hetzij het programma, hetzij voorwaardelijk werden vrijgelaten. Uit de follow-ups van één en vier jaar respectievelijk bleek dat geen enkel programma de recidive significant had doen afnemen.

Het Californische 'Behandeling-in-de-gemeenschap-project' bestond uit een intensieve supervisie in de gemeenschap als alternatief voor plaatsing in een instelling. In een experimentele evalu-

atie concludeerde Palmer (1974) dat er een significante afname was van het aantal keren dat een voorwaardelijke vrijlating werd ingetrokken in een follow-up van twee jaar (44% in de experimentele groep versus 63% in de controlegroep). Lerman (1975) concludeerde echter dat, hoewel het aantal nieuwe arrestaties voor de behandeling in de gemeenschap en institutionele groepen weliswaar vergelijkbaar was, er een verschil zat in de kans die men liep dat de voorwaardelijke vrijlating na arrestatie werd ingetrokken. Zijn conclusie was daarom dat het project succesvol was in het veranderen van het autonome besluitvormingsgedrag van volwassen opvoedingsambten, maar dat het geen duidelijk effect had gehad op het gedrag van de delinquenten.

De daarop volgende vier experimenten vonden plaats in de jaren zeventig in Groot-Brittannië. Dit was een gouden tijd voor Britse criminologische experimenten (Farrington, 2003b). Shaw (1974) wees mannelijke gevangenen willekeurig toe aan hetzij speciale bijstand door welzijnswerkers in de laatste zes maanden van hun straf, hetzij een controleconditie. De gevangenen in het experiment hadden over het algemeen vaker contact met hun welzijnswerkers (dertien maal in plaats van zes) en liepen beduidend minder risico opnieuw veroordeeld te worden in de twee hierop volgende jaren (57% versus 78%). Volgens Shaw was het actieve ingrediënt van deze behandeling de grotere belangstelling van de welzijnswerkers. Er bestond een interactie tussen het type dader en het type behandeling, waarbij introvertere gevangenen het met name in de experimentele conditie goed deden, en extraverte gevangenen in de controleconditie. Cornish and Clarke (1975) wees jongens op Kingswood-opleidingscholen willekeurig toe aan respectievelijk een therapeutische gemeenschap en een traditioneel paternalistisch regime. Tussen de regimes werd geen verschil gevonden in de kans die de jongens liepen op een nieuwe veroordeling in de follow-up periode van twee jaar. Wellicht vanwege dit negatieve resultaat verloren Clarke en Cornish (1972) hun belangstelling voor de gerandomiseerde experimentele methode en concludeerden dat 'het bijzonder onwaarschijnlijk is dat het wijdverspreide gebruik hiervan onze kennis over institutionele behandelingen aanzienlijk zal vergroten, en wel op een manier die anders niet mogelijk is' (p. 21). De kritiek van Clarke-Cornish speelde een grote rol in het beëindigen van de door het Ministerie van Binnenlandse Zaken gesubsidieerde gerandomiseerde experimenten in de 25 jaar die hierop volgden (Nuttall, 2003).

In de IMPACT-studie (Intensive Matched Probation and After-Care Treatment (Intensieve Matched Reclasseringsbehandeling)), werden negenhonderd daders op proeftijd op vier verschillende locaties in Engeland willekeurig toegewezen aan een reguliere proeftijd (met een gemiddelde patiëntenlast van 40-45 per beambte) of aan een intensive proeftijd (waar beambten een patiëntenlast hadden van niet meer dan twintig daders en geen rechtszaken of reclasseringsrapporten hoefden voor te bereiden). Dit was het eerste criminologische experiment dat op dergelijk grote schaal en op zoveel verschillende plekken in Groot-Brittannië werd gehouden. De daders in het experiment hadden maandelijks gemiddeld zo'n twee keer vaker contact met hun reclasseringsbeambten (drie keer in plaats van anderhalf keer in de controlegroep). Toch was er weinig verschil in het percentage dat na een jaar opnieuw werd veroordeeld (Folkard e.a. 1974; Folkard, Smith e.a. 1976). Uit een latere grootschalige evaluatie op meerdere locaties van intensive supervisie in de Verenigde Staten bleek ook dat deze methode niet effectief was in het terugdringen van recidive (Petersilia en Turner, 1993).

Het eerste Britse gerandomiseerde experiment met gerechtelijke uitspraken werd uitgevoerd door Berg e.a. (1978; zie ook Berg, Hullin e.a. 1979). Zij onderzochten de relatieve effectiviteit van twee uitspraken voor spijbelaars, schorsing (continuatie) versus supervisie door maatschappelijk werkers. De kinderrechter in Leeds maakte toentertijd van beide uitspraken gebruik. Uit een retrospectieve studie (Berg e.a., 1977) zou blijken dat schorsing effectiever was in het terugdringen van verder spijbelgedrag dan supervisie. Het is echter mogelijk dat de kinderen die voor de schorsing werden geselecteerd op bepaalde punten verschilden van de kinderen die voor de supervisie werden geselecteerd, en dat deze reeds bestaande verschillen tot een verschil in resultaat leidden.

In samenwerking met de kinderrechter in Leeds werden bijna honderd spijbelaars willekeurig toegewezen aan hetzij schorsing, hetzij supervisie. De inzet en het enthousiasme van de voorzitter van de Jeugdrechtbank (Juvenile Magistrates) (Hullin), een biochemicus aan de Universiteit van Leeds, was essentieel voor de totstandkoming van dit experiment (Berg, Brown e.a. 1988, p. 113). Tijdens een follow-up periode van zes maanden bleken de jongeren onder supervisie vaker te spijbelen en pleegden meer delicten. Schorsing als uitspraak bleek dus effectiever in het voorkomen van spijbelgedrag en delinquentie. De reden hiervoor was waarschijnlijk

dat geschorste zaken werden afgehandeld door educatiewelzijns-
werkers, die er alles aan deden om de kinderen weer naar school te
krijgen, terwijl de maatschappelijk werkers die de supervisie gaven
grotere prioriteit gaven aan de aanpak van de gezinsproblemen.
Volgens Weisburd en Petrosino (2004) was de Cambridge-Somer-
ville Youth Study (Powers en Witmer, 1951) het eerste grootschalige
gerandomiseerde experiment dat ooit op het gebied van de crimino-
logie werd uitgevoerd. McCord (1978) volgde in dit experiment ruim
vijfhonderd mannen die op tienjarige leeftijd (gemiddeld) willekeu-
rig waren toegewezen aan hetzij individuele counseling en thuis-
bezoeken, hetzij een controlegroep. De counselors praatten met de
jongens, namen ze mee op uitstapjes en deden leuke dingen met
hen, gaven hen lees- en rekenonderwijs, moedigden hen aan deel te
nemen aan de YMCA en zomerkampen, deden spelletjes met hen in
het projectcentrum, stimuleerden hen naar de kerk te gaan, hielden
nauw contact met de politie, en gaven advies en algemene steun aan
de gezinnen en families. De behandeling duurde over het algemeen
vijf jaar, en de follow-up vond plaats toen ze rond de 45 jaar waren.
Helaas bleek de behandeling tot ongewenste effecten te hebben
geleid. Uit de behandelgroep werden meer mensen op volwassen
leeftijd veroordeeld voor ernstige misdrijven, en pleegden er signi-
ficant meer daders uit de behandelgroep twee of meer misdrijven,
dan de mensen in de controlegroep. Uit de behandelgroep stierven
er meer mannen op jonge leeftijd, hadden er meer stressgerelateerde
ziekten, of neigden naar alcoholisme, en waren er minder getrouwd.
McCord (1978) speculeerde dat de behandeling wellicht had geleid
tot hoge verwachtingen en een mate van afhankelijkheid die nega-
tief uitpakte toen de behandeling werd gestopt.

Het uiteindelijke experiment evalueerde de effectiviteit van tran-
sitionele hulp aan voormalige gevangenen bij hun reïntegratie in
de maatschappij. Transitionele hulp vond plaats in de vorm van
beperkte financiële hulp of een werkloosheidsuitkering die zij gedu-
rende een korte periode ontvingen (tussen de 13 en 26 weken). In het
eerste van deze experimenten, bekend als Living Insurance for Ex-
Prisoners (LIFE, of Onderhoudsverzekering voor ex-gedetineerden),
uitgevoerd in Baltimore, wezen Rossi, Berk e.a. (1980) meer dan
vierhonderd mannelijke voormalige ex-gedetineerden willekeurig
toe aan een van drie experimentele groepen die financiële hulp,
intensieve loopbaanbegeleiding en plaatsingsassistentie kregen, of
beide, of aan een controlegroep die geen uitkering of begeleiding

kreeg. In de follow-up periode van een jaar werd een significante 13% afname in recidive gevonden voor mensen die financiële hulp hadden ontvangen. Twee replicatie-experimenten, bekend als het Transitional Aid Research Project (TARP), die met een veel grotere steekproef in Georgia en Texas werden uitgevoerd, gaven echter niet de resultaten uit het LIFE-experiment. Het is algemeen bekend dat interventies die in een demonstratieproject worden geëvalueerd, met kleine aantallen, speciaal geselecteerde medewerkers en kwaliteitscontrole door academische onderzoekers, meer resultaat opleveren dan interventies die routinematig en door het vaste personeel op grote aantallen worden toegepast (Lipsey, 2003).

Opmerkelijke criminologische experimenten, 1982-2004

In tabel 2 staan 15 opmerkelijke criminologische experimenten die in de tweede periode zijn uitgevoerd. In een aantal gevallen dateerde de meest significant publicatie van 2005. Het eerste experiment door Lewis (1983) is belangrijk omdat het de schadelijke effecten van het 'Scared Straight'-programma blootlegde. Hierin spraken volwassen gevangenen met jonge delinquenten over de gruwelen van het gevangenisleven om daarmee de jongeren af te schrikken. Het percentage veroordeelden lag in de Scared Straight-conditie echter 21% hoger, en kwam daarmee overeen met de klassieke evaluatie door Finckenauer (1982). Ook bleek uit een systematisch overzicht door Petrosino, Turpin-Petrosino e.a. (2003) dat de recidive na Scared Straight in één op zeven evaluaties groter was. Zij concludeerden (p. 58) dat 'niets doen beter zou zijn geweest dan deze jongeren bloot te stellen aan het programma'.

Het 'Minneapolis huiselijk geweld experiment' (Sherman en Berk, 1984) is van groot historisch belang. Hierin werd aangetoond dat het arresteren van de daders leidde tot minder nieuwe incidenten van huiselijk geweld dan wanneer de dader werd gesommeerd het pand acht uur te verlaten of advies en bemiddeling te ondergaan. De resultaten werden direct opgepikt door het Amerikaanse Ministerie van Justitie, dat politiediensten ging stimuleren om mannelijke daders van huiselijk geweld te arresteren in plaats van hen op een andere manier aan te pakken (Sherman en Cohn, 1989). Volgens Sherman (1992, p. 103) 'waren mensen dankzij de berichtgeving hierover gewend aan het idee van gerandomiseerde experimenten'.

Het bleek het begin van de gouden jaren tachtig in Amerikaanse gerandomiseerde experimenten (Farrington, 2003a). De resultaten van het oorspronkelijke Minneapolis-experiment waren voor het Amerikaanse National Institute of Justice aanleiding een ambitieus programma experimenteel onderzoek te sponsoren om te kijken of de bevindingen in andere omgevingen konden worden gereproduceerd. De resultaten bleken echter nogal variabel (Garner, Fagan e.a., 1995; Maxwell, Garner e.a., 2002; Sherman, 1992).

Het volgende experiment, door Klein (1986), evalueerde de effectiviteit van officiële aanklachten vergeleken met voorwaardelijke vrijlating voor jongeren. In tabel 2 staat alleen de vergelijking tussen de meest extreme condities, dus aanklacht vergeleken met vrijlating. Het recidivepercentage lag voor de rechtzaken significant hoger (49%) dan voor de voorwaardelijke vrijlating, geheel in lijn met de labellingtheorie. Niet in lijn met de labellingtheorie waren echter de bevindingen van Barton en Butts (1990), die tussen intensieve supervisie in de gemeenschap en plaatsing in een instelling weinig verschil constateerden in recidive door jongeren.

Sherman en Weisburd (1995) onderzochten de effecten van intensieve politiepatrouilles in 'hot spots', oftewel plekken met veel criminaliteit (zie Braga, 2001). Dit experiment is opmerkelijk omdat de eenheid die werd gerandomiseerd een gebied was, en geen individu. Uit de resultaten bleek dat intensievere politieaanwezigheid in hot spots leidde tot een significante (11%) afname in het aantal keren dat er vanwege misdaad een beroep op hen werd gedaan.

Net als 'Scared Straight', waren ook *boot camps* bedoeld om jongeren af te schrikken de misdaad in te gaan. Peters, Thomas e.a. (1997) voerden echter een grote experimentele evaluatie op meerdere locaties uit en lieten zien dat deze *boot camps* over het algemeen niet effectief waren. In Cleveland (tabel 2) leidde het *boot camp* tot een significante (44%) toename in recidive vergeleken met opsluiting in een instelling voor jongeren. Deze resultaten zijn niet atypisch. Uit een systematisch overzicht van 29 evaluaties van *boot camps* door MacKenzie, Wilson e.a. (2001) bleek dat in negen studies de recidive onder deelnemers aan een *boot camp* lager lag, in acht onderzoeken hoger, en geen verschil kende in twaalf studies. Ze concludeerden dat 'een meta-analyse geen algemeen significante verschillen in recidive opleverde tussen deelnemers aan een *boot camp* en controlegroepen' (p. 126).

Tabel 2: **Opmerkelijke criminologische experimenten, 1982-2004**

Publicatie, Locatie	Eerste Steekproef	Conditie	Resultaten (n)	% Verandering
Lewis (1983) CA	108 mannelijke delinquenten	E = Scared Straight C = geen behandeling	Arrestaties na 12 maanden E 81,1% (53) C 67,3% (55)	+ 21%
Sherman en Berk (1984) Minneapolis, MN	330 verdachten huiselijk geweld	E = arrestatie C = bemiddeling of separatie	Opnieuw gearresteerd na 6 maanden E 6,5% (93) C 14,8% (237)	- 56%*
Klein (1986) US	306 arrestaties onder minderjarigen	E = vrijlating C = aanklacht	Arrestaties na 27 maanden E 48,8% (82) C 72,8% (91)	-33%*
Barton en Butts (1990) Detroit, MI	511 jongeren in hechtenis	E = intensieve supervisie C = plaatsing	Arrestaties na 2 jaar EM 3,69 (326) CM 3,58 (160)	+3%
Sherman en Weisburd (1995) Minneapolis, MN	110 hot spots	E = intensievere patrouille C = reguliere patrouille	Oproepen na 12 maanden EB 6531 (55) EA 6931 (55) CB 6491 (55) CA 7702 (55)	-11%*
Peters, Thomas, en Zamberlan (1997) Cleveland, OH	Zo'n 354 mannelijke delinquenten	E = <i>boot camp</i> C = opsluiting	Veroordelingen na 9 maanden E 72% (182) C 50% (172)	+44%*
Olds e.a. (1998) Elmira, NY	400 zwangere vrouwen	E = huisbezoeken C = geen huisbezoeken	Veroordelingen na 15 jaar E 16,6% (176) C 36,0% (148)	-54%*

Killias, Aebi, en Ribeaud (2000) Zwitserland	123 veroordeelde daders	E = taakstraf C = gevangenis	Veroordelingen na 2 jaar EB 92,9% (84) EA 21,4% (84) CB 84,6% (39) CA 25,6% (30)	-24%**
Ortmann (2000) Duitsland	228 mannelijke gevangenen	E = sociale therapie gevangenis C = reguliere gevangenis	Veroordelingen na 5 jaar E 60,4% (111) C 67,9% (112)	-11%
Feder en Dugan (2002) Broward County, FL	404 verdachten huiselijk geweld	E = counseling plus proeftijd C = proeftijd	Arrestaties na 12 maanden E 24% (230) C 24% (174)	0%
Gottfredson, Najaka, en Kearley (2003) Baltimore, MD	235 mensen gearresteerd voor drugs	E = drugsrechtbank C = reguliere rechtbank	Arrestaties na 2 jaar E 66,2% (139) C 81,3% (96)	-19%*
van Voorhis e.a. (2004) Georgia	468 voorwaardelijk vrijgelaten mannen	E = Reasoning and Rehabilitation C = geen Reasoning and Rehabilitation	Arrestaties na 9 maanden E 37,8% (229) C 42,4% (232)	- 11%
Schweinhart e.a (2005) Ypsilanti, MI	123 kinderen	E = kleuterschool C = geen kleuterschool	Arrestaties voor misdrijven tot 40 jaar E 31% (58) C 48% (65)	-35%**
Kling, Ludwig, en Katz (2005) 5 locaties in VS	3.079 mensen	E = verhuizing naar betere buurt C = geen middelen	Arrestaties na 5 jaar E 32,9% (1233) C 31,8% (943)	+3%
Strang en Sherman (2006) Canberra, Australië	143 minderjarige winkeldieven	E = herstelbijeenkoms C = reguliere rechtbank	Arrestaties na 12 maanden E 28,6% (77) C 42,4% (66)	-33%

Afkortingen: *p< .05; ** p<.10; E = experiment; C = Controle; EB = voor experiment; CB = voor controle; EA = na experiment; CA = na controle;
EM = experimentgemiddelde; CM = controlegemiddelde

Olds e.a. (1998) voerden een klassieke evaluatie uit van het effect van huisbezoekprogramma's voor zwangere vrouwen. De huisbezoekers (verpleegkundigen) gaven de vrouwen advies over de opvoeding, de ontwikkeling van het kind, de voeding en de noodzaak het gebruik van alcohol en drugs te vermijden. Kortom, een algemeen ouderopvoedprogramma. Een follow-up over een periode van 15 jaar liet zien dat de kinderen van bezochte moeders significant minder werden gearresteerd (54%) dan de kinderen van moeders die niet waren bezocht.

De volgende twee experimenten zijn eveneens opmerkelijk, gedeeltelijk omdat zij buiten de Verenigde Staten zijn uitgevoerd. In Zwitserland meldden Killias, Aebi e.a. (2000) dat het opleggen van een taakstraf in plaats van een gevangenisstraf leidde tot een bijna significante afname in veroordelingen van 24%. In Duitsland wees Ortmann (2000) gevangenen willekeurig toe aan hetzij een speciale sociaetherapiegevangenis, hetzij een reguliere gevangenis en concludeerde dat het aantal nieuwe veroordelingen na de sociaetherapiegevangenis lager lag (maar niet significant veel).

Het experiment van Feder and Dugan (2002) is van historisch belang omdat de implementatie ervan zoveel stof deed opwaaien. De plaatselijke aanklager wilde op juridische en ethische gronden een geding aanspannen tegen het experiment, namelijk wegens misbruik van bevoegdheden (een juridisch punt), en vanwege het feit dat het ontzeggen van een behandeling aan mensen op basis van loting onethisch was. In antwoord hierop stelde Feder (1998) dat het onethisch was een interventie (counseling) op te leggen zonder eerst de positieve en negatieve gevolgen daarvan uitvoerig te testen. Feder (1998, p. 7) stelde dat 'ofschoon vele organisaties ons onderzoek steunden, er slechts één naar voren kwam om ons bij onze rechtszaak te steunen, en dat was de American Society of Criminology'. De drie ASC Presidents die bij de zaak betrokken waren, stelden dat 'academische instellingen een verplichting hebben het principe te respecteren en te verspreiden dat willekeurige toewijzing aan behandelmogelijkheden de beste wetenschappelijke methode is voor het bepalen van de doeltreffendheid van mogelijkheden in een zaak als deze' (Short, Zahn e.a., 2000, p. 295). Uiteindelijk trok de aanklager de zaak in en werd het experiment afgerond.

Feder en Dugan (2002) concludeerden dat door het Hof goedgekeurde counseling voor mannelijke plegers van huiselijk geweld geen effect had op een eventuele latere arrestatie. In hun systema-

tisch programma-overzicht concludeerden Feder en Wilson (2006) echter dat de recidive iets was teruggedrongen. Ook wezen zij erop dat deze gunstige effecten alleen van toepassing zijn op een selecte en gemotiveerde groep mannen die zijn veroordeeld voor een geweldsdelict.

Het volgende experiment evalueerde de effectiviteit van drugsrechtbanken. Gottfredson, Najaka e.a. (2003) constateerden dat vergelijken met de reguliere rechtbanken drugsrechtbanken leidden tot een significante afname van nieuwe arrestaties, en wel met 19%. In een systematisch overzicht van 42 evaluaties van drugsrechtbanken concludeerden Wilson, Mitchell e.a. (2005) dat drugsrechtbankprogramma's effectief waren in het reduceren van delicten.

Het is algemeen bekend dat de meest effectieve behandelprogramma's voor daders denk- en gedragsprogramma's zijn. Van Voorhis e.a. (2004) evalueerden het effect van een populair denk- en gedragsprogramma genaamd 'Reasoning and Rehabilitation' op gedetineerden die voorwaardelijk vrij waren gelaten, en constateerden een kleine (11%) afname in recidive. Tong en Farrington (2005) concludeerden in hun systematisch overzicht van 26 evaluaties van het 'Reasoning and Rehabilitation'-programma echter dat het effectief was in het reduceren van nieuwe veroordelingen, met een gemiddelde afname van 14% in recidive.

De langste follow-up van een ontwikkelingspreventieprogramma werd uitgevoerd door Schweinhart e.a. (2005). In het beroemde Perry-project deden kinderen in het experiment mee aan een cognitief georiënteerd kleuterschoolprogramma dat was opgezet om hun denk- en redeneringsvermogens en schoolprestaties te vergroten. De kinderen in de experimentele en de controlegroep werden gevolgd tot hun veertigste, en uit de resultaten bleek dat er onder de kinderen uit het experiment 35% minder arrestaties voorkwamen.

In een groot gemeenschapsexperiment dat op meerdere locaties werd uitgevoerd (Kling, Ludwig e.a., 2005) ontvingen arme gezinnen middelen waarmee ze als onderdeel van het 'Moving to Opportunity'-programma naar een betere wijk konden verhuizen. Daarna werd onderzocht hoe deze verhuizing van invloed was op het plegen van delicten door hun kinderen. Het effect op het aantal arrestaties in zijn algemeen bleek klein, maar wel was er een duidelijk positief effect op het aantal arrestaties van meisjes voor geweldsdelicten en diefstal en een negatief effect op het aantal arrestaties van jongens voor diefstal. De auteurs speculeerden dat jongens en meisjes ver-

schillend reageerden op hun nieuwe buurt; wellicht dat de meisjes op hun welgesteldere klasgenoten reageerden door op school beter hun best te doen, terwijl jongens met wrok reageerden en gingen stelen van hun klasgenoten en op school niet hun best deden. Tot slot deden Strang en Sherman (2006) een aantal experimenten waarbij ze herstelconferenties vergeleken met de gebruikelijke zittingen in Australië en Engeland. De resultaten van de experimenten waren gemengd. Tabel 2 laat het meest positieve resultaat zien, namelijk een substantiële afname van 33% in recidive door jonge winkeldieven.

Problemen van gerandomiseerde experimenten

Farrington en Welsh (2006) bespraken een aantal problemen van gerandomiseerde experimenten. Zoals eerder gemeld moeten experimenten bijvoorbeeld op theorie-gebaseerde causale hypothesen testen, maar toetsten experimenten vaker een soort techniek waarvan wordt vermoed dat deze effectief zal zijn. Multi-modale interventies zijn gebruikelijk, omdat onderzoekers hun kansen een effect te vinden willen optimaliseren (Wasserman en Miller, 1998). Het blijkt echter vaak moeilijk de verschillende elementen van een complex interventiepakket te ontleden om zo de 'actieve ingrediënten' te kunnen definiëren. Om aan te tonen dat een pakket 'werkt' moeten idealiter verschillende experimenten worden onderzocht, maar dit is niet gebruikelijk.

In criminologische experimenten, vooral waarbij gebruik wordt gemaakt van interviews, is differentiële uitval uit interventie- en controlegroepen een groot probleem. Om de voordelen van randomisatie te kunnen behouden, is het belangrijk een analyse te maken van degenen die oorspronkelijk aan de condities zijn toegewezen (in een 'behandelintentie'-analyse) en degenen die de behandeling voltooiën (in een 'behandeling van behandelde'-analyse). Meestal ziet men dat degenen die de behandeling hebben voltooid het beter doen dan de mensen in de controles, terwijl degenen die de behandeling niet hebben afgemaakt het slechter doen. Deze verschillen kunnen het gevolg zijn van selectie-effecten (reeds bestaande verschillen tussen de mensen onderling die de behandeling wel en niet afmaken) en niet zozeer het effect van het wel of niet afronden van een behandeling. Mensen die de behandeling hebben afgerond, kunnen dus niet goed worden

vergeleken met de controles. Een voordeel van randomisatie binnen gematchte paren (bijvoorbeeld McCord, 1978) is dat, indien de experimentele persoon de behandeling niet afmaakt, beide mensen van het paar uit de analyse kunnen worden gehaald waarmee het probleem van differentiële uitval wordt voorkomen.

In experimenten wordt het effect vastgelegd van een interventie en wordt deze vergeleken met wat de controlegroep is geboden. Als er geen verschil is in uitkomst tussen de experimentele en controlecondities, dan kan het zijn dat beide even effectief zijn geweest. Het is dus belangrijk om net zo uitgebreid vast te leggen wat er met de controles gebeurt als met het experiment. Zoals gezegd had in de evaluatie van de 'intensieve' proeftijd door Folkard e.a. (1976), de 'intensieve' groep gemiddeld drie contacten per maand met een reclasseringsbeambte, vergeleken met anderhalf per maand voor de controlegroep. Het lijkt waarschijnlijk dat de omvang van de interventie (het verschil tussen anderhalf en drie contacten per maand) niet groot genoeg was om een effect te sorteren.

Ook al zijn er inmiddels een groot aantal gerandomiseerde experimenten op het gebied van criminaliteit en justitie uitgevoerd, toch blijven er nog veel vragen voor de onderzoekers. Zo voerden Farrington en Jolliffe (2002) bijvoorbeeld een studie uit naar de mogelijkheid zware criminelen met een ernstige persoonlijkheidsstoornis te evalueren middels een gerandomiseerde gecontroleerde trial. Zij constateerden dat alle klinici tegen een dergelijke trial waren omdat zij van mening waren dat iedereen behandeld moest worden en dat niemand behandeling ontzegd mag worden. Echter, als er veel meer personen behandeld willen of moeten worden dan mogelijk is, is willekeurige toewijzing wellicht de eerlijkste manier om mensen voor behandeling te selecteren (Wortman en Rabinowicz, 1979).

Wellicht geven programma-administrateurs de controle over de opdracht niet graag uit handen aan een onderzoeker verbonden aan het experiment, zelfs niet wanneer men akkoord gaat met randomisatie. Er zijn in de literatuur veel voorbeelden te vinden van randomisatie-experimenten waarbij niet aan de opzet werd voldaan, bijvoorbeeld omdat men uitzonderingen toeliet en de toewijzingsprocedure niet werd gecontroleerd door de onderzoekers. Idealiter mogen programma-uitvoerders en deelnemers niets weten over de experimentele condities en hypothesen. Dit is echter vaak moeilijk te realiseren, ofschoon Berg e.a. (1978) hun experiment voor de jongeren en het behandelend personeel (opvoedingswelzijnswerkers

en maatschappelijk werkers) geheim wisten te houden. De wens om deelnemers onwetend te houden van de condities staat overigens haaks op de ethische eisen van toestemming.

Waar het uiteindelijk op neerkomt, is dat de verwachte baten van een experiment opwegen tegen de verwachte kosten. Volgens Weisburd (2003) dient de belangrijkste vraag te zijn waarom er niet gebruikgemaakt moet worden van een gerandomiseerd experiment: 'Het is aan de onderzoeker om uit te leggen waarom een minder valide methode de basis zou moeten zijn voor het trekken van conclusies over behandeling en uitvoering' (p. 352). Ook andere onderzoekers hebben gepoogd bezwaren tegen gerandomiseerde experimenten uiteen te zetten en te beantwoorden (bijvoorbeeld Cook en Payne, 2002), en Boruch (1997) geeft gedetailleerd en praktisch advies over hoe dergelijke experimenten goed opgezet kunnen worden.

Conclusies

In onze conclusies zijn wij optimistisch. Hoewel gerandomiseerde experimenten in de criminologie vele uitdagingen kennen en nog altijd relatief weinig worden gebruikt, is het aantal redelijk groot-schalige experimenten op het gebied van misdaad in de periodes tussen 1957-1981 en 1982-2004 meer dan verdubbeld. Wat deze experimenten vergeleken met andere methodes zo uniek en aantrekkelijk maakt, is hun grote interne validiteit, of overtuigingskracht, om het echte effect van een interventie op het plegen van delicten vast te stellen. En dit is niet alleen een theoretisch voordeel. In verschillende overzichten is aangetoond dat de omvang van het effect in niet-experimenteel onderzoek typisch (misleidend) groter is dan in gerandomiseerde experimenten (bijvoorbeeld Weisburd e.a., 2001). Misschien wel het belangrijkste punt is dat de resultaten van gerandomiseerde experimenten in de tweede helft van onze tijdperiode over het algemeen bemoedigender zijn dan in de eerste helft, omdat zij suggereren dat sommige interventies effectief zijn (bijvoorbeeld vroege ontwikkelingspreventie, door de rechtbank opgelegde behandelprogramma's voor gewelddadige mannen, drugsrechtbanken, en de politieaanpak van 'hot spots'). Dit is ten dele vanwege de grotere nadruk die er tegenwoordig ligt op de omvang van het effect in plaats van het statistische belang. Experimenten tonen ook aan dat sommige interventies niet effectief zijn (bijvoorbeeld Scared Straight, *boot camps* en intensieve supervisie).

Er bestaan een aantal causale hypothesen, zoals de rol van gebroken gezinnen in het veroorzaken van misdaad, die nooit in gerandomiseerde experimenten getest kunnen worden, tenminste niet in de huidige Westerse maatschappij. Sommige variabelen, zoals leeftijd en geslacht, kunnen niet realistisch worden gemanipuleerd. Gebeurtenissen die zich reeds hebben voorgedaan, zoals wetswijzigingen, kunnen niet experimenteel worden onderzocht omdat experimenten prospectief dienen te zijn. Gerandomiseerde experimenten zijn vooral geschikt wanneer de behandel-effecten onbekend zijn en wanneer het onmogelijk is iedereen te behandelen. Financieel gezien zou het verstandig zijn om sociaal en punitief beleid eerst in kleinschalige experimenten te onderzoeken alvorens te investeren in de grootschalige uitvoering hiervan. Dit is echter zelden gebeurd. We moeten niet te lang stilstaan bij de problemen bij het uitvoeren van gerandomiseerde experimenten in de criminologie. De afgelopen vijftig jaar zijn er ten minste honderd van dergelijke experimenten uitgevoerd. Het is duidelijk dat experimenten als deze mogelijk zijn, en wij hopen dat ons werk een steun is voor onderzoekers die in de toekomst gerandomiseerde experimenten willen opzetten. We eindigen met de conclusies van Weisburd en Petrosino (2004): wetenschappers, uitvoerders en beleidsmakers zijn het er steeds meer over eens dat misdaadbestrijding en beleid zoveel mogelijk gestoeld moeten zijn op wetenschappelijk onderzoek. Dit zou erop wijzen dat gerandomiseerde experimenten, ook wel de 'gouden standaard' voor evaluaties genoemd, een steeds belangrijker onderdeel van criminologische studies zullen vormen. Wij hopen dat andere landen, waaronder Nederland, samen met de Verenigde Staten de effectiviteit van criminologische interventies middels experimentele opzetten willen evalueren.

Literatuur

Barton, W., J.A. Butts
Viable options; intensive supervision programmes for juvenile Delinquents
 Crime and delinquency, 36e jrg.,
 1990, p. 238-256

Berg, I., I. Brown e.a.
Off school, in court; an experimental and psychiatric investigation of severe school attendance problems
 New York, Springer-Verlag, 1988

- Berg, I., M. Consterdine e.a.**
The effect of two randomly allocated court procedures on truancy
British journal of criminology, 18e jrg., 1978, p. 232-244
- Berg, I., R. Hullin e.a.**
A randomly controlled trial of two court procedures in truancy
In: D.P. Farrington, K. Hawkins e.a. (red.), *Psychology, law and legal processes*, London, Macmillan, 1979, p. 143-151
- Berg, I., R. Hullin e.a.**
Truancy and the courts; research note
Journal of child psychology and psychiatry, 18e jrg., 1977, p. 359-365
- Boruch, R.F.**
Randomized experiments for planning and evaluation; a practical guide
Thousand Oaks, CA, Sage, 1977
- Braga, A.A.**
The effects of hot spots policing on crime
Annals of the American Academy of Political and Social Science, 578, 2001, p. 104-125
- Campbell, D.T., J.C. Stanley**
Experimental and quasi-experimental designs for research
Chicago, Rand McNally, 1966
- Clarke, R.V.G., D.B. Cornish**
The controlled trial in institutional research
London, Her Majesty's Stationery Office, 1972
- Cook, T.D., D.T. Campbell**
Quasi-experimentation; design and analysis issues for field settings
Chicago, Rand McNally, 1979
- Cook, T.D., M.R. Payne**
Objecting to the objections to using random assignment in educational research
In: F. Mosteller, R.F. Boruch (red.), *Evidence matters; randomized trials in education research*, Washington, DC, Brookings Institution Press, 2002, p. 150-178
- Cornish, D.B., R.V. Clarke**
Residential treatment and its effects on delinquency
London, Her Majesty's Stationery Office, 1975
- Empey, L.T., M.L. Erickson**
The Provo experiment; evaluating community control of delinquency
Lexington, MA, D. C. Heath, 1972
- Empey, L.T., S.G. Lubeck**
The Silverlake experiment; testing delinquency theory and community intervention
Chicago, Aldine, 1971
- Farrington, D.P.**
The effects of public labelling
British journal of criminology, 17e jrg., 1971, p. 112-125
- Farrington, D.P.**
Randomized experiments on crime and justice
In: M. Tonry, N. Morris (red.), *Crime and justice; vol. 4,*

- Chicago, University of Chicago Press, 1983, p. 257-308
- Farrington, D.P.**
A short history of randomized experiments in criminology; a meagre feast
 Evaluation review, 27e jrg., 2003a, p. 218-227
- Farrington, D.P.**
British randomized experiments on crime and justice
 Annals of the American Academy of Political and Social Science, 589, 2003b, p. 150-167
- Farrington, D.P.**
Methodological quality standards for evaluation research
 Annals of the American Academy of Political and Social Science, 587, 2003c, p. 49-68
- Farrington, D.P. (red.)**
Integrated developmental and life-course theories of offending
 Advances in criminological theory, 14e jrg., 2005, New Brunswick, NJ, Transaction
- Farrington, D.P.**
Building developmental and life-course theories of offending
 In: F.T. Cullen, J.P. Wright e.a. (red.), *Taking stock; the status of criminological theory*, Advances in criminological theory, 15e jrg., 2006, New Brunswick, NJ, Transaction (in druk)
- Farrington, D.P., D. Jolliffe**
A feasibility study into using a randomized controlled trial to evaluate treatment pilots at HMP Whitemoor
 London, Home Office, 2002 (Online report 14/02: www.homeoffice.gov.uk/rds/pdfs2/rdsolr1402.pdf).
- Farrington, D.P., B.C. Welsh (red.)**
What works in preventing crime? Systematic reviews of experimental and quasi-experimental research
 Annals of the American Academy of Political and Social Science, 578, 2001
- Farrington, D.P., B.C. Welsh**
Family-based prevention of offending; a meta-analysis
 Australian and New Zealand journal of criminology, 36e jrg., 2003, p. 127-151
- Farrington, D.P., B.C. Welsh**
Randomized experiments in criminology; what have we learned in the last two decades?
 Journal of experimental criminology, 1e jrg., 2005, p. 9-38
- Farrington, D.P., B.C. Welsh**
A half-century of randomized experiments on crime and justice
 In: M. Tonry (red.), *Crime and justice*, 34e jrg., 2006, Chicago, University of Chicago Press (in druk)
- Feder, L.**
Using random assignment in social science settings
 Professional ethics report, 11e jrg., nr. 1, 1998, p. 1-7
- Feder, L., L. Dugan**
A test of the efficacy of court-mandated counselling for

- domestic violence offenders; the Broward experiment*
Justice quarterly, 19e jrg., 2002, p. 343-375
- Feder, L., D.B. Wilson**
Mandated batterer programs to reduce domestic violence
In: B.C. Welsh, D.P. Farrington (red.), *Preventing crime; what works for children, offenders, victims and places*, Dordrecht, Springer, 2006
- Finckenaue, J. O.**
Scared straight and the panacea phenomenon
Englewood Cliffs, NJ, Prentice-Hall, 1982
- Folkard, M.S., A.J. Fowles e.a.**
IMPACT (Intensive Matched Probation and After-Care Treatment): vol. 1; The design of the probation experiment and an interim evaluation
London, Her Majesty's Stationery Office, 1974
- Folkard, M.S., D.E. Smith**
IMPACT; vol. 2
London, Her Majesty's Stationery Office, 1976
- Garner, J., J. Fagan e.a.**
Published findings from the spouse assault replication program; a critical review
Journal of quantitative criminology, 11e jrg., 1995, p. 3-28
- Gottfredson, D.C., S.S. Najaka**
Effectiveness of drug treatment courts; evidence from a randomized trial
Criminology and public policy, 2e jrg., 2003, p. 171-196
- Hindelang, M.J., T. Hirschi e.a.**
Measuring delinquency
Beverly Hills, CA, Sage, 1981
- Kassebaum, G., D. Ward e.a.**
Prison treatment and parole survival
New York, Wiley, 1971
- Killias, M., M. Aebi e.a.**
Does community service rehabilitate better than short-term imprisonment? Results of a controlled experiment
Howard journal, 39e jrg., 2000, p. 40-57
- Klein, M.W.**
Labelling theory and delinquency policy; an experimental test
Criminal justice and behavior, 13e jrg., 1986, p. 47-79
- Kling, J.R., J. Ludwig e.a.**
Neighborhood effects on crime for female and male youth; evidence from a randomized housing voucher experiment
Quarterly journal of economics, 120e jrg., 2005, p. 87-130
- Lerman, P.**
Community treatment and social control
Chicago, University of Chicago Press, 1975
- Lewis, R.V.**
Scared straight - California style; evaluation of the San Quentin Squires Program
Criminal justice and behavior, 10e jrg., 1983, p. 209-226

- Lipsey, M.W.**
Those confounded moderators in meta-analysis; good, bad, and ugly
Annals of the American Academy of Political and Social Science, 587, 2003, p. 69-81
- Lipsey, M.W., D.B. Wilson**
Practical meta-analysis
Thousand Oaks, CA, Sage, 2001
- McCord, J.**
A thirty-year follow-up report on the Cambridge-Somerville Youth Study
American psychologist, 33e jrg., 1978, p. 284-289
- MacKenzie, D.L., D.B. Wilson e.a.**
Effects of correctional boot camps on offending
Annals of the American Academy of Political and Social Science, 578, 2001, p. 126-143
- Maxwell, C.D., J.H. Garner e.a.**
The preventive effects of arrest on intimate partner violence; research, policy, and theory
Criminology and public policy, 2e jrg., 2002, p. 51-79
- Nuttall, C.**
The Home Office and random allocation experiments
Evaluation review, 27e jrg., 2003, p. 267-289
- Olds, D.L., C.R. Henderson e.a.**
Long-term effects of nurse home visitation on children's criminal and antisocial behavior; 15-year follow-up of a randomized controlled trial
Journal of the American Medical Association, 280e jrg., 1998, p. 1238-1244
- Ortmann, R.**
The effectiveness of social therapy in prison; a randomized experiment
Crime and delinquency, 46e jrg., 2000, p. 214-232
- Palmer, T.B.**
The Youth Authority's Community Treatment Project
Federal Probation, 38e jrg., nr. 1, 1974, p. 3-14
- Peters, M., D. Thomas e.a.**
Boot camps for juvenile offenders
Washington, DC, Office of Juvenile Justice and Delinquency Prevention (Program Summary), 1997
- Petersilia, J., S. Turner**
Intensive probation and parole
In: M. Tonry (red.), *Crime and justice*, 17e jrg., 1993, p. 281-335, Chicago, University of Chicago Press
- Petrosino, A.J., C. Turpin-Petrosino e.a.**
Scared straight and other juvenile awareness programs for preventing juvenile delinquency; a systematic review of the randomized experimental evidence
Annals of the American Academy of Political and Social Science, 589, 2003, p. 41-62
- Powers, E., H. Witmer**
An experiment in the prevention of delinquency

- New York, Columbia University Press, 1951
- Reimer, E., M. Warren**
Special intensive parole unit
NPPA Journal, 3e jrg., 1957, p. 222-229
- Rose, G., R.A. Hamilton**
Effects of a juvenile liaison scheme
British journal of criminology, 10e jrg., 1970, p. 2-20
- Rossi, P.H., R. Berk e.a.**
Money, work, and crime; experimental evidence
New York, Academic Press, 1980
- Schweinhart, L.J., J. Montie**
Lifetime effects; the High/Scope Perry Preschool Study through age 40
Ypsilanti, MI, High/Scope Press, 2005
- Shadish, W.R., T.D. Cook**
Experimental and quasi-experimental designs for generalized causal inference, Boston, MA, Houghton Mifflin, 2002
- Shaw, M.**
Social work in prison
London, Her Majesty's Stationery Office, 1974
- Sherman, L.W.**
Policing domestic violence; experiments and dilemmas
New York, Free Press, 1992
- Sherman, L.W., R.A. Berk**
The specific deterrent effects of arrest for domestic assault
American sociological review, 49e jrg., 1984, p. 261-272
- Sherman, L.W., E.G. Cohn**
The impact of research on legal policy; the Minneapolis domestic violence experiment
Law and society review, 23e jrg., 1989, p. 117-144
- Sherman, L.W., D. Weisburd**
General deterrent effects of police patrol in crime 'hot spots'; a randomized controlled trial
Justice quarterly, 12e jrg., 1995, p. 625-648
- Short, J.F., M. Zahn e.a.**
Experimental research in criminal justice settings; is there a role for scholarly societies?
Crime and delinquency, 46e jrg., 2000, p. 295-298
- Strang, H., L.W. Sherman**
Restorative justice to reduce victimization
In: B.C. Welsh, D.P. Farrington (red.), *Preventing crime; what works for children, offenders, victims and places*, Dordrecht, Springer, 2006
- Tong, L.S.J., D.P. Farrington**
How effective is the 'Reasoning and Rehabilitation' programme in reducing reoffending? A meta-analysis of evaluations in four countries
Psychology, crime and law, 2005 (in druk)
- Voorhis, P. van, L.M. Spruance e.a.**
The Georgia Cognitive Skills Experiment; a replication of 'Reasoning and Rehabilitation

- Criminal justice and behavior, 31e jrg., 2004, p. 282-305
- Wasserman, G.A., L.S. Miller**
The prevention of serious and violent juvenile offending
In: R. Loeber, D.P. Farrington (red.), *Serious and violent juvenile offenders; risk factors and successful interventions*, Thousand Oaks, CA: Sage, 1998, p. 197-247
- Weisburd, D.**
Ethical practice and evaluation of interventions in crime and justice; themoral imperative for randomized trials
Evaluation review, 27e jrg., 2003, p. 336-354
- Weisburd, D., C.M. Lum e.a.**
Does research design affect study outcomes in criminal justice?
Annals of the American Academy of Political and Social Science, 578, 2001, p. 50-70
- Weisburd, D., A. Petrosino**
Experiments, Criminology
In: K. Kempf-Leonard (red.), *Encyclopaedia of social measurement*, San Diego, CA, Academic Press, 2004, p. 877-884
- Welsh, B.C., D.P. Farrington (red.)**
Preventing crime; what works for children, offenders, victims, and places
Dordrecht, Springer, 2006
- Wilson, D.B., L.C. Allen e.a.**
Quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders
Criminal justice and behavior, 32e jrg., 2004, p. 172-204
- Wilson, D.B., O. Mitchell e.a.**
A systematic review of drug court effects on recidivism
Manuscript wordt gereviewed, 2005
- Wortman, C.B., V.C. Rabinowitz**
Random assignment; the fairest of them all
In: L. Sechrest, S.G. West e.a. (red.), *Evaluation Studies Review Annual*, Beverly Hills, CA, Sage, 4e jrg., 1979

Evaluatie van complexe programma's

Een theoriegestuurde aanpak

*R. Pawson en C. Klein Haarhuis**

In dit nummer wordt regelmatig ingegaan op het belang van quasi-experimenteel onderzoek voor het verkrijgen van inzicht in de effecten van beleidsinterventies. Bij experimenteel evaluatieonderzoek ligt de nadruk op een zo groot mogelijke interne validiteit. Met andere woorden, met het onderzoeksdesign moet de zekerheid dat een uitkomst aan beleid kan worden toegeschreven, zo groot mogelijk worden gemaakt. Aan de verdiensten van deze benadering valt niets af te doen: andere invloeden worden zoveel mogelijk uitgeschakeld en eventuele effecten van beleid worden zo 'hard gemaakt'. Een potentiële blinde vlek van zulke experimentele designs is echter dat beleidsprogramma's (hierna: programma's) worden gezien als afgebakende en op zichzelf staande *interventies* op een bepaalde tijd op een bepaalde plaats, en minder als 'complexen' van regels, bevoegdheden of 'gedragsprikkelers'. Met een experimenteel design is dan ook niet na te gaan wát het nu is in een programma dat de gevonden uitkomsten veroorzaakt, en *hoe*, via welke processen, dit gebeurt. In deze bijdrage wordt een theoriegestuurde evaluatiebenadering gepresenteerd die zich richt op de innerlijke werking (het *hoe en waarom*) van programma's en op de tijdige aanpassing ervan. Programma's werken niet in isolatie: het zijn open, veranderlijke complexen van interventies. Daarom dient een evaluator aandacht te besteden aan de 'theorieën' van programmamakers: waarom denken zij dat hun programma zal werken? De basislogica voor theoriegestuurde evaluatie is heel simpel. Programma's zijn theorieën. Daaruit volgt dat evaluatie het testen van theorie is: welk onderdeel van een programma veroorzaakt welke uitkomsten en hoe? Deze benadering zal worden geïllustreerd aan de hand

* Dr. Ray Pawson is als Reader verbonden aan de Universiteit van Leeds. Dr. Carolien Klein Haarhuis is junior-onderzoeker bij het WODC.

van *Megan's Law*, die burgergemeenschappen in de VS actief moet betrekken bij het toezicht houden op potentiële seksdelinquenten en het afstraffen van ongewenst gedrag.

Programma's zijn theorieën

Programma's zijn zoals gezegd geen relatief simpele 'stimuli' zoals in het geval van laboratoriumexperimenten, maar complexe systemen van interventies die in continue interactie staan met hun omgeving. De gedachtegangen achter programma's zijn dan ook vaak complex en meerlagig. Er is sprake van lange implementatieketens en vele betrokkenen. De programma's kunnen alleen effect sorteren als die mensen of groepen mensen (organisaties) *actief* betrokken zijn. Vergelijk de aanleg van een dijk (de dijk werkt als waterkering of niet, ongeacht het gedrag van de burger) met een campagne die tweemaal daags tanden poetsen moet bevorderen. Die publiciteit kan wel of niet worden geaccepteerd door de doelgroep, laat staan met het gewenste gedrag worden nageleefd. Actieve programma's werken dus in de eerste plaats via de denkwijze en het gedrag van de aangesprokene.

Programma's zijn dus in de eerste plaats *theorieën*. Deze beginnen in de hoofden van beleidsontwerpers, belanden in de handen van uitvoerders, en soms in de hoofden van de doelgroepen. Alle programma's, ongeacht de inhoud, hebben één ding gemeen: de kernhypothese is altijd dat als wij deze mensen deze middelen aanreiken, hun gedrag kan veranderen.

Voor de sceptici onder ons willen we iets dieper ingaan op een geliefd stokpaardje: de 'Lekkere-David-Beckham'-hypothese. Voor wie niet van voetbal houdt: Beckham heeft samen met zijn vrouw een status als icoon in de wereld van 'beroemdheden' verworven. Zozeer zelfs dat hij de inspiratie was voor een recent gehouden opvoedingsprogramma over gezondheid. De achterliggende redenering was dat tienermeisjes hun tijd grotendeels zonder lichaamsbeweging doorbrengen en dat hun levensstijl wordt beïnvloed door de tijdschriften die zij lezen. Wellicht zou het mogelijk zijn hun houding ten opzichte van fitness en gezondheid te beïnvloeden door de uitgevers van deze tijdschriften te stimuleren andere rolmodellen te nemen. Zo begon Beckhams carrière als fotomodel en zag de theorie het licht. Het uiteindelijke resultaat blijkt echter duidelijk uit

het volgende stukje uit een evaluatieverslag (Mitchell, 1997):

Interviewer: Maar denk je dat het feit dat deze mooie mannen voetballers zijn enige invloed heeft op de houding van meisjes ten opzichte van voetballen?

Meisje: Nee, ik denk dat het meer van invloed is op het *kijken naar* voetbal, nou ja, niet naar het voetbal, maar naar de mannen (gelach en instemmende geluiden).

De meisjes, zo lijkt het, kunnen alleen maar lachen om de theorie van het programma en laten liever hun hersens dan hun lichamen werken. Trouwens, Beckham is als vader van twee kinderen inmiddels een opvoedkundige programmatheorie, met de naam 'lads-make-good-dads', oftewel 'jongens zijn goede vaders'.

Evaluatie is het testen van programmatheorieën

Als programma's theorieën zijn, is evaluatie dus een proces om die programmatheorieën te testen. Een van de grotere onderzoeksapproches die hieruit is voortgekomen, is de strategie van de 'verandertheorie' (Connell e.a., 1995). Het idee hierachter is dat de programma's iteratieve theoriereeksen zijn: als we A invoeren, dan moet dit leiden tot ons eerste interventiedoel B, en zodra B is geïmplementeerd, zijn we op de juiste plek voor C, waardoor ook resultaat D mogelijk wordt, enzovoort. De belangrijkste belanghebbenden worden geraadpleegd over de werktheorieën van het initiatief, en de beoogde 'tussenschappen' worden blootgelegd en uiteengezet.

Realistische Evaluatie: Context-Mechanisme-Uitkomstpatroon

De werking van een programma kan verder worden uitgewerkt aan de hand van de Context-Mechanismen-Outcome ('C-M-O')-configuraties (Pawson en Tilley, 1997).

De processen of 'tussenschappen' komen overeen met de 'mechanismen' in de zogenoemde 'realistische evaluatie'-benadering. Een van de kernelementen van deze specifieke vorm van theoriegestuurde evaluatie is om oog te hebben voor die mechanismen, en deze te articuleren. Bij mechanismen gaat het om

verschillende mogelijke determinanten van gedrag. Mechanismen maken dat 'dingen werken' en kunnen daarom beschouwd worden als de motoren achter beleidsinterventies. Ook de werking van Megan's Law in de VS stoelt op mechanismen. Met deze wet wordt gepoogd om burgergemeenschappen in de VS te betrekken bij het observeren en afstraffen van het gedrag van seksdelinquenten. Dit vergt besluiten van uitvoerders van de wet over wie de risicogevoelers zijn, welke informatie over deze personen wanneer en hoe vaak zou moeten worden opgeslagen, wiens identiteit naar buiten wordt gebracht en naar wie toe, hoe de reacties van de gemeenschap te controleren en gemeenschapstoezicht te stimuleren, enzovoort. De effectiviteit van een dergelijk programma als geheel zal afhangen van de gecombineerde werkzaamheid van dit soort 'tussenstappen' in de programmatheorie.

De werking van mechanismen is afhankelijk van *context*.

Elke verandering in gedrag, gebeurtenissen en maatschappelijke omstandigheden komt tot stand via het gehele systeem van sociale relaties. Naast oog voor mechanismen heeft een 'realist' – evaluator daarom ook oog voor de verschillende aspecten van de context waarbinnen een programma plaatsvindt. Een voorbeeld van een contextueel aspect dat de werking van Megan's Law kan beïnvloeden is de mate van sociale cohesie en betrokkenheid in de burgergemeenschappen die het gedrag van seksdelinquenten moeten helpen beheersen.

'*Uitkomstpatronen*' omvatten de bedoelde en onbedoelde gevolgen van programma's, die volgen uit de activering van verschillende mechanismen in verschillende contexten. Deze uitkomstpatronen zijn er in vele soorten en maten. Voor evaluatoren van programma's is het zaak om de meest adequate uitkomstindicatoren te kiezen. Zo zijn recidivecijfers weliswaar relevant voor het beoordelen van de effecten van Megan's Law, maar bleek de snelheid van arrestatie van voormalige gevangenen een even belangrijke uitkomstindicator. Gezamenlijk resulteren de vier concepten in '*context-mechanisme-uitkomstpatroon-configuraties*' (CMOC's). Deze omvatten modellen die aangeven hoe programma's mechanismen activeren, bij wie en onder welke omstandigheden, en die zo resulteren in uitkomsten: veranderingen in gedragpatronen, of patronen van gebeurtenissen. Met deze modellen kunnen uiteindelijk verklaringen worden geboden voor verschillen in programma-uitkomsten tussen mensen, tussen groepen, en tussen contexten.

Evaluatie bestaat uit het in een vroeg stadium maken van proces-observaties om te kijken of de theorieën voldoen aan de realiteit. Is het resultaat zoals verwacht, dan kunnen we bijna met zekerheid stellen dat het programma hiervoor verantwoordelijk was. We hebben immers gekeken naar de 'binnenwerking'. Als wij in een bepaalde programmatheorie een fout ontdekken, kan deze worden geïdentificeerd, worden verbeterd en opnieuw worden getest op zijn vermogen het programma verder te brengen.

Tegenwoordig wordt veel minder tijd, moeite en geld besteed aan de 'live' evaluaties van programma's. In plaats daarvan is er veel meer aandacht voor het onderzoeken van reeds beschikbaar bewijs. De redenering hiervoor ligt voor de hand. Programma's zijn keer op keer uitgevoerd en keer op keer 'geresearched' en nu is het de hoogste tijd om voor eens en altijd te bepalen wat ze hebben opgeleverd. Echter, de bestaande kennis over de praktijk van een programma is over het algemeen veelsoortig en gefragmenteerd. Bovendien is onderzoeksmateriaal waarin een (quasi-)experimentele opzet wordt gehanteerd, schaars of ontbreekt zelfs. Dit betekent dat de effecten van een programma dus zelden 'hard' zijn te maken. Daar komt bij dat we inzicht willen krijgen in de werking van de programmatheorie: welk onderdeel van een programma veroorzaakt welke uitkomsten en hoe?

Om die reden is het van belang om zoveel mogelijk gebruik te maken van *andere* bronnen van informatie, die ons iets leren over de (soms vele parallele of samenhangende) processen waarlangs beleid zijn uitwerking heeft. Te denken valt aan beleidsdocumenten, verslagen van (diepte-) interviews, notulen van vergaderingen, observaties van publieke bijeenkomsten, dossiers van de doelgroep (bijvoorbeeld delinquentenprofielen), enzovoort.

In het vervolg van dit artikel illustreren we het nut van de evaluatiestrategie van de 'verandertheorie' aan de hand van het voorbeeld van Megan's Law in de VS.

Megan's Law: wat vertelt de theorie ons?

Na de moord op Megan Kanka door een seksdelinquent die na het uitzitten van zijn gevangenisstraf anoniem in haar buurt woonde, werd het Amerikaanse seksdelinquenten registratie- en notificatieprogramma, in het kort Megan's Law, versneld in de wetboeken opgeno-

men. Het initiatief gaf uitdrukking aan het ongenoegen dat er onder de bevolking leefde en dat zich later in Groot-Brittannië herhaalde na vergelijkbare drama's, met name de moord op Sarah Payne. Een evaluatie van de doeltreffendheid van Megan's Law volgde echter pas mondjesmaat na de opname in het wetboek. De verschillende data waren niet vergelijkbaar qua dekking, onderzoeksmethode en kwaliteit van het onderzoek. Dus hoe was het mogelijk om aan de hand van een dergelijke ongefundeerde kennisbasis te bepalen of een beleidsverandering als deze de moeite waard is geweest? Een strategie voor systematisch onderzoek is gebaseerd op de methodologie van de 'verandertheorie', die hierboven kort is toegelicht. Deze methode is bij programma-evaluatie tot dusver beperkt gebleven tot het hier en nu. Kan de klok worden teruggedraaid door het retrospectief toepassen van de verandertheorie-strategie bij het onderzoeken van *eerder gevonden* bewijs?

Het achterhalen van de programmatheorie

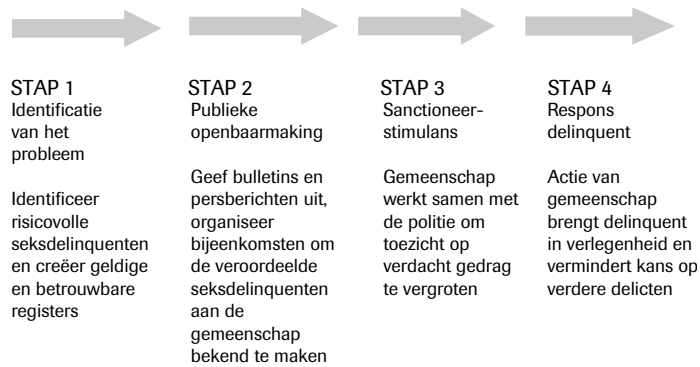
De strategie van de verandertheorie biedt een uitweg door het moeras aan tegenstrijdig en variabel bewijs over Megan's Law. Normaal gesproken is de eerste fase van de methode het blootleggen en verbinden van de 'live' programmatheorie door middel van diepgaande discussies met de belanghebbenden.

In de toepassing van de strategie op basis van *bestaand bewijs* zijn de beoogde werkingsmechanismen in Megan's Law gereconstrueerd aan de hand van officiële, administratieve en wetgevende documenten. Meer specifiek zijn de volgende 'bronnen' van informatie gebruikt:

- vele publieke commentaren;
- wettelijk en grondwettelijk materiaal;
- ambtelijke rapporten;
- waarnemingen van de publieke opinie;
- studie, observatie van publieke vergaderingen;
- simulaties van profielen van delinquenten;
- studies van de reactie van potentiële overtreeders op correcties door de gemeenschap;
- diepte-interviews met ex-delinquenten;
- inhoudsanalyse van rapportage in de media;
- quasi-experimenteel onderzoek.

Figuur 1 is een vereenvoudigde versie van de programmatheorie. Hierin wordt een reeks van vier tussenstappen geïdentificeerd die doorlopen moeten worden om op de plek van bestemming te komen, namelijk de handhaving binnen de leefgemeenschap van seksdelinquenten die hun gevangenisstraf hebben uitgezeten.

Figuur 1: Het beoogde proces voor Megan's Law



In het onderzoek werden deze deeltheorieën stuk voor stuk getest, waarbij honderden bestaande studies werden doorgenomen om een goed onderbouwde beslissing te kunnen nemen of de wet werkte. De drie cases hieronder geven een indicatie van de kracht van een theoriegestuurde aanpak in synthetiserend onderzoek.

Een quasi-experimenteel onderzoek

Een voor de hand liggend uitgangspunt voor een dergelijk onderzoek is kijken of de interventie inderdaad tot het gewenste resultaat heeft geleid. Er bestaan echter slechts twee studies die hebben getracht het effect van de introductie van Megan's Law vast te stellen aan de hand van de hoeveelheid seksdelinquenten die opnieuw in de fout gaan. Helaas zijn de uitkomsten van de twee studies ongeveer vergelijkbaar en niet overtuigend. We concentreren ons hier op het quasi-experimentele onderzoek van Schram and Milloy (1995) in Washington. De uitkomsten in het kort:

- Aan het eind van de 54 maanden in de gemeenschap, kende de notificatiegroep slechts een iets lagere seksuele recidive (19%) dan

de vergelijkingsgroep (22%). Gezien de geringe aantallen werd dit verschil niet als statistisch afwijkend gezien.

- Ofschoon er geen significante verschillen waren in de algemene recidive, was er wel een duidelijk verschil in het tijdstip waarop de nieuwe arrestatie van de 'voor'- en 'na'-groepen plaatsvonden. Voormalige gevangenen waarop de gemeenschapsnotificatie van toepassing was, werden veel sneller voor nieuwe seksmisdaden gearresteerd dan mensen die zonder waarschuwing waren vrijgelaten.

Het is echter de conclusie bij het rapport dat boekdelen spreekt. De auteurs hechten veel belang aan de experimentele aanpak, omdat zij die als de enige manier zien om een duidelijke indruk van het programma-effect te krijgen. Hun teleurstelling over de non-significantie van het resultaat over nieuwe arrestaties is duidelijk voelbaar. De bevindingen over de nieuwe arrestaties worden bijna terloops gegeven, min of meer als goedmakertje. Gezien vanuit de theorieopbouw, geeft het verschil in de twee uitkomsten echter een verrassend inzicht in wat nu precies het effect van de interventie is: Megan's Law bevordert wellicht detectie (arrestatie), en niet zozeer protectie (voorkomen van recidive).

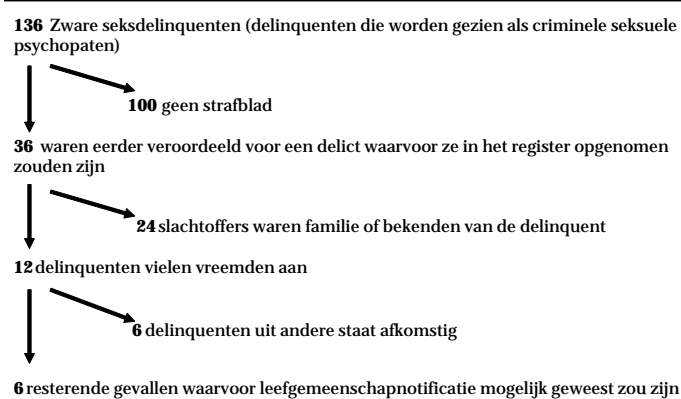
Een prospectieve simulatie

De volgende studie door Petrosino en Petrosino (1999) biedt een veel dieper inzicht in de lange implementatieketen zoals uiteengezet in figuur 2. De auteurs hebben zichzelf de moeilijke taak gesteld te kijken in hoeverre Megan's Law het publiek in staat stelt zichzelf te verdedigen tegen aanvallen als deze. Door het inlichten van de gemeenschap hoopte men vooral seksuele geweldsmisdrijven te voorkomen waarbij dader en slachtoffer elkaar niet kennen. Deze zijn relatief zeldzaam en natuurlijk moeilijk te voorspellen. Het onderzoek poogt dus een antwoord te vinden op de vraag: 'in welk percentage seksuele misdrijven biedt notificatie het slachtoffer (of zijn of haar familie of leefomgeving) een reële kans de dreiging te herkennen en daarmee te vermijden of te voorkomen?'

Naar deze vraag werd onderzoek gedaan in Massachusetts, de laatste van alle staten waar Megan's Law in het wetboek werd opgenomen. Op het moment van het onderzoek beschikten de onderzoekers niet over recente registraties waarmee ze aan de slag konden. Het was een obstakel waarvoor zij een ingenieuze oplossing bedach-

ten, namelijk door vanuit een recente set delicten vooruit te werken, waarbij ze keken naar het aantal delinquenten dat onder toezicht zou hebben gestaan, als de wet reeds in werking was getreden. Hun prognose is samengevat in figuur 2.

Figuur 2: De afnemende 'target' van Megan's Law in Massachusetts



Van de 136 delinquenten waren er slechts 36 veroordeeld voor een eerder delict waarvoor ze anders in aanmerking zouden zijn gekomen voor het notificatieproces. Hiervan hadden er 12 een vreemde aangevallen. 24 slachtoffers waren familie, vrienden of collega's van de delinquent. Bedenk ook dat notificatie weinig zin heeft als beschermingsmiddel voor mensen in de nabije omgeving van de delinquent. Zij zijn immers waarschijnlijk reeds op de hoogte van eerdere veroordelingen.

De volgende stap bestond uit het verzamelen van de gegevens over de twaalf delicten waarbij de delinquent vreemden had aangevallen. Zo kon er een inschatting worden gemaakt van de mogelijkheid om een proactieve, gelokaliseerde waarschuwing uit te laten gaan naar potentiële slachtoffers, waarmee het slachtoffer zich zou kunnen verdedigen. In zes gevallen leek het heel onwaarschijnlijk dat het slachtoffer door notificatie zou zijn gewaarschuwd of voorbereid zou zijn geweest, omdat de delinquenten van buiten de staat afkomstig waren. De gesimuleerde notificatieketen eindigt dus met zes slachtoffers die een realistische kans zouden hebben gehad op de waarschuwingen te reageren.

Deze studie is van groot belang voor de verklarende syntheses. Zelfs al zijn Petrosino en Petrosino zelfs maar bij benadering juist in hun onvolledig onderbouwde schatting van het aantal potentiële delinquenten dat daadwerkelijk onder toezicht in de leefgemeenschap wordt geplaatst, dan nog verklaart dit veel over het nulresultaat uit de recidivestudie van Schram en Milloy. De reden dat de introductie van de wet weinig effect lijkt te hebben gehad op recidive is mogelijk dat leefgemeenschapsurveillance misschien een zwak wapen is in de strijd tegen een zeldzaam delict waarvan de daders wellicht onvindbaar blijven.

Een onderzoek onder mensen uit de praktijk

De laatste primaire studie is het onderzoek van Zevitz en Farkas (2000) naar proefverlof- en reclasseringspersoneel in Wisconsin dat dagelijks is belast met het houden van toezicht op seksdelinquenten. In de studie worden hun ervaringen met Megan's Law besproken, met name hoe de *Special Bulletin Notificatie-cases* (SBNs), de risicovolle seksdelinquenten, bijdroegen aan hun caseload. Er werd gekozen voor het houden van een enquête met zowel gesloten als open vragen. Het verslag bevat dus een grote hoeveelheid kwantitatieve informatie over veranderingen in personeelsproblemen, trainingseisen en werkbelasting. Toch maakt het rapport voornamelijk gebruik van kwalitatieve methodes voor de registratie van de inzichten van de mensen uit de praktijk, en het is dit kwalitatieve bewijs dat hier van belang is.

Megan's Law is volgens de reclasseringsbeambte, een 'mandaat zonder budget'. Door de nieuwe opdracht verschoof de nadruk in hun dagelijkse werk naar SBN-cases. Zo werd er bijvoorbeeld door een respondent gesteld: 'Ik denk niet dat het management begrip heeft voor de enorme hoeveelheid bijkomstige contacten die nodig zijn voor het behandelen van een seksdelinquent: familie van de delinquent, de familie van het slachtoffer, de openbaar aanklager, de klinisch arts of psycholoog, werkgever, enzovoort'. Misschien verwijzen de meest significante opmerkingen over de caseload (zowel wat betreft de omvang als het gevoel) naar de problemen die het beheren van de reacties vanuit de leefgemeenschap met zich meebrengen. Kleine plagerijtjes zijn een dagelijks feit, huisvestingsproblemen blijken veelvoorkomend, en soms worden mensen zelfs met de dood bedreigd. Zoals een reclasseringsbeambte het spottend

verwoordt: 'de druk om meer tijd aan SBN-cases te besteden (te babysitten) is groter, alleen omdat het SBN-gevallen zijn.' Het biedt een goed inzicht in fase 3 van de Megan's Law programmatheorie. De bedoeling was om de leefgemeenschap met de overheidsinstellingen te laten samenwerken om toezicht te houden op delinquenten die na een gevangenisstraf zijn vrijgelaten. Het lijkt er echter op dat mensen uit de praktijk hun tijd vooral besteden aan het beschermen van de delinquenten tegen de ruigere aspecten van de aandacht die ze krijgen van de leefgemeenschap. Als Zevitz en Farkas ook maar een beetje gelijk hebben in het beeld dat zij schetsen van de toegenomen aandacht voor SBN-gevallen, bieden zij de perfecte verklaring voor de grote toename in arrestatiecijfers in Schram en Milloy's quasi-experiment: geregistreerde seksdelinquenten lopen een grotere kans opnieuw gearresteerd te worden vanwege significante 'bijkomende' contacten met ordehandhavingpersoneel.

Conclusie

Uit deze korte voorbeelden van een verandertheorie-onderzoek blijkt dat er waardevol bewijs gehaald kan worden uit materiaal dat in eerste instantie niet veelbelovend lijkt en dat methodologisch gezien beneden de maat is. Iedere studie helpt de bevindingen van andere te verduidelijken. Een ongewijzigd recidivepatroon kan bijvoorbeeld worden verklaard aan de hand van de extreem weinige mogelijkheden die buurttoezicht na notificatie biedt. Een verbetering in detectiecijfers is wellicht mogelijk vanuit de veranderingen in de werkbelasting van de reclasseringsbeambte. Om hiervan zeker te zijn zou er meer bewijs moeten worden verzameld om deze claims te onderbouwen en dient er een vollediger overzicht (Pawson, 2002) gemaakt te worden.

Het belangrijkste methodologische punt dat hieruit naar voren komt, is dat verklarende vooruitgang wordt geboekt omdat een *programmatheorie* eerst duidelijk uiteen wordt gezet alvorens het onderzoek wordt uitgevoerd. Met behulp van de verandertheorie-structuur weten we waar en wat we zoeken dat als bewijs kan dienen.

Zonder deze ruggesteun resulteert onderzoek meestal slechts in verwarring, in het aloude liedje dat we 'simpelweg onvoldoende weten' en dat er dus meer onderzoek gedaan moet worden (Lovell, 2001).

Literatuur

Connell, J. A. Kubish e.a.

New approaches to evaluating community initiatives

Washington, DC, Aspen

Institute, 1995

Lovell, T.

Megan's Law: does it protect children? A review of evidence on the impact of community notification as legislated for through Megan's Law in the United States: recommendations for policy makers in the United Kingdom

London, NSPCC, 2001

Mitchell, K.

Encouraging young women to exercise; can teenage magazines play a role?

Health education journal, 56^e jrg., nr. 2, 1997, p. 264-273

Pawson, R., N. Tilley

Realistic evaluation

Londen/Thousand Oaks/New Delhi, Sage Publications, 1997

Pawson, R.

Does Megan's Law work? A theory-driven systematic review

London, ESRC Centre for Evidence Based Policy and Practice, Working Paper 8, 2002

Petrosino, A., C. Petrosino

The public safety potential of Megan's law in Massachusetts; an assessment from a sample of criminal sexual psychopaths

Crime and delinquency, 45^e jrg., nr. 1, 1999, p. 140-158

Schram, D., C. Milloy

Community notification; a study of offender characteristics and recidivism

Olympia, WA, Washington State Institute for Public Policy, 1995

(www.wa.gov/wsipp/crime/pdf/chrrec.pdf)

Zevitz, R., M. Farkas

The impact of sexoffender community notification on probation/parole in Wisconsin

International journal of offender therapy and comparative criminology, 44^e jrg., nr. 1, 2000, p. 8-21

Rechtshandhaving geëvalueerd

Een synthese van 31 beleidsevaluaties

*C. Klein Haarhuis, M. van Ooyen-Houben, E. Kleemans en F. Leeuw**

In de laatste decennia is het aantal beleidsonderzoeken, prestatie-metingen en evaluatieonderzoeken sterk in omvang toegenomen. Cijfers laten zien dat overheden en non-gouvernementele organisaties tegenwoordig vele honderden beleids- en evaluatieonderzoeken per jaar verrichten (Leeuw en Cooksey, 2004; Rist en Stame, 2005). Daarbij komen nog de 'monitoren' en 'indicatoren' die in toenemende mate worden ontwikkeld ten behoeve van de beleidsbepaling en -uitvoering (Zall Kusek en Rist, 2004).

Bij een dergelijke overvloed aan onderzoek doet zich de vraag voor of het niet mogelijk is om al deze onderzoeksresultaten te synthetiseren tot een overzichtelijk geheel. Dat dergelijke syntheses wenselijk zijn, wordt door bijna niemand betwist. Maar over hoe dit zou moeten gebeuren, lopen de meningen uiteen.

De ene stroming, die nauw verbonden is met de Campbell Collaboration Group on Crime and Justice, probeert overzicht te creëren door hoge eisen te stellen aan de kwaliteit van onderzoeksopzetten. Daarbij wordt sterk de nadruk gelegd op het belang van (quasi)-experimentele designs voor het valide en betrouwbaar meten van effecten. Of een interventie (een beleidsmaatregel) wel of niet werkt, dient in het ideale geval te worden vastgesteld door 'at random' een experimentele groep en een controlegroep samen te stellen, door vervolgens de experimentele groep te confronteren met de interventie, en door ten slotte de resultaten tussen de experimentele groep en de controlegroep met elkaar te vergelijken. De basisgedachte binnen deze stroming is dat onderzoeken die niet aan bepaalde eisen voor (quasi)-experimenteel onderzoek voldoen, gevoeglijk buiten beschouwing kunnen worden gelaten (onder anderen Farrington en Welsh, 2005).

* Dr. Carolien Klein Haarhuis, dr. Marianne van Ooyen-Houben, dr. Edward Kleemans en prof. dr. Frans Leeuw zijn allen werkzaam bij het WODC. Dit artikel is gebaseerd op het WODC-rapport *Kennis voor beleid: een synthese van 58 onderzoeken op het gebied van rechtshandhaving*, door dezelfde auteurs (Den Haag, WODC, Cahier 2005-5).

De andere stroming, waarvan Pawson (2002) één van de voortrekkers is, stelt dat bij een dergelijke benadering veel te veel bruikbare kennis buiten beschouwing wordt gelaten en dat syntheses op een bredere basis gebaseerd moeten zijn, afhankelijk van de aard van de vraagstelling én onder uitdrukkelijke bekendmaking van de aard en wijze van dataverzameling en -analyse. Pawson pleit voor een zogenoemde 'realist synthesis', omdat kennis over *hoe* interventies werken (werkzame mechanismen) minstens zo belangrijk is voor beleidsmakers als kennis over *of* interventies werken (waarbij de interventie verder een 'black box' blijft). Bovendien is deze kennis over werkzame mechanismen (binnen bepaalde contexten) veel breder toepasbaar, ook bij andere interventies en op andere beleidsterreinen.

In dit artikel willen we heel concreet aangeven hoe het maken van een dergelijke synthese in zijn werk gaat en wat de resultaten daarvan kunnen zijn. Daarbij richten wij ons op de resultaten van 31 beleidsevaluaties, die tussen januari 2002 en mei 2004 in opdracht van het Ministerie van Justitie zijn uitgevoerd op het gebied van rechtshandhaving. Met betrekking tot deze 'dwarsdoorsnede' aan evaluatieonderzoeken beantwoorden we de volgende drie vragen:

- Om wat voor typen interventies (beleidsmaatregelen) gaat het eigenlijk bij deze onderzoeken?
- Op welke manier dient de kwaliteit van deze onderzoeken te worden beoordeeld en hoe bruikbaar zijn de resultaten voor een synthese?
- Hoe kunnen de resultaten uit deze onderzoeken inhoudelijk worden gesynthetiseerd? Met andere woorden: wat hebben we inhoudelijk geleerd van twee jaar evaluatieonderzoek op het gebied van rechtshandhaving?

Rechtshandhaving: vier typen interventies

In pleidooien voor (quasi)-experimenteel onderzoek wordt vaak een vergelijking getrokken met het testen van medicijnen. Slechts weinig mensen accepteren immers een medicijn zonder dat dit medicijn uitvoerig is getest in goed opgezet experimenteel onderzoek. Waarom zouden dan voor overheidsbeleid en onderzoek naar beleidsinterventies andere normen moeten gelden?

Voordat wij een genuanceerd antwoord geven op deze vraag, is het misschien goed om eerst na te gaan hoe die beleidsinterventies er

concreet uitzien. Relatief zelden is er sprake van maatregelen die rechtstreeks zijn gericht op verandering van het nalevingsgedrag van burgers. Veel vaker blijkt er sprake te zijn van:

- interventies in de vorm van speciale projecten en nieuwe organisatievormen;
- indirecte sturing door de financiering van organisaties;
- indirecte sturing van ketens van organisaties;
- stimulering van innovatie in (ketens van) organisaties;
- invoering of wijziging van wet- en regelgeving.

In het kader van deze synthese is daarom een typologie ontwikkeld van soorten (evaluatie)onderzoek, die recht doet aan de verschillende typen beleidsinterventies en daarmee ook aan de verschillende manieren waarop dergelijke maatregelen kunnen worden geëvalueerd. Daarbij onderscheiden wij de volgende vier typen evaluaties:

- *Projectevaluaties*: Onderzoek naar de uitvoering en resultaten van concrete projecten (bijvoorbeeld het instellen van een anoniem meldpunt voor misdrijven, of een project ter bestrijding van drugsgerelateerde overlast en criminaliteit door intensivering van handhavingsinspanningen).
- *Organisatie-evaluaties*: Onderzoek naar het functioneren van specifieke organisaties (bijvoorbeeld de rol van het Openbaar Ministerie in de aanpak van milieudelicten).
- *Ketenevaluaties*: Onderzoek naar het functioneren van de gehele handhavingsketen, of van aspecten/maatregelen die deze keten aangaan (bijvoorbeeld: het functioneren van de keten 'ongebruikelijke transacties' die is geënt op de melding, verwerking, opsporing en vervolging van praktijken die kunnen duiden op bijvoorbeeld het 'witwassen' van gelden).
- *Evaluatie van wet- en regelgeving*: Onderzoek naar de werking van nieuwe wet- en regelgeving, of van wijzigingen in wet- en regelgeving (bijvoorbeeld het opheffen van het bordeelverbod en het reguleren van de prostitutiesector, of een wetswijziging ter verruiming van de toepassingsmogelijkheden van DNA-onderzoek).

In tabel 1 zijn de 31 geselecteerde evaluatieonderzoeken ingedeeld volgens de bovenstaande systematiek. De tabel geeft goed zicht op de diversiteit van de onderzochte beleidsmaatregelen. Acht onderzoeken vallen in de categorie 'projectevaluaties'. Het gaat hierbij

Tabel 1: De geselecteerde onderzoeken ingedeeld naar type (evaluatie) onderzoek

Project-evaluaties	<p><i>Overzichtelijke interventies</i></p> <ul style="list-style-type: none"> * Project Meld Misdaad Anoniem (Blauw Research, 2003). * Evaluatie team opsporing oorlogsmisdrijven; het 'NOVO-team' (Beijer e.a., 2002). * Evaluatie Kernteams, opgericht voor de uitvoering van grotere en complexere opsporingsonderzoeken (Klerks e.a., 2002). <p><i>Multimodale interventies</i></p> <ul style="list-style-type: none"> * Integratie van Financieel Rechercheren in de handhavingsketen (Faber en Van Nunen, 2002). * Drugsoverlast in Venlo; indicatoren en nulmeting (Snippe en Bieleman, 2002). * Drugsoverlast in Venlo; tussenmeting (Snippe, Bieleman en Kruize, 2004). * Indicatoren voor het monitoren van de XTC-aanpak (Snippe, Oldersma, Bieleman, 2002). * Inspannen tegen XTC: eerste meting (Snippe, Kruize en Bieleman, 2003).
Organisatie-evaluaties	<ul style="list-style-type: none"> * Openbaar Ministerie en de aanpak van milieuzaken (DHV Milieu en infrastructuur, 2002).
Keten-evaluaties	<p><i>Beschrijving huidige situatie – breed</i></p> <ul style="list-style-type: none"> * Quick scan van de handhaving in het landelijk gebied (Bos e.a., 2001). * Staat van de handhaving door gemeenten (Andersson Elffers Felix, 2002). * Handhaving afvalstoffenregeling (Frederiks e.a., 2003). <p><i>Beschrijving inefficiënties</i></p> <ul style="list-style-type: none"> * Naar redelijke doorlooptijden in de strafrechtsketen (Berenschot, 2002). <p><i>Beschrijving nieuwe vormen (eventueel pilot)</i></p> <ul style="list-style-type: none"> * Integrale handhaving van het beleid ten aanzien van openbare inrichtingen (Bekkers e.a., 2003). * Evaluatie van de keten Ongebruikelijke Transacties (Faber en Van Nunen, 2004). * Handhaving leerplichtwet via lokale partners (BBSO, 2003).
Evaluatie van wet- en regelgeving	<ul style="list-style-type: none"> * (<i>In totaal 7 studies</i>) Gevolgen van de opheffing van het bordeelverbod (6 deelonderzoeken + synthese) (Daalder, 2002). * Invoering van het Transactiebesluit Milieudelicten (Blomberg e.a., 2002). * Evaluatie van de Wet Bijzondere Opsporingsbevoegdheden (eerste fase) (Bokhorst e.a., 2002). * Evaluatie van de Wet Herziening Gerechtelijk Vooronderzoek (Verrest en Beenackers, 2002). * Handhaafbaarheid van de Meststoffenwet (Prinsen e.a., 2002). * Naleving van de Wet verontreiniging oppervlaktewateren (Prinsen en Vossen, 2002). * DNA-onderzoek in opsporing en bewijsvoering in strafzaken, nulmeting (Buiter, Dubelaar, Haesen e.a., 2003). * Doping en handel. Onderzoek naar aard en omvang van dopinghandel en ontwikkeling van indicatoren (Oldersma, Snippe en Bieleman, 2002). * Gebruik en aanschaf van cannabis door 16- en 17-jarigen (Van Ooyen-Houben en Siepermann, 2002).

concreet om zes projecten die zijn gelanceerd door Nederlandse organisaties op het gebied van rechtshandhaving. Eén onderzoek betreft een zogenoemde organisatie-evaluatie. Zeven evaluaties betreffen de handhavingsketen als geheel, soms toegespitst op deelaspecten van de handhaving. Vijftien evaluaties slaan terug op de invoering of wijziging van wet- en regelgeving. Opmerkelijk is dat het bij de evaluatieonderzoeken steeds gaat om ex-post evaluaties; ex-ante evaluaties zijn niet aangetroffen.

Kwaliteit van de onderzoeken en bruikbaarheid voor een synthese

Voordat de inhoudelijke synthese heeft plaatsgevonden, hebben we eerst de kwaliteit en bruikbaarheid van de onderzoeken beoordeeld. Dit is gedaan aan de hand van de volgende vier beproefde kwaliteitscriteria.

In de eerste plaats is de onderzoekskwaliteit beoordeeld aan de hand van het meer algemene criterium van 'descriptieve validiteit' (Wittebrood en Van Beem, 2004). Gefocust is op de beschrijving van de onderzoeksopzet, de meetmethode en van de eventuele selectiemethode. Ook is nagegaan of meerdere informatiebronnen zijn gebruikt, omdat dit de betrouwbaarheid van de onderzoeksbevindingen verhoogt (Yin, 1994).

In de tweede plaats hebben we beoordeeld of in een onderzoek is nagegaan of en in welke mate een beleid daadwerkelijk is geïmplementeerd. Want beleid dat nooit is uitgevoerd, kan ook niet zinvol worden geëvalueerd. Eerder onderzoek laat zien dat de programma-integriteit, de daadwerkelijke implementatie van een beleid conform plan, vaak te wensen overlaat (Lösel, 1995). Ook een recent rapport van de Algemene Rekenkamer (2003) toont aan dat er veel meer beleidsmaatregelen worden gemaakt dan er daadwerkelijk worden uitgevoerd.

In de derde plaats is geanalyseerd of in de onderzoeken aandacht wordt besteed aan de *veronderstellingen* die ten grondslag liggen aan beleidsinterventies (onder anderen Leeuw, 2003; Pawson, 2002). Bij het ontwerpen en ontwikkelen van beleid worden reacties op interventies verwacht van de doelgroep (burgers, [potentiële] criminelen, politieagenten enzovoort). Beleidsmakers veronderstellen verder ook gedrag van derden (bijvoorbeeld 'buren', advocaten) en ze veronderstellen effecten. Van evaluatoren wordt verwacht dat

zij de assumpties over hoe beleid zou moeten werken, articuleren (de 'programmatheorie'), en deze vervolgens testen. *Hoe* (via welke mechanismen) wordt beleid verondersteld te werken, en tot welke resultaten zou dit moeten leiden? Pawson en Tilley (1997) houden een pleidooi om beleidsinterventies én evaluatieonderzoek te structureren aan de hand van Context- Mechanismen-Outcome ('C-M-O')-configuraties. Beleid(interventies) worden verondersteld zich te onttollen in bepaalde contexten. Denk aan hoe de rechtshandhaving is georganiseerd, of er sprake is van centrale én mede-overheden of juist niet, hoe omvangrijk de misdaad is enzovoort. Deze contexten hebben op zichzelf genomen geen 'wervingskracht'. Dat hebben de gedragsmechanismen die in de verschillende interventies besloten liggen. Denk aan interventies die de werking van (financiële) prikkels veronderstellen, interventies die juist voortborduren op vertrouwen tussen organisaties en medewerkers, interventies die regels en sancties centraal stellen, interventies die een beroep doen op de behoefte van mensen cognitieve dissonantie te voorkomen of te reduceren enzovoort (voor een categorisering zie Hedström en Swedberg, 1998). Ten slotte wordt verondersteld dat de combinatie van Contexten & Mechanismen bepaalde resultaten ('outcome') tot gevolg zal hebben. Denk aan een betere opsporing of het terugdringen van misdaad. Het belang van een dergelijke benadering is tweërlei. Enerzijds leidt het tot betere onderzoeksopzetten, omdat ook de werkzame mechanismen en de tussenstappen tussen interventie en effect worden onderzocht. Anderzijds is kennis over deze mechanismen veel breder toepasbaar, ook bij andere interventies en op andere beleidsterreinen.

In de vierde en laatste plaats is bij evaluatieonderzoek de vraag van belang naar interne validiteit: of de beoogde resultaten van de onderzochte beleidsinterventie door die interventie tot stand zijn gekomen, of door iets anders, met andere woorden, of een bepaalde beleidsinterventie de beoogde resultaten heeft veroorzaakt (Cook en Campbell, 1979; Sherman e.a., 2002). Dit criterium kan worden geplaatst in de benadering van de Campbell Collaboration Group on Crime and Justice, die we hierboven introduceerden. Hoe zekerder dit kan worden vastgesteld, des te hoger de interne validiteit, en des te meer gesproken kan worden van effecten in strikte zin. Voor de beoordeling van deze interne validiteit van de onderzoeken hebben we gebruikgemaakt van de Maryland Scientific Methods Scale (MSMS), een vijf-puntsschaal (Sherman e.a., 2002; Farrington e.a., 2002) (zie tabel 2).

Tabel 2: Vijf niveaus van interne validiteit conform de Maryland Scientific Methods Scale

Niveau 1	Samenhang tussen een beleidsinterventie en uitkomstvariabele na invoering van de beleidsinterventie
Niveau 2	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, zonder (vergelijkbare) controlegroep
Niveau 3	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, gemeten in een experimentele en vergelijkbare controlegroep (quasi-experimenteel design)
Niveau 4	Niveau 4: Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie gemeten in een quasi-experimentele setting, met controle voor andere variabelen die de uitkomstvariabele beïnvloeden
Niveau 5	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, waarbij de beleidsmaatregel 'at random' is toebedeeld aan een experimentele en een controlegroep (experimenteel design).

Elk niveau hoger op deze schaal betekent een grotere mate van zekerheid waarmee resultaten aan beleid kunnen worden toegeschreven (interne validiteit). Op niveau 1 is die zekerheid nihil. In feite wordt alleen gekeken naar de uitkomstvariabele (bijvoorbeeld mate van jeugdcriminaliteit) op een bepaald punt in de tijd, zonder een vergelijking te trekken met de periode vóór invoering van de beleidsinterventie, of met situaties waarin deze niet is toegepast. Op niveau 2 wordt de uitkomstvariabele ná invoering van de beleidsinterventie wél vergeleken met de situatie ervóór. Maar ook dan kan niet meer worden geclaimd dan een eventuele verandering in de uitkomstvariabele *ten tijde van* het beleid, die dus ook aan iets anders zou kunnen worden toegeschreven dan aan de beleidsinterventie. Op niveau 3 spreken we van een quasi-experimenteel design. De uitkomstvariabele wordt niet alleen vergeleken met de situatie vóór invoering van de beleidsinterventie, maar ook met een 'controlegroep' die niet met het beleid is geconfronteerd. Volgens Cook en Campbell (1979) vormt dit het minimumniveau waarmee op verantwoorde wijze conclusies kunnen worden getrokken over de effectiviteit van een interventie. Op niveau 4 wordt daarbij nog eens gecontroleerd voor specifieke andere factoren, die volgens de onderzoeker de uitkomstvariabele mogelijk beïnvloeden (bij criminaliteit valt te denken aan sociaal-psychologische kenmerken, opvoeding, opleiding). Farrington en Welsh (2005) hanteren het criterium van

experimenteel design (niveau 5) voor het opnemen van studies in hun onderzoekssynthese. Met het 'at random' toebedelen van de beleidsinterventie aan experimentele en controlegroep worden automatisch alle mogelijke externe invloeden uitgeschakeld, óók van factoren die de onderzoeker niet bekend zijn.

Met de bovengenoemde vier criteria beoordelen wij de onderzoeken breed op hun merites. Er ontstaat zo meer aandacht voor de inhoud en de beschrijvende kwaliteit van de onderzoeken dan wanneer we alléén de MSMS als maatstaf zouden gebruiken.

De kwaliteit van de 31 onderzoeken

Hoe goed zijn de onderzoeken als we ze beoordelen aan de hand van de vier criteria? Het blijkt dat de onderzoeken op het ene criterium beter scoren dan op het andere, maar al met al komen ze er redelijk uit. Appendix 1 illustreert dit aan de hand van de projectevaluaties. De meeste onderzoeken maken gebruik van gegevens uit meerdere informatiebronnen. Door verschillende gegevens te combineren, geven de onderzoekers vaak een genuanceerd beeld van hun onderwerp. Dit is een sterk punt van de onderzoeken, het maakt hun resultaten stevig en valide.

Ook brengen alle onderzoeken in beeld of de onderzochte beleidsmaatregel is geïmplementeerd. Hierdoor kan met enige zekerheid aangenomen worden dat een gevonden resultaat verband houdt met de beleidsinterventie. Dit helpt ons om de causaliteitsvraag beter te kunnen beantwoorden. Ook op dit punt zitten de onderzoeken gedegen in elkaar.

Verder vonden we dat in ruim éénderde van de evaluatieonderzoeken expliciet is beschreven langs welke weg een nieuwe beleidsinterventie tot de beoogde resultaten zou moeten leiden. Met andere woorden, deze onderzoeken bevatten een rudimentaire weergave van de 'programmatheorie'. De onderzoeken geven vaak heldere analyses van de werking van het beleid in de praktijk. Maar een expliciete analyse waarmee 'veronderstelde' met 'werkzame' mechanismen worden vergeleken, die dieper gaat dan een beschrijving van beoogde interventies en beoogde doelen, hebben we nauwelijks aangetroffen. Dit punt is minder sterk.

Afgemeten aan de MSMS-schaal bevinden de meeste evaluaties zich op niveau 1 of 2. Meestal verantwoordden de onderzoekers niet waarom ze geen controleconditie (nulmeting of controlegroep)

hebben gebruikt. Het is duidelijk dat de onderzoeken op het punt van interne validiteit niet sterk zijn. Het is wel zo dat de MSMS soms niet van toepassing is omdat een effectmeting niet is beoogd. Ook is het soms niet mogelijk om een geschikte controleconditie te vinden, bijvoorbeeld omdat een maatregel landelijk is ingevoerd. Maar hoe het ook zij, de score op de MSMS is laag, en we kunnen ons afvragen of dit in een aantal onderzoeken niet beter had gekund.

Waarom synthese?

Tussen de 31 evaluatieonderzoeken zijn geen evaluatiestudies van niveau 3 of hoger gevonden. Menigeen zal daarom stellen dat het synthetiseren van bevindingen niet raadzaam is. Wij willen dit bestrijden. Waarom?

In de eerste plaats komt de descriptieve validiteit van de meeste studies tegemoet aan algemene evaluatiestandaarden (meerdere onderzoeksmethoden, accuraatheid van bevindingen). In de tweede plaats wordt in vrijwel elk van de 31 studies aandacht besteed aan het implementatieproces. In de derde plaats wordt in veel studies aandacht besteed aan onderliggende werkingsmechanismen; zodra het beleid is geïmplementeerd. Hiermee valt enig zicht te krijgen op de 'black box' van beleidsinterventies: wat is het in een beleid dat de uitkomst veroorzaakt, hoe komt deze tot stand, voor wie, en onder welke condities? Als het doel van een synthese van evaluatieonderzoeken is erachter te komen of interventies daadwerkelijk zijn ingevoerd of als de taak is achter de werkingsmechanismen te komen, dan zou het onverstandig zijn om bevindingen beneden niveau 3 van de MSMS niet te synthetiseren (vergelijk Pawson, 2002). Dit wordt nog eens onderstreept door het toenemende belang dat wordt gehecht aan een 'what works-' of kennisgestuurde aanpak, vooral ook op het gebied van criminaliteit en rechtshandhaving. Een reconstructie van beleids-theorieën en een blik op werkzame mechanismen verhoogt niet alleen de kwaliteit van evaluaties, maar ook de bruikbaarheid van evaluatieonderzoek voor beleid.

Inhoudelijke synthese

Hoe gaat nu een inhoudelijke synthese van de onderzoeksresultaten in zijn werk en wat levert dit uiteindelijk op aan kennis? Voor elk evaluatieonderzoek hebben we geïnventariseerd:

- de inhoud van het beleid (respectievelijk de interventies), zoals afgeleid uit de evaluatierapporten;
- de mechanismen op basis waarvan het beleid wordt verondersteld te werken, en de condities waaronder het beleid tot resultaten zou moeten leiden;
- de feitelijke werking van het geëvalueerde beleid, al dan niet gekoppeld aan voornoemde mechanismen;
- contextuele condities waaronder het beleidsproces al dan niet effectief blijkt te zijn;
- het feitelijk behalen van de met het beleid beoogde doeleinden: welke resultaten zijn bereikt?

Appendix 2 illustreert dit aan de hand van de projectevaluaties. Vervolgens zijn de bevindingen uit de onderzoeken inhoudelijk gesynthetiseerd, dat wil zeggen samengebracht op een niveau dat de individuele onderzoeken overstijgt, vorsend naar algemene lijnen en rode draden.

Wat leert deze inhoudelijke synthese van onderzoeken ons nu over de rechtshandhaving in Nederland? Wat kunnen we leren over de uitvoering van projecten, over maatregelen die de handhavingsketen willen verbeteren en over de invoering van nieuwe wetten en regels?¹

Projecten op het gebied van rechtshandhaving bestaan vaak uit verhoogde input van mankracht, expertise en middelen, en uit intensivering van samenwerking. Beleidsmakers vinden het belangrijk dat projecten goed aansluiten bij de bestaande ketenorganisatie. Daarom worden ketenpartners zoals politie en Openbaar Ministerie actief bij een project betrokken. Een aantal projecten behelzen extra sturing of overleg. Door dit alles zou betere informatie beschikbaar moeten komen en zouden handhavingsinstrumenten beter gebruikt moeten worden. Dit moet dan leiden tot meer opsporingsonderzoeken, aanhoudingen, veroordelingen, en uiteindelijk tot een verhoogde pakkans en minder criminaliteit.

Wat zien we in de praktijk? De verhoogde input en de betere samenwerking blijken goed gerealiseerd te worden. De informatiepositie van betrokken organisaties wordt vaak versterkt. Maar we vonden een aantal systematische knelpunten. Er treden vaak problemen op in de aansturing, wat leidde tot fragmentatie en inefficiënties.

¹ Op de categorie organisatie-evaluaties gaan we niet in omdat er maar één onderzoek in de steekproef zit.

Ook was er in enkele projecten gebrek aan scholing en expertise van medewerkers of verloop onder personeel. Organisaties blijken vaak terughoudend met het oppakken van de aan nieuwe projecten verbonden activiteiten. De projectkennis wordt dan niet omgezet in daden. Het nieuwe dat het project heeft opgeleverd raakt moeilijk ingebed in de bestaande werkprocessen. Dit kan duiden op een variant van het bekende 'kennis-attitude-gedragsprobleem': via communicatie opgedane projectkennis in een organisatie hoeft nog niet te betekenen dat die kennis ook in daden wordt omgezet.

In de meeste projecten komen meer opsporingsonderzoeken, aanhoudingen en andere 'producten' tot stand. De einddoelen zijn vaak een zaak van lange adem. De evaluatieonderzoeken van projecten die wij analyseerden, laten op het punt van de einddoelen nog nauwelijks resultaten zien. Soms is het nog te vroeg daarvoor.

Interventies in de keten beslaan meestal de hele keten van preventie, opsporing en vervolging tot en met sanctietoepassing. Het gaat vaak om de invoering van nieuwe regulering, handhavingsinstrumenten en -methoden en om verbetering van de communicatie en samenwerking tussen ketenpartners. Dit zou moeten leiden tot betere opsporings- en vervolgingsresultaten en uiteindelijk tot minder criminaliteit.

Wat zeggen de onderzoeken hierover? De onderzoeken stellen (onder meer) vast dat er systematische knelpunten zijn die bij verschillende vormen van ketenhandhaving telkens terugkeren. Dit zijn met name:

- gebrek aan aansturing van de organisatie; te weinig integrale besturing;
- gebrek aan coördinatie tussen ketenpartners en met externe partners;
- gebrek aan gelijkvormigheid in beleid tussen ketenpartners (verschillende stijlen, werkwijzen, beleidscycli en prioriteiten); deelbewerkingen in de keten zijn te weinig in lijn met elkaar;
- gebrek aan organisatiecapaciteit;
- gebrek aan deskundigheid (opleidingsniveau);
- teveel focus op misdadbestrijding, ten koste van preventie en kennisverwerving.

Door deze knelpunten werken de mechanismen in de praktijk niet zoals dat vooraf door het beleid werd beoogd.

Wat zijn de resultaten van keteninterventies? Het blijkt dat de organisatie en de uitvoering van de handhaving in een aantal gevallen onder de norm worden bevonden. Op gemeentelijk niveau wordt

bijvoorbeeld veel gedoogd. De ketenstructuur vertoont over het algemeen veel inefficiënties. Daarnaast lijken zowel het beleid als de evaluaties een blinde vlek te vertonen. Rechtshandhavingsbeleid draait uiteindelijk om het nalevingsgedrag van burgers en rechtspersonen, maar veel interventies en ook veel evaluatieonderzoek is daar niet rechtstreeks op gericht. De focus betreft voornamelijk de uitwerking van beleidsingrepen op de handhavingsketen, terwijl de uiteindelijke effecten hiervan op het nalevingsgedrag veelal buiten beschouwing blijven.

Wat kunnen we, ten slotte, leren uit de evaluaties van wet- en regelgeving en wijzigingen hierin? Hoewel de veelvormigheid aan geëvalueerde wet- en regelgeving groot is, komen er enkele algemene lijnen naar voren.

Meestal blijkt het bij nieuwe wet- en regelgeving te gaan om uitbreiding van het handhavingsinstrumentarium, soms gaat het om stroomlijning van bevoegdheden. Beleidsmakers verwachten dat dit zal leiden tot ruimere toepassing van opsporings- en handhavingsmogelijkheden en tot verhoogde efficiëntie van de handhaving.

Hogere ophelderingspercentages en uiteindelijk minder crimineel gedrag moeten uiteindelijk het gevolg zijn.

Het daadwerkelijke gebruik van een aantal verruimde handhavingsmogelijkheden is in de praktijk gering en vertoont knelpunten. Dit kan te maken hebben met het feit dat de evaluatieonderzoeken kort na invoering van de wetten of wetswijzigingen plaatsvonden. Betrokkenen in de praktijk zelf zijn positief over de doelbereiking van de wet- en regelgeving; het handhavingsinstrumentarium is completer geworden en bepaalde vormen van criminaliteit zijn beter te bestrijden.

Besluit

Onze exercitie heeft laten zien dat het evaluatieonderzoek op het gebied van rechtshandhaving goede descriptieve kwaliteit heeft. Ook geeft het zicht op implementatie en resultaten van beleid. Het vormt dan ook een geschikte basis om kennis voor beleid uit op te doen die het niveau van het individuele onderzoek overstijgt. Het belang van het verhogen van de kwaliteit van evaluatieonderzoek staat niettemin buiten kijf. Het gedachtegoed van de Maryland-standards over het waarborgen van de interne validiteit aan de hand

van (quasi-)experimentele designs vormt daarbij een belangrijke leidraad.

Dit neemt niet weg dat deze concrete exercitie ons leert dat er meer onder de zon is dan Maryland-standards. In de eerste plaats hebben we daarom gewezen op het belang van onderzoek naar de feitelijke implementatie van maatregelen. Want beleid dat nooit begonnen is, kan ook niet zinvol worden geëvalueerd. Bovendien kan bijsturing van beleid ook al in de implementatiefase plaatsvinden.

In de tweede plaats hebben wij gewezen op het belang van onderzoek naar de werkzame mechanismen, aan de hand waarvan men kan laten zien *waarom* beleid wel of niet werkt. De zogenoemde 'black box' van beleidsinterventies en hun uitwerking dient door beleidsmakers en onderzoekers te worden geopend. Een reconstructie van beleidstheorieën en een blik op werkzame mechanismen verhoogt niet alleen de kwaliteit van evaluaties, maar ook de bruikbaarheid van evaluatieonderzoek voor beleid. Kennis over deze mechanismen laat namelijk zien welke elementen van beleid essentieel zijn. Daarbij maakt het in feite niet uit of deze kennis berust op 'succesvol' of 'mislukt' beleid.

Vooraf wanneer evaluatieonderzoek inzicht verschaft in *hoe* beleid werkt *én of* het effectief is, biedt het kennis over preventie en bestrijding van criminaliteit waarop beleidsmakers beslissingen kunnen baseren.

Literatuur

Algemene Rekenkamer

Tussen beleid en uitvoering; lessen uit recent onderzoek van de Algemene Rekenkamer

Den Haag, Kamerstukken II, 2002-2003, 28 831 nrs. 1-2, 2003

Cook, T.D., T.D. Campbell

Quasi-experimentation; design and analysis issues in field settings

Chicago, Rand McNally/ Boston, Houghton-Mifflin, 1979

Farrington, D.P., D.C.

Gottfredson e.a.

The Maryland Scientific Methods Scale

In: Sherman, L. W., Farrington e.a. (red.), *Evidence-based crime prevention*, London, Routledge, 2002, p. 13-21

Farrington, D.P., B.C. Welsh

Randomized experiments in criminology; what have we learned in the last two decades?

- Journal of experimental criminology, nr. 1, 2005, p. 9-38
- Hedström, P., R. Swedberg**
Social mechanisms; an introductory essay
In: Hedström, P., R. Swedberg (red.), *Social mechanisms; an analytical approach to social theory*, New York, Cambridge University Press, 1998
- Klein Haarhuis, C.M., M. van Ooyen, E.R. Kleemans en F.L. Leeuw**
Kennis voor beleid; een synthese van 58 onderzoeken op het gebied van rechtshandhaving
Den Haag, WODC, Cahier 2005-5
- Leeuw, F.L.**
Reconstructing program theories; methods available and problems to be solved
American journal of evaluation, 24e jrg., nr. 1, 2003, p. 5-20
- Leeuw, F.L., L. Cooksey**
Evaluating the performance of development agencies; the role of meta-evaluations
In: World Bank, *Proceedings of the Fifth Conference on Evaluations*, Washington, World Bank, 2004
- Lösel, F.**
The efficacy of correctional treatment; a review and synthesis of meta-evaluations
In: McGuire, J. (red.), *What works; reducing re-offending*, Chichester, UK, John Wiley & Sons, 1995, p. 79-114
- Pawson, R., N. Tilley**
Realistic evaluation
Londen/Thousand Oaks/New Delhi, Sage Publications, 1997
- Pawson, R.**
Evidence-based policy; the promise of 'realist synthesis'
Evaluation, 8e jrg., nr. 3, 2002, p. 340-358
- Rist, R.C., N. Stame**
From studies to streams
New Brunswick, Transaction Publishers, 2004
- Sherman, L.W., D.P. Farrington e.a. (red.)**
Evidence-based crime prevention
London, Routledge, 2002
- Wittebrood, K., M. van Beem**
Sociale veiligheid vergroten door gelegenheidsbeperking; wat werkt en wat werkt niet?
Den Haag, Sociaal Cultureel Planbureau i.o.v. Raad voor Maatschappelijke Ontwikkeling, 2004, Werkdocument 2
- Yin, R.K.**
Case study research, design and methods
Newbury Park, Sage Publications, 1994
- Zall Kusek, J., R.C. Rist**
Ten steps to a results-based monitoring and evaluation system; a handbook for development practitioners
Washington, The World Bank, 2004

Appendix 1: De kwaliteit van projectevaluatieonderzoek

Interventie (X) en beoogd effect (Y)	Opzet evaluatie (descriptieve validiteit)	Beschrijving van de implementatie?	Reconstructie en evaluatie van beleidsveronderstellingen (mechanismen, resultaten)	MSMS (1-5) Toelichting
Project M. Instelling van een anoniem meldpunt misdrijven moet leiden tot meer opsporingen en meer vervolgingen.	<ul style="list-style-type: none"> * Inventarisatie en proces-beschrijving van binnengekomen meldingen in 5 politieregio's gedurende de eerste 8 maanden van het project. * Media-analyse van ongeveer 300 artikelen * Steekproef onder de bevolking voor interviews (voor en na) t.a.v. attitudes en perceptie van M. 	Ja.	Beschrijving van de werking met strakke proceslijn (van communicatie tot meldingen, tot en met verwerking en gebruik van de meldingen). Hierdoor kunnen effecten met meer zekerheid worden toegeschreven aan het beleid.	Niveau 2 Nulmeting en tussenmeting beschikbaar. Geen controlecondities.
NOVO-team voor oorlogsmisdrijven moet leiden tot opsporing, intermediëring en bewijsvergaring.	Tussenevaluatie van effecten, aan de hand van documentstudie en interviews. Bestudering van de condities waaronder NOVO heeft gefunctioneerd: <ul style="list-style-type: none"> * Contextanalyse * Interne analyse * Triangulatie en rapportage. 	Ja.	Beschrijving van de werkingsmechanismen, met confrontatie tussen geformuleerde en uitgevoerde taken.	Niveau 2 Nulmeting en tussenmeting beschikbaar. Geen controlecondities.
Kernteam: gespecialiseerde opsporings-teams dienen complexer en grootschaliger zaken aan te pakken en op te lossen.	Beschrijvende studie van de 'toegevoegde waarde' van de Kernteams. <ul style="list-style-type: none"> * Deskonderzoek * Géén dossieronderzoek (is verantwoord) * Interviews in alle 7 Kernteams onder functionarissen (slechts 8 per team) en enquêtes onder alle medewerkers. 	Ja.	Geen reflectie op centrale mechanismen van beleid, maar focus op implementatieproces en organisatie.	Niveau 1 Nulmeting ontbreekt. Het ontbreken van een nulmeting is niet expliciet verantwoord. Geen controlecondities. <i>De facto</i> dus een doelbereikings-onderzoek. Wel wordt meestal goed verantwoord waarom informatie gebrekkig is.

<p>Financieel Recherchen: opleiding en communicatie tussen ketenpartners (politie, O.M.) moet leiden tot acceptatie, integratie en implementatie van financieel rekenen. Dit ter bestrijding van georganiseerde criminaliteit en fraude.</p>	<p>Studie naar de mate van implementatie na 5 jaar, met als hoofdindicator: de integratie van FR in de handhavingsketen.</p> <ul style="list-style-type: none"> * Kwalitatief: Interviews (n=155), enquêtes (n=633), trefwoorden-brainstorm, documentstudie * Kwantitatief: Compas-gegevens, IFT-registraties, BOOM, CJIB. Voor evaluatie van het onderwonderdeel is een steekproef getrokken. <p>Niet representatief voor politieregio's en parketten.</p>	<p>Ja.</p>	<p>De feitelijke werkingsmechanismen worden niet sterk teruggekoppeld naar veronderstelde mechanismen, maar staan verdeeld over de deelrapporten.</p>	<p>Niveau 1 Nulmeting ontbreekt. Geen controlecondities. Omdat FR het enige project over dit onderwerp is, zijn sommige resultaten automatisch met enige zekerheid aan het project FR toe te schrijven.</p>
<p>Project 'Hektor' moet drugsoverlast in Venlo terugdringen met behulp van een multimodale aanpak.</p>	<ul style="list-style-type: none"> * Verzameling van registratiegegevens politie, belastingdienst, O.M. * Interviews met ondernemers * Peiling drugsoverlast onder bevolking * Observaties door stadswachten aan de hand van schema. 	<p>Ja, bij eindmeting.</p>	<p>Beschrijving van de veronderstelde aanpak en uitkomsten, indicatoren op basis hiervan vastgesteld. Er vindt terugkoppeling naar veronderstellingen plaats.</p>	<p>Niveau 2 Opzet met nulmeting, tussenmeting en eindmeting. Geen controlecondities.</p>
<p>'Samenspannen tegen XTC' moet de XTC-productie en -handel in en vanuit Nederland terugdringen.</p>	<ul style="list-style-type: none"> * Eerst zijn indicatoren geïnventariseerd voor meten van input, samenwerking en resultaten * De indicatoren zijn gescreend op validiteit en beschikbaarheid * 59 indicatoren gekozen voor de eerste meting gebaseerd op plannen van de aanpak XTC (beoogde input, processen, output en 'outcome') * Gegevens uit registratiesystemen van betrokken actoren * Schriftelijke vragenlijst over samenwerking. 	<p>Ja, bij tussenmeting.</p>	<p>Veronderstellingen worden beschreven en indicatoren op basis hiervan vastgesteld. Bevindingen worden teruggekoppeld naar veronderstellingen. De studie laat zien wat er van de geplande aanpak tot stand is gekomen.</p>	<p>Niveau 2 Opzet met nulmeting, tussenmeting en eindmeting. Geen controlecondities.</p>

Appendix 2: De inhoud van projectevaluatieonderzoek: interventie(s) en beoogde effecten, veronderstelde mechanismen, bevindingen m.b.t. mechanismen en andere factoren, resultaten

Projecten	Interventie(s) en beoogde effecten	Veronderstelde mechanismen	Feitelijke werking van mechanismen; overige factoren	Resultaten
Project 'M': (Meld Misdaad Anoniem)	Instelling van een anoniem meldpunt misdrijven moet leiden tot meer opsporingen en meer vervolgingen.	<ul style="list-style-type: none"> * Communicatie 'M' * Bekendheid bij burger * Attitudes * Intentie tot gedrag * Feitelijk gedrag bij burgers * Frontoffice: meldpunt * Backoffice: verwerken van bruikbare meldingen * Resultaat: opsporing, vervolging. 	<ul style="list-style-type: none"> * Knelpunt is de beslissingsbevoegdheid van 'backoffices' * Van welke <i>typen</i> delicten meldingen zijn gewenst, is beter te sturen. 	<ul style="list-style-type: none"> * Frontoffice, 2000 gesprekken * Burgerattitudes positief * 6 tot 9% onderzoek * 1-2% resultaat in de opsporing * Geen langetermijneffecten beschikbaar.
Evaluatie NOVO-team oorlogsmisdrijven	Het NOVO-team is opgericht ter opsporing van oorlogsmisdrijven, intermediering en bewijsvergaring.	<ul style="list-style-type: none"> * Aangifte * Bewijsvergaring * Verhoor * Institutionele inbedding: aansluiting bij ketenpartners * Rapportage. 	<ul style="list-style-type: none"> * Aard van de zaken is problematisch (moeilijk bewijsbaar) * Goede kwaliteit NOVO-team * Lage prioriteit bij ketenpartners. 	Het NOVO-team heeft geen nieuwe opsporingen gegenereerd.
Evaluatie Kernteams	Instelling van gespecialiseerde opsporingsteams moet leiden tot meer, complexere, en grootschalige opsporingsonderzoeken.	<ul style="list-style-type: none"> * Expertise en mankracht verenigen in speciale teams. 	<ul style="list-style-type: none"> * Veel verschillen tussen kernteams * Gebrek aan landelijke strategische aansturing * Positie 'uit de wind' blijkt gunstig voor complexiteit, expertise, samenhang en continuïteit * Medewerkers blijken te licht opgeleid * Kernteams blijken onontbeerlijk als communicatieknooppunt. 	Geen eenduidige conclusie mogelijk over de toegevoegde waarde van de kernteams.

Financieel Rechercheren	Opleiding en communicatie tussen ketenpartners (politie, O.M.) moet leiden tot acceptatie, integratie en implementatie van financieel rechercheren. Ter bestrijding van georganiseerde criminaliteit en fraude.	<i>Sub-doelstellingen.*</i> Bestuurlijk-organisatorische adviezen * Communicatieverbetering * Opleidingen * Methodenontwikkeling * Aanpassing wet- en regelgeving * Ontwikkeling van preventieve maatregelen.	<ul style="list-style-type: none"> * Doelstelling niet helder en te ambitieus * Beleving van delicten als niet ernstig * Te gering zaaksaanbod * Gebrek aan aansturing. 	<ul style="list-style-type: none"> * Financieel rechercheren is op de agenda gezet bij politie en Openbaar Ministerie * Wel geïntegreerd onder specialisten, maar niet in de gemiddelde politieregio.
Project 'Hektor' tegen drugs-overlast in Venlo	Drugsoverlast terugdringen met behulp van een multimodale aanpak.	<ul style="list-style-type: none"> * Intensivering van de handhaving: verhoogde input van mensen en middelen (hogere pakkans) * Aanpak van drugsgerelateerd vastgoed * Aanpassing coffeeshopbeleid. 	<ul style="list-style-type: none"> * Inspanningen en 'productie' (bijv. aanhoudingen) van betrokken diensten zijn toegenomen * Aanpassing coffeeshopbeleid nog niet gerealiseerd ten tijde van tussenmeting. 	<ul style="list-style-type: none"> * Overlast en criminaliteit bij tussenmeting nog niet verminderd * Geregistreerde drugspanden, -dealers en -runners afgenomen * Groter gevoel van veiligheid en meer economische activiteit.
Evaluatie Nota 'Samen-spannen tegen XTC'	Reductie van productie en handel XTC in en vanuit Nederland.	<ul style="list-style-type: none"> * Intensivering van de aanpak van XTC: verhoogde input van mensen en middelen (hogere pakkans) * Aanpak van precursoren en hulpmiddelen, productielocaties en handel. 	<p>Eerste resultaten overwegend in de beoogde richting:</p> <ul style="list-style-type: none"> * Inspanningen en 'productie' (bijv. ontmantelingen) door betrokken diensten toegenomen * De informatiepositie van Douane, FIOD-ECD is versterkt * De samenwerking van opsporings- en controlediensten is geïntensiveerd. 	<ul style="list-style-type: none"> * Meer inbeslagnames in het eerste jaar van implementatie * Meer productieplaatsen ontmanteld * Minder milieudumpingen * Vooralsnog geen nadere resultaten (alleen eerste meting verricht).

Causale effecten van beleid

Over gecontroleerde en natuurlijke experimenten

*H.D. Webbink**

Een goede evaluatie van de effecten van beleid kan veel opleveren. Daarmee kan kennis worden verkregen over wat werkt en wat niet werkt. Beleid dat niet werkt of zelfs schadelijk is, kan worden gestopt. Beleid dat wel werkt, kan breed worden ingevoerd. Goede evaluaties zijn echter niet eenvoudig en de literatuur geeft tal van voorbeelden van slechte evaluaties, met de daarbij horende desastreuze gevolgen. Een beroemd voorbeeld uit de medische wetenschap betreft het effect van de zogenoemde hormoonvervangings therapie voor vrouwen na de menopauze (U.S. Department of Education, 2003). In de afgelopen dertig jaar werd in meer dan 25 studies met controlegroep, echter zonder toewijzing op basis van loting, gevonden dat deze therapie de kans op een hartinfarct of hartaanval vermindert. Maar toen uiteindelijk twee grootschalige experimenten met aselechte toewijzing werden uitgevoerd, bleek het tegendeel: de therapie vergroot de kans op hartziekten, beroerte en borstkanker. Dit voorbeeld illustreert niet alleen het belang van valide kennis, maar laat ook zien dat zelfs herhaalde replicatie van studies met controlegroepen tot volkomen verkeerde conclusies kan leiden. Vertekening van de resultaten door verschillen tussen de controle- en de experimentele groep is een groot probleem. De ideale aanpak voor het vaststellen van causale effecten van beleid is het klassieke experiment. Door deelnemers aselechte toe te wijzen aan een experimentele groep die te maken krijgt met de beleidsinterventie, en een controlegroep die niet te maken heeft met de beleidsinterventie, kan de invloed van andere factoren sterk worden gereduceerd. Het is echter lastig om echte experimenten te realiseren vanwege praktische problemen of ethische bezwaren. Onderzoekers zijn dan ook vaak aangewezen op alternatieve

* Dr. Dinand Webbink is programmaleider Onderwijs & Wetenschap bij het Centraal Planbureau. Hij dankt Free Huizinga, Bert Smid en Daniël Waagmeester voor commentaar op een eerdere versie van dit artikel.

methoden. In een groot aantal recente studies worden quasi-experimentele onderzoeksdesigns toegepast. Deze studies maken gebruik van toevallige situaties in de werkelijkheid waardoor aselecte controlegroepen worden gevormd. Dit artikel geeft een overzicht van dit nieuwe onderzoek, de daarin gebruikte methoden en recente Nederlandse toepassingen. De meeste voorbeelden zijn ontleend aan de onderwijs- en arbeidseconomie, en enkele aan het justitiële werkkterrein. Echter, deze methoden zijn breed toepasbaar. Ongetwijfeld zijn er op het terrein van strafrechtelijke interventies of interventies gericht op preventie van criminaliteit veel nieuwe toepassingsmogelijkheden.

De opbouw van dit artikel is als volgt. Allereerst wordt ingegaan op het kernprobleem bij het vaststellen van de effecten van een beleidsinterventie. Daarna worden twee oplossingen besproken: gecontroleerde experimenten en natuurlijke experimenten. Vervolgens wordt een overzicht gegeven van recente Nederlandse studies die gebruikmaken van (quasi-)experimentele designs. Tot slot worden enige lessen geformuleerd en mogelijkheden voor het juridische werkkterrein aangegeven.

Het evaluatieprobleem

Het kernprobleem bij de evaluatie van beleid is dat de groep die te maken heeft met de beleidsinterventie meestal niet 'random' is. Vaak vindt er selectie plaats bij deelname aan een bepaalde beleidsinterventie, ofwel door de deelnemer ofwel door de uitvoerders van de beleidsinterventie. Door deze selectie kunnen er systematische verschillen bestaan tussen de groep die wel en de groep die niet te maken heeft met de beleidsinterventie. Een standaardvoorbeeld betreft het vaststellen van de effecten van training voor werklozen of werkenden. Om het effect van de training vast te stellen dient de groep deelnemers te worden vergeleken met een groep niet-deelnemers. Daarbij dient rekening te worden gehouden met verschillen tussen beide groepen. Het cruciale probleem is dan dat deelnemers zich selecteren voor de training. Anders gezegd, de groep die zich aanmeldt en bereid is om de training te volgen wijkt af van de groep die zich niet aanmeldt. Voor een deel zal een onderzoeker rekening kunnen houden met de verschillen. Voor een deel van de verschillen, zoals bijvoorbeeld motivatie of bepaalde vaardigheden, is dat

moelijk, omdat het lastig is om deze verschillen waar te nemen. In dat geval kan niet worden vastgesteld of bepaalde uitkomsten na afloop van de training, bijvoorbeeld het sneller vinden van een baan, veroorzaakt worden door de training dan wel door verschillen in motivatie of vaardigheden. De vertekening door het negeren van de invloed van niet waargenomen factoren kan aanzienlijk zijn. In onderzoek naar het financieel rendement van het volgen van training door werknemers, zoals in dit voorbeeld, werden in het verleden zeer hoge rendementen gevonden (Frazis en Loewenstein, 1999). Echter, in recente studies die gebruikmaken van experimentele designs worden rendementen gevonden die niet significant verschillen van nul (Leuven en Oosterbeek, 2002). Dit suggereert dat de hoge rendementen vooral het gevolg zijn van selectiviteit bij de deelname aan training.

Een ander beroemd voorbeeld uit de onderwijseconomie betreft het effect van de klassengrootte op de prestaties van leerlingen. Het direct vergelijken van de resultaten van leerlingen in grote klassen met die in kleine klassen kan een vertekend beeld opleveren. Het is immers aannemelijk dat de samenstelling van klassen niet door toeval (loting) gebeurt maar op basis van beslissingen en inschattingen van leraren en mogelijk ook ouders. Het gevolg kan zijn dat leerlingen in kleine klassen systematisch verschillen van leerlingen in grote klassen. Als de onderzoeker niet al deze verschillen observeert, kan een schatting van het effect van de beleidinterventie 'klassenverkleining' vertekend zijn door de niet waargenomen verschillen. De Amerikaanse onderwijseconoom Eric Hanushek publiceerde in 1986 een invloedrijke overzichtsstudie waarin hij 277 schattingen van het effect van klassengrootte op prestaties van leerlingen verzamelde. Het belangrijkste resultaat was dat alle mogelijke uitkomsten waren gevonden: 15% van de schattingen was significant positief, 13% significant negatief, 27% niet-significant positief, 25% niet-significant negatief en 20% onbekend. Hanushek concludeerde daaruit dat er geen sterk of systematisch verband was tussen onderwijsuitgaven en prestaties van leerlingen. Echter, al deze schattingen kunnen vertekend zijn door niet-waargenomen factoren. Recente schattingen in studies die gebruikmaken van (quasi-)experimentele designs geven een veel consistentere beeld (Webbink, 2005). Ook blijkt in sommige studies het gevonden effect geheel te veranderen als een (quasi-)experimenteel design wordt toegepast, bijvoorbeeld van significant positief naar significant

negatief. De bekendste criticaster van de conclusie van Hanushek is Alan Krueger. Hij acht de 'vote counting exercise' ongeschikt: onderzoek is niet democratisch, het gaat om de kwaliteit van de studie (Krueger, 2003).

Oplossingen

In principe kan het evaluatieprobleem worden opgelost door rekening te houden met meer factoren. In de praktijk is dat vaak niet mogelijk doordat sommige factoren niet of onvoldoende zijn waar te nemen – denk aan motivatie of aanleg – en door restricties op tijd en geld. In het economisch onderzoek is de laatste jaren een sterke groei te zien van studies die gebruikmaken van (quasi-)experimentele onderzoeksdesigns. Deze designs, die gebaseerd zijn op gecontroleerde of natuurlijke experimenten, bieden een oplossing voor het evaluatieprobleem. Experimenten kunnen gecreëerd worden door onderzoekers (gecontroleerde experimenten) maar ook ontstaan als gevolg van toevallige situaties in de werkelijkheid (natuurlijke experimenten).

Gecontroleerde experimenten

Door deelnemers op basis van toeval toe te wijzen aan een experimentele en een controlegroep wordt de kans sterk verkleind dat andere factoren dan de beoogde interventie de resultaten beïnvloeden. Het verschil in uitkomsten tussen de experimentele groep en de controlegroep kan dan worden toegewezen aan de interventie. De aselecte toewijzing aan de experimentele en de controlegroep zorgt er immers voor dat niet waargenomen kenmerken van de deelnemers aan het onderzoek niet systematisch zullen verschillen tussen beide groepen. Een prominent voorbeeld is het Star-experiment met klassenverkleining in het basisonderwijs in Tennessee in de jaren tachtig. In dit project werden leerlingen en docenten aselect toegewezen aan drie soorten klassen: klassen met 13 tot 17 leerlingen, klassen met 22 tot 26 leerlingen en klassen met 22 tot 26 leerlingen en een klassenassistent. Het project duurde vier jaar. Aan het project namen tachtig scholen deel met in totaal 11.600 leerlingen. De belangrijkste bevinding van dit experiment was dat leerlingen in

kleinere klassen betere testcores halen, en dat de winst het grootst was voor Afro-Amerikaanse leerlingen (Krueger, 1999).

Met gecontroleerde experimenten kan overtuigend bewijs worden verkregen over de effecten van beleid. Echter, gecontroleerde experimenten zijn soms moeilijk. Het kan bijvoorbeeld lastig zijn om leerlingen aselekt een jaar extra onderwijs te laten volgen. Daarnaast kunnen experimenten soms gehinderd worden door ethische bezwaren of door restricties in tijd of geld.

Natuurlijke experimenten

Natuurlijke experimenten zijn een belangrijk alternatief voor gecontroleerde experimenten. Een natuurlijk experiment is een toevallige situatie in de werkelijkheid waardoor een aselekte controlegroep wordt gecreëerd. Deze situaties kunnen worden benut om causale effecten vast te stellen. Een fraai voorbeeld is een studie naar het effect van klassengrootte in het Israëliësch onderwijs (Angrist en Lavy, 1999). De klassengrootte in Israël is deels gebaseerd op een regel van de 12e-eeuwse geleerde Maimonides. Deze regel bepaalt dat scholen bekostiging ontvangen voor één docent zolang zij niet meer dan 40 leerlingen hebben. Bij 41 leerlingen ontvangt een school bekostiging voor twee docenten en bij 81 leerlingen voor drie docenten. Het gevolg hiervan is dat de gemiddelde klassengrootte sterk verandert rond eenheden van 40 leerlingen: bij 40 leerlingen is de gemiddelde klassengrootte 40, bij 41 is dat 20,5 en bij 81 is dat 27. Angrist en Lavy (1999) gebruiken deze variatie in klassengrootte om het causale effect op leerprestaties vast te stellen.

Het kenmerk van natuurlijke experimenten is dat een bepaalde omstandigheid toevallige variatie creëert in een bepaalde interventievariabele. In het bovenstaande voorbeeld leidt de bekostigingsregel van Maimonides tot toevallige variatie in klassengrootte. Belangrijk is dat deze variatie niet het gevolg is van selectie van deelnemers, maar voortkomt uit toevallige omstandigheden. Deze variatie kan vervolgens benut worden om het causale effect van een interventie vast te stellen. De techniek die hiervoor meestal wordt gebruikt, is instrumentele variabelen regressie. Met deze techniek wordt het effect in twee stappen geschat. In de eerste stap wordt de invloed van het natuurlijk experiment op de interventievariabele geschat. In het voorbeeld is dat de invloed van de bekostigingsregel

op de klassengrootte. In de tweede stap wordt het effect van de interventievariabele geschat op een bepaalde uitkomst. Daarbij wordt alleen de variatie in de interventievariabele gebruikt die het gevolg is van het natuurlijke experiment. Deze schatting kan geïnterpreteerd worden als het causale effect van de beleidsinterventie. Op hoofdlijnen kunnen twee soorten natuurlijke experimenten worden onderscheiden, gebaseerd op specifieke kenmerken van institutionele regels of op natuurlijke variatie.

Institutionele regels

Specifieke kenmerken van instituties kunnen aselecte controle-groepen creëren. Veel recente studies gebruiken discontinuïteiten (knippen) in regels om causale effecten te identificeren. In Nederland werd bijvoorbeeld enkele jaren geleden extra geld ingezet voor scholen met veel achterstandsleerlingen. Scholen met 70% of meer achterstandsleerlingen ontvingen extra geld, scholen met minder dan 70% achterstandsleerlingen ontvingen geen extra geld. In Leuven e.a. (2003) werd het causale effect van de extra middelen geïdentificeerd door de leerprestaties op scholen net beneden de knip van 70% te vergelijken met die op scholen net boven deze knip. Dit design heet een regressie discontinuïteits design (RD-design) en kent al een lange historie in de sociale wetenschappen. Het werd al toegepast door Campbell (1969).

Natuurlijke variatie

De tweede groep natuurlijke experimenten maakt gebruik van 'natuurlijke variatie'. Typische voorbeelden hiervan zijn twee studies van Caroline Hoxby. Het eerste voorbeeld is een studie naar (wederom) het effect van klassengrootte op leerprestaties (Hoxby, 2000a). In deze studie wordt gebruikgemaakt van demografische veranderingen. Een toename of afname van het aantal kinderen in een leeftijdscohort leidt tot variatie in de klassengrootte. Anders gezegd, de omvang van cohorten levert toevallige variatie in klassengrootte. Het tweede voorbeeld is een studie naar het effect van concurrentie tussen scholen op de leerprestaties (Hoxby, 2000b). In deze studie wordt gebruikgemaakt van variatie in concurrentie die het gevolg is van de ligging van grote rivieren. Het idee is dat grote rivieren het aantal keuzealternatieven van ouders beperken omdat

de reistijd en kosten aanzienlijk toenemen bij de aanwezigheid van grote rivieren. Deze variatie is benut om het effect op de leerprestaties te bepalen.

Op veel beleidsterreinen zijn recent studies uitgevoerd die gebruikmaken van natuurlijke experimenten, bijvoorbeeld onderwijs, zorg, veiligheid en arbeidsmarkt. Tabel 1 geeft ter illustratie een aantal voorbeelden van toepassingen van natuurlijke experimenten.

Tabel 1: Voorbeelden van studies die gebruikmaken van natuurlijke experimenten

Het causaal effect van	Op	Natuurlijk experiment	Referentie
Arbeidsongeschiktheidsverzekering	Arbeidsaanbod	Variatie in regelgeving naar tijd en regio	Gruber (2000)
Politie-inzet	Criminaliteit	Verkiezingscycli	Levitt (1997)
WW-premie	Lonen / werkgelegenheid	Wetgeving in staten	Anderson en Meyer (2000)
Hartoperaties	Gezondheid	Afstand tot ziekenhuis	McClellan e.a. (1994)
Lengte gevangenisstraf	Werkgelegenheid en inkomen	Aselect toegewezen federale rechters	Kling (1999)
Roken moeder	Geboortegewicht	Accijns op tabak	Evans en Ringel (1999)
Onderwijs	Inkomens	Geboortekwartaal	Angrist en Krueger (1991)
Onderwijs	Inkomens	Afstand tot school	Card (1995)
Onderwijs	Inkomens	Variatie in bouw van scholen naar tijd / regio	Duflo (2001)
Klassegrootte	Onderwijsprestaties	Regelgeving over maximale klassegrootte	Angrist en Lavy (1999)
Studiebeurs	Deelname hoger onderwijs	Regelgeving over studiebeurzen	Van der Klaauw (2001)

Bron: Angrist en Krueger (2002), p. 82.

Twee studies hebben betrekking op het terrein van justitie. Kling (1999) analyseert het effect van de lengte van de gevangenisstraf op de latere arbeidsmarktpositie van de delinquent in de Verenigde Staten. Het probleem daarbij is dat de lengte van de gevangenisstraf niet door toeval bepaald wordt. Voor een onderzoeker zal het lastig zijn om rekening te houden met alle verschillen tussen veroordeel-

den. Kling lost dit probleem op door gebruik te maken van verschillen tussen rechters. Sommige rechters zijn geneigd om systematisch zwaardere straffen uit te delen dan andere rechters. Dit geeft variatie in de lengte van de gevangenisstraf die onafhankelijk is van de kenmerken van de veroordeelden. Kling vindt geen effect van deze variatie op de latere arbeidsmarktpositie. De studie van Levitt (1997) analyseert het effect van politie-inzet op criminaliteit. Het probleem bij een dergelijke studie is dat de omvang van de politie-inzet in een bepaalde regio niet toevallig is. Voor een onderzoeker zal het echter lastig zijn om te controleren voor alle relevante verschillen tussen regio's. Levitt (1997) maakt gebruik van de invloed van verkiezingscycli op de omvang van politie-inzet in de Verenigde Staten. In de aanloop naar verkiezingen blijken veel politici de politie-inzet uit te breiden, ongetwijfeld omdat men verwacht dat dat gunstig is voor de uitkomst van de verkiezingen. De variatie in politie-inzet door verkiezingscycli wordt in deze studie benut om de invloed van politie op criminaliteit te meten.

Recent Nederlands onderzoek

De afgelopen jaren zijn (quasi-)experimentele onderzoeksdesigns toegepast in een aantal Nederlandse studies. Deze studies richten zich vooral op de effecten van interventies in het onderwijs of de opbrengsten van het volgen van onderwijs of training. Hieronder wordt kort besproken welke toevallige omstandigheden (natuurlijke experimenten) worden benut in deze studies.

Twee recente studies maken gebruik van een design dat sterk lijkt op een gecontroleerd experiment. Luginbuhl, Webbink e.a. (2005) maken gebruik van een random steekproef van de Inspectie van het Onderwijs. Deze studie richt zich op het effect van een schoolbezoek door de Inspectie van het Onderwijs op de prestaties van leerlingen. Het probleem daarbij is dat schoolbezoeken door de Inspectie niet random hoeven te zijn. Een inspecteur kiest de school die hij/zij wil bezoeken en daarbij kunnen vele factoren een rol spelen. Dit probleem wordt in deze studie opgelost door gebruik te maken van een speciale steekproef van de Onderwijsinspectie. Voor het jaarlijkse Onderwijsverslag trekt de Inspectie een random steekproef van scholen die worden bezocht. Daarmee kan de Inspectie een representatief beeld geven van de stand van het Nederlands onder-

wijs. Deze steekproef kan echter ook worden benut om het effect van schoolbezoeken vast te stellen. In Cornet, Vroomen e.a. (2005) wordt gekeken naar de invloed van innovatievouchers¹ op de contacten tussen bedrijven en kennisinstellingen. In deze studie wordt gebruikgemaakt van het feit dat de innovatievouchers door loting zijn toegewezen aan bedrijven, waarmee een experimentele en een controlegroep zijn gecreëerd.

Enkele studies maken gebruik van discontinuïteiten in financiële regelingen. De studie naar het extra geld voor scholen met 70% of meer achterstandsleerlingen kwam al eerder aan de orde (Leuven e.a., 2003). Leuven en Oosterbeek (2004) analyseren het effect van een fiscale aftrek voor scholing van werknemers. In de regeling was een additionele aftrek ingebouwd voor werknemers van 40 jaar en ouder. Deze discontinuïteit, het verschil in behandeling tussen werknemers jonger en ouder dan 40 jaar, wordt in deze studie benut om het effect op training van werknemers vast te stellen en om het financiële rendement van training te meten. Dobbelsesteen e.a. (2003) schatten het effect van de klassengrootte op leerprestaties door gebruik te maken van de Nederlandse bekostigingsregels voor het primair onderwijs, analoog aan de studie van Angrist en Lavy (1999).

Een andere bron voor natuurlijke experimenten betreft specifieke institutionele kenmerken. Leuven e.a. (2004) analyseren het effect van extra tijd in school op de leerprestaties van jonge leerlingen. Daarvoor wordt gebruikgemaakt van toevallige variatie in schooltijd als gevolg van schoolvakanties. Kinderen mogen naar school als ze vier jaar worden. Als de verjaardag in de zomervakantie valt, kan het enkele weken duren voordat het kind naar school kan. Deze variatie in schooltijd blijkt van invloed te zijn op de prestaties van jonge leerlingen uit achterstandsgroepen. In Wassenberg en Webink (2004) wordt gebruikgemaakt van de zogenoemde '1-oktober-regeling' voor instroom in het basisonderwijs. Scholen gebruiken deze regeling om leerlingen te laten doorstromen naar groep 3. Degenen die voor 1 oktober zeven jaar worden mogen naar groep 3, degenen die na 1 oktober zeven jaar worden mogen pas een jaar later instromen. Deze regeling blijkt van invloed te zijn op het uiteindelijk bereikte onderwijsniveau. Jonge leerlingen, degenen die nog net wel mochten instromen, halen een hoger niveau dan oudere leerlin-

1 De innovatievoucher is een tegoedbon ter waarde van 7500 euro die een bedrijf kan besteden bij een kennisinstelling.

gen. Deze toevallige variatie in onderwijsniveau wordt gebruikt om de invloed van onderwijs op beloning te meten. Ook institutionele hervormingen kunnen worden gebruikt om causale effecten vast te stellen. Webbink (2004) maakt gebruik van de verkorting van de nominale studieduur in het wetenschappelijk onderwijs in 1982 om het financieel rendement van een jaar wetenschappelijk onderwijs vast te stellen. Daarvoor worden de lonen vergeleken van degenen met respectievelijk een vijf- en vierjarige studieduur. Dezelfde benadering wordt toegepast in Oosterbeek en Webbink (2004). Deze studie richt zich op het financieel rendement van een extra jaar lager beroepsonderwijs. Daarvoor wordt gebruikgemaakt van de verlenging van verschillende opleidingen in het lager beroepsonderwijs van drie naar vier jaar in 1975.

Lessen

De belangrijkste opbrengst van de (quasi-)experimentele benadering in recente studies is dat daarmee de effecten van beleidsinterventies veel overtuigender kunnen worden vastgesteld. De vertekening in 'traditionele' studies, die geen rekening houden met de selectie van deelnemers aan interventies, kan aanzienlijk zijn. Een belangrijke vraag is hoeveel toevallige variatie kan worden gevonden om beleidsinterventies te evalueren. Michael Hurd zei al in 1990: 'Nature's stream of experiments seems limited.' Afgaande op de recente stroom aan studies lijkt Hurd's opmerking nog niet heel relevant. Bovendien kan de stroom van (natuurlijke) experimenten ook worden gemanipuleerd. In de eerste plaats door het starten van nieuwe gecontroleerde experimenten. In de tweede plaats door het handig vormgeven van beleid.

Bij de toekenning van subsidies kan bijvoorbeeld gebruik worden gemaakt van loting zoals recent is toegepast bij de toewijzing van innovatievouchers aan bedrijven. Een andere mogelijkheid is om bewust verschillen (discontinuïteiten) in te bouwen in subsidie-regelingen. Dat maakt een overtuigende evaluatie mogelijk in de toekomst. Sinds 2002 wordt 'nature's stream of experiments' actief gestimuleerd in het Amerikaans onderwijsonderzoek (Angrist, 2004). Het Amerikaanse Department of Education kiest namelijk nadrukkelijk voor het ondersteunen van onderzoek op basis van geloofwaardige onderzoeksdesigns. Experimentele designs met

aselecte toewijzing staan daarbij aan de top van de hiërarchie. In 2002 en 2003 zijn ongeveer een dozijn experimenten met random toewijzing gefinancierd, in 2000 ging het om één experiment op een totaal van 84 projecten.

Tot slot, de baten van meer overtuigende evaluaties van beleidsinterventies kunnen zeer hoog zijn. Dit geldt niet alleen voor het onderwijs- en arbeidsmarktbeleid maar voor het beleid in den brede, inclusief het justitiële werkterrein. Recent Amerikaans onderzoek geeft ook enkele toepassingen van het benutten van natuurlijke experimenten voor het evalueren van beleid. Ongetwijfeld zijn er meer mogelijkheden op dit terrein. Zo kunnen bijvoorbeeld capaciteitsrestricties op de tenuitvoerlegging van straffen of behandelingen controlegroepen creëren. Uiteraard kan ook hier selectie optreden op basis van de zwaarte van de gevallen (de zwaarste gevallen worden behandeld). In het onderzoeksdesign kan daarmee rekening worden gehouden door vooral te kijken naar 'de marginale gevallen': de groep die net wel wordt behandeld te vergelijken met de groep die net niet wordt behandeld, en daarbij te controleren voor 'de zwaarte van het geval'. Andere mogelijkheden kunnen wellicht gevonden worden in veranderingen in regels of in discontinuïteiten/knippen in financieringsregels: beneden een specifieke afkapgrens krijgt men wel middelen, onmiddellijk daarboven niet. Het benutten van 'nature's stream of experiments' en het uitvoeren van gecontroleerde experimenten kan ook op dit terrein leiden tot beter inzicht in het effect van beleidsinterventies.

Literatuur

- | | |
|--|---|
| <p>Angrist, J.D., V. Lavy
 <i>Using Maimonides Rule to estimate the effect of class size on scholastic achievement</i>
 <i>Quarterly journal of economics</i>, CXIV, 1999, p. 533-575</p> <p>Angrist, J.D., A.B. Krueger
 <i>Instrumental variables and the search for identification; from supply and demand to natural experiments</i></p> | <p><i>Journal of economic perspectives</i>, 15e jrg., nr. 5, 2002, p. 69-87</p> <p>Angrist, J.D.
 <i>American education research changes tack</i>
 <i>Oxford review of economic policy</i>, 2004, p. 198-212</p> <p>Campbell, D.T.
 <i>Reforms as experiments</i>
 <i>American psychologist</i>, 24e jrg., 1969, p. 409-429</p> |
|--|---|

- Cornet, M., B. Vroomen**
De effectiviteit van de innovatievoucher 2004
 Centraal Planbureau,
 Document nr. 95, 2005
- Dobbelsteen, S., J. Levin e.a.**
The causal effect of class size on scholastic achievement: distinguishing the pure class size effect from the effect of changes in class composition
 Oxford bulletin of economics and statistics, 64e jrg., 2002, p. 17-38
- Hanushek, E.A.**
The economics of schooling: production and efficiency in public schools
 Journal of economic literature, 24e jrg., 1986, p. 1141-77
- Frazis, H., M. Loewenstein**
Reexamining the returns to training: functional form, magnitude, and interpretation
 Bureau of Labor Statistics,
 Working Paper 367, 1999
- Hoxby, M.C.**
The effects of class size and composition on student achievement: new evidence from natural population variation
 Quarterly journal of economics, 2000a, p. 1239-1285
- Hoxby, C.**
Does competition among public schools benefit students and taxpayers?
 American economic review, 90e jrg., nr. 5, 2000b, p. 1209-1238
- Krueger, A.B.**
Experimental estimates of education production functions
 Quarterly journal of economics, 114e jrg., nr. 2, 1999, p. 497-532
- Krueger, A.B.**
Economic considerations and class size
 The economic journal, 113e jrg., 2003, F34-F63
- Leuven, E., H. Oosterbeek**
A new approach to estimate the wage returns to work-related training
 SCHOLAR working paper, 27/02, 2002
- Leuven, E., H. Oosterbeek**
Evaluating the effect of tax deductions on training
 Journal of labor economics, 22e jrg., nr. 1, 2004, p. 461-488
- Leuven, E., H. Oosterbeek e.a.**
The effect of extra funding for disadvantaged students on achievement
 SCHOLAR working paper 39/03, 2003
- Leuven, E., H. Oosterbeek e.a.**
New evidence on the effect of time in school on early achievement
 SCHOLAR Working Paper 47/04, 2004
- Levitt, S.D.**
Using electoral cycles in police hiring to estimate the effect of police on crime
 American economic review, 87e jrg., nr. 3, 1997, p. 270-290

Luginbuhl, R, H.D. Webbink e.a.

Measuring the effect of school inspections on primary school performance; a study based on cito test scores

Centraal Planbureau,
Discussion paper, 2005 (in druk)

Kling, J.

The effect of prison sentence length on the subsequent employment and earnings of criminal defendants

Woodrow Wilson School,
Economics Discussion Paper 208,
1999

Oosterbeek, H., H.D. Webbink

Wage effects of an extra year of lower vocational education; evidence from a simultaneous change of compulsory school leaving age and program length
SCHOLAR working paper 44/04,
2004

U.S. Department of Education

Identifying and implementing educational practices supported by rigorous evidence; a user-friendly guide

Institute of Education Sciences,
National Center for Education
Evaluation and Regional
Assistance, 2003

Webbink, H.D.

Returns to university education; evidence from an institutional reform

Centraal Planbureau,
Discussion Paper 34, 2004

Webbink, D.

Causal effects in education

Journal of economic surveys,
19e jrg., nr. 4, 2005, p. 535-560

Evaluatieonderzoek en het integratiebeleid

*R. Bijl**

Onlangs heeft de minister voor Vreemdelingenzaken en Integratie een convenant afgesloten met 21 zogeheten Antillianengemeenten. In dat convenant is geregeld dat de rijksoverheid deze gemeenten extra financiële middelen toekent. De gemeenten worden zo in staat gesteld om extra maatregelen te nemen om de hoge werkloosheid, de hoge schooluitval en de verhoudingsgewijs hoge criminaliteit onder de Antillianen een halt toe te roepen en liefst ook terug te dringen. Afgesproken is ook dat jaarlijks gerapporteerd zal worden over de voortgang en de behaalde successen. Evaluatie van het beleid maakt dus deel uit van de gemaakte afspraken.

Zo'n combinatie van beleid en evaluatie komen we nog niet vaak tegen op het gebied van het integratiebeleid. Waarom is dat? In deze bijdrage geven we een aantal verklaringen. En we geven een eerste aanzet op welke wijze evaluatieonderzoek in het integratiebeleid uitgevoerd zou kunnen worden, uitgaande van het bijzondere karakter van wat integratie 'is'. Immers, het object van het beleid – integratie – dient eenduidig en meetbaar omschreven te zijn, om dat beleid te kunnen toetsen.

Het doel van het integratiebeleid is 'gedeeld burgerschap'. 'Meedoen vanuit verscheidenheid, daar gaat het om'.¹ Volgens de minister kan een groep als 'geïntegreerd' worden beschouwd bij:

- beheersing van de Nederlandse taal;
- evenredige deelname aan structurele maatschappelijke terreinen;
- onderhouden van interetnische contacten;
- onderschrijven van basale Nederlandse normen.

* Dr. Rob Bijl is werkzaam bij het WODC bij de afdeling Asiel, Migratie, Integratie en Internationale aangelegenheden (ASMI).

¹ Zie de 'brief Integratiebeleid Nieuwe Stijl' die de Minister voor Vreemdelingenzaken en Integratie op 16 september 2003 aan de Tweede Kamer heeft gestuurd (TK 2003-2004, 29 203, nr. 1, p. 9).

Integratie kan volgens het kabinet bereikt worden via de weg van 'toerusting' van de migranten zodat zij de kennis en vaardigheden ontwikkelen om zich in de samenleving een positie te verwerven; via de weg van de 'toenadering' tussen migranten en autochtone bevolking, zodat migranten en autochtonen elkaar via het onderhouden van sociale contacten leren kennen en waarderen; en via de weg van 'toegankelijkheid', waarmee bedoeld wordt dat maatschappelijke sectoren zich open moeten stellen voor de migranten. Het Nederlandse integratiebeleid richt zich zowel op de verschillende categorieën 'nieuwkomers' (vluchtelingen, arbeidsmigranten, gezinsherenigers, gezinsvormers) alsook op migranten die al vele jaren in Nederland verblijven ('oudkomers') en op de tweede generatie van migranten. De verschillen tussen deze groepen in migratiegeschiedenis, motieven om te migreren, en vele andere achtergrondkenmerken (sociaal-economische en politieke omstandigheden in het land van herkomst, opleiding, enzovoort), impliceren dat verwacht mag worden dat de mate en de snelheid van integratie en de wegen waarlangs dat proces verloopt sterk uiteen kunnen lopen.

Het is duidelijk dat het evalueren van een beleid dat zo veel uiteenlopende maatschappelijke terreinen bestrijkt en zo veel doelgroepen wil bereiken hooggegrepen is. In het navolgende proberen we eerst het begrip 'integratie' concreter te beschrijven en aan te geven wat de belangrijkste aspecten zijn waar het integratiebeleid – op landelijk dan wel op lokaal niveau – op zou kunnen aangrijpen. Daarna beschrijven we op welke wijze dergelijk beleid beoordeeld kan worden op zijn kwaliteiten.

Dimensies van integratie en actoren in het integratieproces

Bij integratie draait het in de kern om het proces van het verwerven van burgerschap en het participeren in de samenleving waar de migrant zich vestigt. Met deze globale omschrijving benadrukken we het procesmatige karakter van integratie en geven we aan dat er niet een objectieve minimale of gewenste eindsituatie bestaat (Bijl e.a., 2005).

Het verwerven van burgerschap en het participeren in de samenleving bestrijkt ten minste een drietal dimensies of domeinen: legaal/politiek, sociaal-economisch, en sociaal-cultureel (zie onder meer Penninx e.a., 2004). De centrale vraag bij de legaal/politieke

dimensie van participatie en burgerschap is in welke mate allochtonen worden beschouwd en feitelijk participeren als volwaardige leden van de politieke gemeenschap. In welke mate hebben zij formele politieke rechten en plichten en kunnen zij posities verwerven zoals de autochtone bevolking? De sociaal-economische dimensie verwijst naar de sociale en economische rechten, plichten en prestaties. In welke mate hebben allochtonen (gelijke) rechten en mogelijkheden om werk te accepteren en gebruik te maken van institutionele voorzieningen om het te vinden? Hebben zij toegang tot arbeidsgerelateerde voorzieningen, zoals werkloosheidsuitkeringen en verzekeringen, en door de overheid geregelde andere socialezekerheidsvoorzieningen? De derde dimensie verwijst naar het domein van sociaal-culturele en religieuze rechten en mogelijkheden van migranten. In welke mate hebben zij (gelijke) rechten en mogelijkheden om zich te organiseren en te manifesteren als culturele, etnische of religieuze groepen? In hoeverre maken allochtone groepen deel uit van de ontvangende samenleving of blijven zij zich daarvan onderscheiden? Is er sprake van enigerlei vorm van toenadering tussen migranten en de autochtone bevolking?

Een tweede aspect is welke actoren betrokken zijn bij integratieprocessen. In feite gaat het om twee partijen: de migranten en de ontvangende samenleving. De interactie tussen deze twee bepaalt de richting en de uitkomsten van het integratieproces. Uiteraard zijn dit geen gelijkwaardige partners in termen van (politieke) macht en middelen. De ontvangende samenleving, haar institutionele structuur en haar reacties op nieuwkomers zijn veel meer bepalend voor de uitkomst van het integratieproces.

Processen van integratie spelen zich dus niet alleen af op het niveau van de individuele immigrant, waarbij diens integratie wordt gemeten in termen van huisvesting, werk en opleiding, en zijn sociale en culturele aanpassing aan en participatie in de nieuwe samenleving. Integratie vindt ook plaats op het collectieve niveau van de immigrantengroep(en). Organisaties van migranten zijn de uitdrukking van gemobiliseerde middelen en ambities, en zij kunnen een geaccepteerd deel van de samenleving worden, of zij kunnen zichzelf isoleren en zelfs uitgesloten worden door de samenleving.

Een derde niveau waarop integratieprocessen plaatsvinden, is dat van de instituties, zowel op lokaal als landelijk niveau. Algemene instituties zoals het onderwijssysteem, de arbeidsmarkt of het politieke systeem worden geacht alle burgers op gelijkwaardige wijze

te dienen. Wetten, voorschriften, maar ook ongeschreven regels en praktijken maken deel uit van deze instituties. Deze instituties kunnen echter de toegang of gelijkwaardige uitkomsten voor immigranten – formeel of informeel – bemoeilijken. Bijvoorbeeld doordat er van – informele – ‘beperkingen tot de toegang’ tot instituties sprake is of door vormen van discriminatie.

Context

Integratieprocessen spelen zich dus af op vele terreinen en er zijn vaak meerdere actoren bij betrokken. Integratie is niet een individueel fenomeen. Integratieprocessen spelen zich niet in een vacuüm af, maar in een complexe en steeds wisselende context. We geven bij wijze van voorbeeld twee belangrijke omgevingsfactoren kort weer.

Demografie

Integratie van migranten speelt zich af in een demografische context van een weliswaar afnemende, maar nog altijd relatief sterke groei van de niet-westerse bevolking als gevolg van verschillen in vruchtbaarheidspatronen (gemiddeld meer kinderen) tussen de niet-westerse en autochtone groepen. Demografische ontwikkelingen zijn van belang, omdat de macht van het getal telt – dat kan in positieve zin bijdragen aan de politieke invloed en de maatschappelijke emancipatie van de allochtone groepen; maar dat kan ook negatief uitpakken in de zin dat de autochtone bevolking de omvang van de allochtone groepen als bedreiging kan ervaren.

Relatie tussen immigratie en integratie

Er is een duidelijke relatie tussen immigratie en integratie. In de Jaarnota 2005 wordt het verband tussen integratiebeleid en immigratiebeleid sterk benadrukt daar waar het om huwelijksmigratie gaat. Gesteld wordt dat de huidige massale huwelijksmigratie onder grote groepen van de niet-westerse bevolking een belangrijke rem op de integratie vormt (Jaarnota Integratiebeleid 2005, p. 13). De conclusie die daaruit wordt getrokken, is dat er een selectief en integratiegericht immigratiebeleid nodig is. En dat maatregelen ontwikkeld worden (zoals de Wet Inburgering in het Buitenland) die tot

doel hebben garanties te bieden dat de nieuwkomers in Nederland die zich blijvend vestigen ook daadwerkelijk integreren. Behalve aan deze twee omgevingsfactoren kunnen we ook nog denken aan de economische situatie en de tolerantie in de samenleving ten opzichte van minderheden. Ook die bepalen de mogelijkheden en de kans van slagen van migranten om een positie te verwerven in de Nederlandse samenleving.

Hoe te evalueren?

Evaluatieonderzoek heeft als belangrijk doel bij te dragen aan meer 'evidence based' beleid en aan de behoefte om te weten of het gevoerde beleid opbrengt wat ervan verwacht werd toen men ermee startte. Welk beleid of welke maatregelen en interventies zijn het meest effectief, welke zijn minder effectief en welke zijn wellicht zelfs schadelijk voor de integratie? Voordat we iets zeggen over de stand van zaken in het evaluatieonderzoek op het gebied van integratie, geven we in het kort de kernbegrippen weer die bij dat soort onderzoek relevant zijn.

Evaluatieonderzoek omvat (RPE Justitie, 2004):

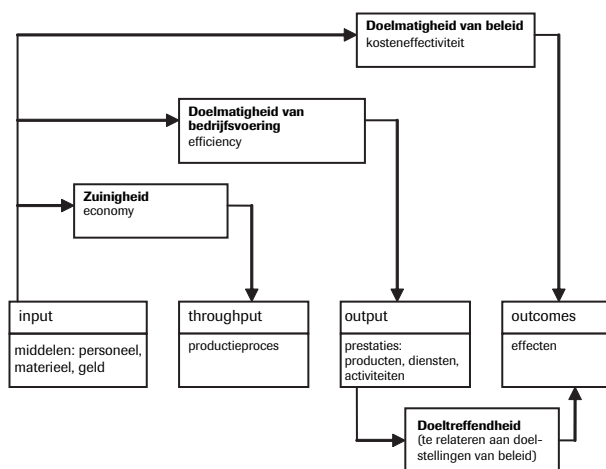
- evaluatieonderzoek ex ante: systematisch onderzoek naar de verwachte kosten en baten van mogelijke beleidsalternatieven (in termen van inzet van middelen en effecten);
- evaluatieonderzoek ex post: systematisch onderzoek naar de effecten van bestaand beleid, de wijze waarop beleid wordt uitgevoerd en/of de kosten en kwaliteit van geleverde producten en diensten;
- organisatieonderzoek: onderzoek waarbij afzonderlijke organisatieonderdelen geheel of ten dele worden getoetst op hun functioneren. Dit type onderzoek laten we hier buiten beschouwing.

In het bekende schema 1 (RPE, 2004) zijn de opeenvolgende stappen van beleid en de samenhang tussen beleidsdoelen, middelen en resultaten weergegeven.

De vier basisstappen bij beleid(sprogramma's) zijn de input, de throughput en de output. En als een punt aan de horizon wordt er ook nog wel gesproken over outcome, dat zijn de gevolgen voor de samenleving van het gevoerde beleid. In termen van het integratiebeleid kunnen we bijvoorbeeld denken aan input in termen van het aantal medewerkers van het COA of Vluchtelingenwerk Nederland

die nieuwkomers ondersteunen bij hun inburgeringsproces; bij throughput aan de feitelijke activiteiten, cursussen en dergelijke die deze organisaties verzorgen; en bij output aan de resultaten van die inspanningen: hebben de migranten een vooraf geëxpliciteerde doelstelling gerealiseerd, zijn de migranten bijvoorbeeld in staat een Nederlandse krant te lezen of zijn zij in staat om zelfstandig te wonen? Bij dit alles kan dan ook nog de doelmatigheidsafweging gemaakt worden in hoeverre er met de ingezette middelen meer bereikt had kunnen worden, dan wel met minder middelen dezelfde resultaten geboekt hadden kunnen worden. Bij outcome gaat het om de vraag of al deze genoemde activiteiten en uitkomsten resulteren in beter geïntegreerde migranten in de Nederlandse samenleving. In essentie gaat het daarbij om de causaliteitsvraag: in hoeverre heeft het gevoerde beleid geleid tot meer of betere integratie van migranten in Nederland?

Schema 1



Evaluatieonderzoek *ex ante*, dat bedoeld is om vooraf aan te geven of voorgenomen beleid of een voorgenomen interventie effect zou kunnen sorteren, is voor zover bekend op het terrein van het integratiebeleid nog niet uitgevoerd. Bij dergelijk onderzoek kan men denken aan scenario's waar diverse alternatieve ontwikkelingen worden doorgerekend en op hun consequenties en eventuele

neveneffecten worden doordacht. Dergelijk onderzoek is bij uitstek geschikt om de invloed van onzekere omgevingsfactoren mee te wegen voordat een definitieve keuze voor een bepaalde maatregel wordt gemaakt. Een bekend voorbeeld is het doorrekenen door het Centraal Planbureau van de verschillende verkiezingsprogramma's van de politieke partijen op hun financieel-economische gevolgen ('koopkrachtplaatjes').

Theoriegestuurde evaluaties

In schema 1 blijft nog duister waarom het in gang gezette beleid of de interventie effectief is. In de wereld van het evaluatieonderzoek is een krachtige opkomst van de zogeheten theoriegestuurde evaluaties te bespeuren (Pawson e.a., 2005). Het uitgangspunt bij deze benadering is dat aan beleid en interventies bepaalde (vaak impliciete) veronderstellingen ('beleids- of programmatheorie') ten grondslag liggen over de wijze waarop mensen (of organisaties) zullen reageren op de interventie gegeven de context waarbinnen zij leven. Met andere woorden, de beleidsmaker veronderstelt dat zijn beleid of interventie mechanismen kent die zullen leiden tot succes. Bij de evaluatie gaat het dan om het achterhalen en expliciteren van die 'working mechanisms' en de context waarbinnen die mechanismes werkzaam zijn. De vraag voor de evaluator is of de programmatheorie 'deugt' op grond van wat er bekend is uit de wetenschap of 'best practices' en of de gehanteerde aannamen valide zijn. De vraag is dus ook of de interventie of het beleid op zich enige kans van slagen heeft. Pas als daarover met enige zekerheid een positief antwoord gegeven kan worden, heeft het zin naar uitkomsten (effecten) te kijken.

Kwaliteitscriteria bij evaluatieonderzoek

Politici en beleidsmakers zijn primair geïnteresseerd in de vraag of de door hen geïnitieerde maatregelen 'effect hebben'. Nog los van het punt, zoals hierboven beschreven, dat het noodzakelijk is om een interventie of maatregel niet als een 'black box' te beschouwen, maar om de werkingsmechanismen helder te krijgen, steken hier moeilijke vragen de kop op. In evaluatieonderzoek dat tracht effecten vast te stellen, is het experimentele onderzoeksdesign het hoogst in aanzien, omdat dat met de meeste zekerheid uitspraken toelaat over de oor-

zaak-gevolgrelatie tussen een interventie en de gemeten effecten. In de veelgebruikte Maryland Scientific Methods Scale (Farrington, 2002) is het hoogste niveau van effectmeting een zogeheten experimenteel design met een meting vóór en een meting ná de interventie en waarbij de effecten in de groep die de interventie heeft ondergaan wordt vergeleken met die in een controlegroep en waarbij de kandidaten volstrekt willekeurig (at random) aan één van beide groepen zijn toegewezen. Het zal duidelijk zijn dat een dergelijke onderzoeksaanpak, die in de medische wereld de gouden standaard is, op het terrein van integratiebeleid ondenkbaar is. En zelfs minder strikte methodologische eisen worden in dit veld zelden gehaald.

Er zijn meerdere redenen te bedenken waarom dat zo is: landelijk beleid kent geen uitzonderingen, zodat een controlegroep construeren die het beleid niet ondergaat, onmogelijk is; en bij lokale initiatieven en interventies geldt meestal dat deze gericht zijn op gemotiveerde mensen of op een groep met een specifieke achtergrond – zoals de in de inleiding genoemde Antillianen – zodat ook daar een vergelijkbare controlegroep moeilijk te realiseren is.

Dit laatste kan worden geïllustreerd aan de hand van een recent onderzoek van het Britse Home Office (Schibel e.a., 2002). In een onderzoek waar de vraag centraal stond in hoeverre research-syntheses relevant konden zijn voor beleid gericht op de integratie van vluchtelingen (dus een beperkte doelgroep), deden de auteurs een speurtocht naar relevante studies op dat terrein vanaf 1990. In internationale databases vonden zij uiteindelijk 4481 vluchtelingenstudies. Slechts zeven studies konden worden geselecteerd omdat ze voldoende kwaliteit hadden – bijvoorbeeld de interventie helder omschreven – en daarvan hadden drie studies een quasi-experimenteel onderzoeksdesign (dat wil zeggen effecten in de interventiegroep werden vergeleken met een controlegroep). De rest werd omschreven als ‘descriptieve’ studies, waar dus niets met zekerheid was te concluderen over het mogelijke effect van de betreffende interventie.

Recent evaluatieonderzoek gericht op integratie

Er komt langzaam maar zeker meer evaluatieonderzoek op het gebied van integratie. Enkele van de meest in het oog springende actuele onderzoeken met een landelijke reikwijdte beschrijven we hier.

- Het Jaarrapport Integratie (SCP/WODC/CBS, 2005) tracht voor een tiental onderwerpen de positie van allochtonen in de samenleving te beschrijven en te analyseren. Deze tien onderwerpen – onder meer demografie, inburgering, opleidingsniveau en taalvaardigheid, onderwijs, arbeid en inkomen, sociaal-culturele integratie, criminaliteit – sluiten aan bij de beleidsprioriteiten van het kabinet. Een belangrijke plaats is ingeruimd voor opvattingen en opinies over de mate waarin de integratie van allochtonen zich ontwikkelt. Een dergelijke ‘state of the art’ is een relevante vorm van evaluatieonderzoek, maar heeft ook zijn beperkingen. Het stelt ons niet in staat om uitspraken te doen of, en zo ja, hoe sterk het gevoerde beleid heeft bijgedragen aan de huidige maatschappelijke positie van bepaalde groepen allochtonen. We weten de ‘outcome’, maar of de beleidsinput daarbij een positieve, negatieve of misschien wel helemaal geen rol heeft gespeeld, kunnen we niet met zekerheid zeggen.
- De sinds 2004 door het WODC ontwikkelde *Integratiekaart* (Bijl e.a., 2005; Van Rijn e.a., 2004) is een monitorinstrument met behulp waarvan de integratie van allochtonen van de eerste en tweede generatie door de tijd heen gevolgd kan worden. Met behulp van longitudinale data, afkomstig uit het Sociaal Statistisch Bestand van het CBS, geeft de *Integratiekaart* inzicht in het verloop van de maatschappelijke integratie van grote en kleine herkomstgroepen over een langere periode en op verschillende maatschappelijke gebieden. Dit gebeurt zowel door trends in de tijd weer te geven als door cohorten migranten die zich in een bepaald jaar in Nederland hebben gevestigd door de jaren heen te volgen om te zien hoe het hun is vergaan. Ook voor dit onderzoek geldt dat het ‘outcomes’ in beeld brengt maar niet rechtstreeks een verband kan tonen tussen het gevoerde beleid en die outcomes. Wel biedt de longitudinale aanpak de mogelijkheid om de invloed van contextuele factoren – economische, politieke of sociaal-culturele omslagen – bij de interpretatie van de gegevens mee te wegen. Zodat er als het ware ‘circumstantial evidence’ ontstaat over de al dan niet relevante invloed van die omgevingsfactoren op de maatschappelijke positie van allochtonen.
- Het op dit moment nog in uitvoering zijnde project *Evaluatie Preventiebeleid 2001-2004*, uitgevoerd door het IVA in Tilburg. Het Preventiebeleid 2001-2004 heeft tot doel te voorkomen dat jeugd uit etnische minderheidsgroepen marginaliseert of in de criminaliteit belandt. Het bouwt voort op ervaringen die zijn opgedaan

in de CRIEM-pilots die in 1999 en 2000 in acht gemeenten plaatsvonden. Gelijktijdig met de uitvoering van de projecten (232 projecten in 38 gemeentes) vindt de evaluatie van de projecten plaats. Dit gebeurt door middel van een proces- en een effect- ('output')evaluatie in de periode medio 2003 tot en met medio 2006. Het onderzoek loopt op dit moment dus nog. De effectevaluatie vindt plaats bij een beperkt aantal projecten die zich richten op een tweetal thema's: reduceren van overlast en criminaliteit, voorkomen van schooluitval op gebied van integratie. Interessant, maar tegelijk voor een evaluatie ingewikkeld, is dat de vele lokale initiatieven zich op uiteenlopende doelgroepen (zowel jongeren zélf als intermediairen) richten en zeer uiteenlopende benaderingen gebruiken. Omdat dit evaluatieonderzoek zich richt op heel concrete integratieprogramma's, kan het de 'throughput' in kaart brengen. Worden de programma's uitgevoerd op de wijze zoals bedoeld was ('programma-integriteit'); wat zijn de (veronderstelde) werkingsmechanismen? Dergelijke informatie is van het grootste belang om in 2006 met zekerheid te kunnen stellen of de vele lokale programma's hebben bijgedragen aan een reductie van criminaliteit en een verminderde schooluitval.

- Ontwikkeling van een lokale *Integratiekaart*. Ook dit project is nog in uitvoering. Omdat integratiebeleid zich vooral op lokaal niveau afspeelt, is er een grote behoefte bij gemeenten om hun inspanningen te kunnen toetsen. Het project heeft tot doel om zowel de positie van allochtonen op relevante maatschappelijke terreinen te beschrijven en te kunnen monitoren in de tijd als ook de op gemeentelijk niveau uitgevoerde projecten en beleidsinspanningen in kaart te brengen. Om gemeenten met elkaar te kunnen vergelijken wat betreft de uitkomsten van hun beleidsinspanningen, is het noodzakelijk een eenheid van taal te creëren. Om die reden is aansluiting gezocht bij de definities en operationalisering van begrippen van de landelijke *Integratiekaart*.
- Hoewel op dit moment de lijst met kwalitatief hoogstaand evaluatieonderzoek op het gebied van integratie nog kort is, is het van belang om een goed overzicht te hebben welke onderzoeken er de afgelopen jaren in Nederland in opdracht van de landelijke overheid zijn uitgevoerd op het gebied van integratie, wat de methodologische kwaliteit daarvan was, en welke beleidstheorieën/veronderstelde werkzame mechanismen wat betreft (succesvolle) integratie daar zijn uit te halen. Het WODC is recent met

een dergelijke *researchsynthese* gestart. Het doel is dus niet om concrete beleidsmaatregelen of interventies te evalueren, maar het onderzoek te evalueren dat daarnaar in het verleden is gedaan. Dat kan ons inzicht geven in zowel de (on)mogelijkheden om goed evaluatieonderzoek naar integratie uit te voeren als in de tot op heden gehanteerde opvattingen over wat waarom effectief is om integratie te bevorderen.

Evaluatieonderzoek van integratiebeleid: kansen

In toenemende mate komt kwantitatieve informatie beschikbaar over de positie van allochtonen in de Nederlandse samenleving. Er komen steeds meer mogelijkheden om bestaande registraties en andersoortige databestanden van uiteenlopende organisaties aan elkaar te koppelen. Het Sociaal Statistisch Bestand van het CBS is daarvan een krachtig voorbeeld. Overigens is enige bescheidenheid wel op zijn plaats: voor belangrijke onderwerpen die we in de inleiding genoemd hebben, zijn nu nog weinig tot geen landelijke gegevens beschikbaar. Op het gebied van de structurele integratie (arbeidsmarkt, onderwijs) zijn veel data voorhanden, maar over de sociaal-culturele integratie (sociale contacten, houding en tolerantie) én over de wederkerigheid van het integratieproces tussen migrant en Nederlandse samenleving is veel minder informatie beschikbaar. Aan de andere kant is het gegeven dat met het beschikbaar komen van longitudinale gegevens we steeds beter in staat zijn om processen van integratie in de tijd te volgen, een heel mooie ontwikkeling. Want dat maakt het mogelijk om de 'integratiecarrières' van verschillende groepen migranten met elkaar te vergelijken: wie is op welk terrein succesvol, welke groep stagneert op welk terrein, zien we verschillen in de snelheid waarmee verschillende groepen nieuwkomers een plek weten te vinden in de Nederlandse samenleving? Deze informatie ondersteunt het beleid in het formuleren van betere doelgroepgerichte 'maatwerk'-maatregelen.

Behalve dat de rijkdom aan beschikbare 'outcome'-gegevens toeneemt, ontstaat er langzaam maar zeker ook een grotere eenduidigheid van begrippen en definities. Dat is een grote winst in een veld waarin het centrale concept 'integratie' een allesbehalve heldere inhoud heeft.

Voor onderzoek betekent dit dat de vergelijkbaarheid van onderzoeksresultaten groter wordt. Als uniforme definities bestaan over, bijvoorbeeld, wat werkloosheid is, kan een adequate vergelijking gemaakt worden tussen de uitkomsten van onderzoeken naar de arbeidsmarktpositie van verschillende herkomstgroepen. Eenheid van taal is essentieel.

Om beleid te kunnen evalueren dienen de doelen van dat beleid 'SMART' te zijn geformuleerd. SMART staat voor specifiek, meetbaar, acceptabel, realistisch en tijdgebonden. De eerste vraag is dan of het integratiebeleid, de integratieprogramma's en de concrete interventies voldoende SMART zijn geformuleerd. Wat betreft arbeidsmarktparticipatie van allochtonen zijn er bijvoorbeeld streefpercentages geformuleerd die binnen een vastgestelde periode gerealiseerd moeten worden. Een dergelijk goed voorbeeld zien we echter nog maar zelden. Bij veel integratieprojecten gaat het om wat diffuse en impliciete doelstellingen als migranten ergens bij betrekken, begrip kweken voor elkaars situatie, ondersteunen bij toetreding tot de arbeidsmarkt, en dergelijke. Wel sympathiek, maar weinig concreet vast te stellen. Het gevolg is dat een goede evaluatie van alle inspanningen vaak niet mogelijk is.

Een mogelijkheid om beleids- of interventiedoelstellingen voldoende SMART te krijgen, is om vaker een ex ante evaluatie uit te laten voeren. Dat type onderzoek is primair bedoeld om diverse alternatieve ontwikkelingen door te rekenen op hun consequenties en eventuele neveneffecten te doordenken. Maar het kan ook bijdragen – op een moment dat het er nog toe doet – aan de verheldering en aanscherping van de beoogde doelen, op een zodanige wijze dat een proces- en effectevaluatie mogelijk wordt.

Ex ante evaluatie kan ook vaker ingezet worden om te verhelderen wat de beleidstheorie is van het uit te voeren beleid: waarom zal het werken, en zal het ook over langere tijd nog steeds werken als de initiatiefnemers zijn vertrokken of het sociale klimaat in Nederland ten opzichte van allochtonen is verslechterd? En heeft de beleidsmaker nagedacht over de mogelijkheid dat géén beleid ook wel eens effectief zou kunnen zijn? Op het gebied van integratie zouden wat vaker Integratie Effect rapportages uitgevoerd moeten worden, vergelijkbaar met de al bestaande milieu-effectrapportages en de emancipatie-effectrapportages. In de kern is ex ante evaluatie simpelweg 'eerst denken dan doen'.

Aan antwoorden op dit soort basale vragen kunnen evaluatieonderzoekers bijdragen op basis van hun kennis van menselijk en institutioneel gedrag, zonder op de stoel van de beleidsmaker te gaan zitten. De rol van de onderzoekers is het stellen van de goede kritische vragen, zodat er beleid of interventies ontwikkeld worden die én kans van slagen hebben én goed te evalueren zijn.

Bescheiden pretenties

Zoals gezegd zijn experimentele of quasi-experimentele aanpakken van evaluatieonderzoek zelden mogelijk op het terrein van integratie (met een enkele internationale uitzondering op het gebied van taalverwerving en arbeidsmarktkansen: Delander e.a., 2005). Óf een maatregel bestrijkt meteen de hele bevolking, en laat geen ruimte om eerst eens op kleinere schaal – en dus met de mogelijkheid een controlegroep te volgen – te bekijken of de maatregel effectief is. Of een maatregel of interventie is gericht op (de situatie van) een heel specifieke groep waarbij het niet goed mogelijk is een vergelijkbare controlegroep te bedenken. Het evaluatieonderzoek moet wat bescheidener zijn in zijn pretenties.

Er is in het integratieonderzoek behoefte aan een methodologie die een kwantitatieve en kwalitatieve benadering weet te verenigen (Ager en Strang, 2004b). Want ook al zijn er nu landelijk steeds meer kwantitatieve ‘outcome’-gegevens beschikbaar die de positie en de carrières van allochtonen op verschillende maatschappelijke gebieden beschrijven, inzicht in de oorzaak-gevolgrelatie tussen een beleidsmaatregel of een programma en die maatschappelijke positie vereist andere informatie.

De op zich terechte aandacht voor ‘harde’ cijfers mag er niet toe leiden dat kwalitatieve informatie verkregen via diepte-interviews, groepsdiscussies, Delphi-panels, en dergelijke – uiteraard uitgevoerd conform wetenschappelijke standaarden – als bron van evaluatie vergeten wordt.

Daarom is een wedstrijd als ‘Parels van de Integratie’ interessant. Op 28 november 2005 zijn de prijzen uitgereikt in het kader van een door het Kennisnet Integratiebeleid en Etnische Minderheden (KIEM, www.integratie.net/kiem) uitgeschreven wedstrijd voor de beste lokale integratieprojecten. 62 inzendingen werden na een oproep voor het inzenden van toonaangevende projecten door 26 gemeenten ingediend. Een zogenoemd ‘Comité van Parelvisser’s’

heeft uit deze grote variëteit aan praktijkgerichte projecten acht 'Parels van integratie' geselecteerd. Bij zijn selectie lette het comité in de eerste plaats op de mate waarin de projecten effectief en spannend zijn. Vervolgens is kritisch gekeken naar de uitvoerbaarheid en kopieerbaarheid van de projecten. De gebruikte criteria mogen dan weliswaar voor evaluatieonderzoekers niet 'wetenschappelijk' klinken, het is toch de moeite waard om dergelijke *best practices* goed te bekijken. Want zij hanteren mogelijke andere relevante criteria om de effecten van beleid of projecten te beoordelen. Criteria die dan wel niet overeenkomen met Maryland-standaarden, maar wel de sociale processen belichten die van belang zijn bij het al dan niet welslagen van een initiatief.

Evaluatieonderzoek van integratiebeleid en van integratieprojecten begint van de grond te komen. In het voorgaande zijn enkele kernproblemen van dat onderzoek beschreven en zijn enkele richtingen aangegeven hoe die problemen zouden kunnen worden aangepakt. Het zou goed zijn als er een op de integratiesector toegeschreven handleiding komt voor goed opgezette (en dus te evalueren) lokale en nationale integratie-initiatieven. Zodat gemeenten en landelijke overheid die kunnen hanteren voordat zij groen licht/geld geven voor nieuwe initiatieven.

Literatuur

Ager, A., A. Strang

Indicators of integration; final report

Londen, Home Office, Home Office Development and Practice report 28, 2004a

Ager, A., A. Strang

The experience of integration; a qualitative study of refugee integration in the local communities of Pollokshaws and Islington

Londen, Home Office, On-line Report 55/04, 2004b

Bauböck, R., A. Heller (red.)

The challenge of diversity; integration and pluralism in societies of immigration
Avebury, Aldershot, 1996

Bijl, R.V., A. Zorlu e.a.

Integratiekaart 2005; de maatschappelijke integratie van migranten in de tijd gevolgd: trend- en cohortanalyses
Den Haag, WODC, 2005

Delander L., M. Hammarstedt e.a.

Integration of immigrants; the role of language proficiency and experience

- Evaluation review, 29e jrg., nr. 1, 2005, p. 24-41
- Entzinger, H., R. Biezeveld**
Benchmarking in immigrant integration; report to the European Commission
Rotterdam, Erasmus Universiteit/ERCOMER, 2003
- Esser, H.**
Does the new immigration require a new theory of intergenerational integration?
Mannheim, Universität Mannheim/Mannheimer Zentrum für Europäische Sozialforschung, Working paper nr. 71, 2003
- Europese Commissie**
Mededeling van de Commissie over immigratie, integratie en werkgelegenheid
Brussel, EU, COM(2003)336, 3 juni 2003
- Farrington, D.P., D.C. Gottfredson e.a.**
The Maryland Scientific Methods Scale
In: Sherman, L. e.a. (red.), *Evidence based crime prevention; chapter 2*, New York, Routledge, 2002
- Hagendoorn, L., J. Veenman e.a. (red.)**
Integrating immigrants in the Netherlands; cultural versus socio-economic integration.
Aldershot, Ashgate, 2003
- Kennisnet Integratiebeleid en Etnische Minderheden
KIEM, www.integratie.net/kiem
- Tweede Kamer**
Rapportage Integratiebeleid Etnische Minderheden 2003
Den Haag, Tweede Kamer, Kamerstukken II, 2003-2004, 29 203, nr. 1
- Tweede Kamer**
Jaarnota Integratiebeleid 2005
Den Haag, Tweede Kamer, Kamerstukken 2005-2006, 30 304, nrs. 1 en 2
- Ministerie van Justitie, Directie Financieel-Economische Zaken**
Regeling Prestatiegegevens en Evaluatieonderzoek Justitie 23 juli 2004
Den Haag, Ministerie van Justitie, 2004
- Pawson, R., T. Greenhalgh e.a.**
Realist review; a new method of systematic review designed for complex policy interventions
Journal of health services research and policy, 10e jrg., 2005, supplement 1, p. 21-34.
- Penninx, R., K. Kraal e.a.**
Citizenship in European cities
Ashgate, Aldershot, 2004
- Rijn, A.S. van, A. Zorlu e.a.**
De ontwikkeling van een Integratiekaart
Den Haag, WODC/CBS, 2004
- Schibel, Y., M. Fazel e.a.**
Refugee integration; can research synthesis inform policy?
Londen, Home Office/RDS, RDS
On-line report 13/02, 2002
- SCP/WODC/CBS**
Jaarrapport Integratie
Den Haag, SCP/WODC/CBS, 2005

De rol van evaluaties in het begrotingsproces

*A.H.M. de Jong en D.L. Kabel**

‘Wat krijgt de belastingbetaler terug voor het geld dat de overheid uitgeeft?’ Dat is een belangrijke vraag waarop het antwoord door het Ministerie van Financiën gezocht wordt. Voor een deel van de begroting is het directe effect van de besteding van belastinggeld nog vrij gemakkelijk aan te geven: er worden bijvoorbeeld onderwijzers, politieagenten en wegen van betaald. Wat precies de beoogde effecten zijn, is echter bij lang niet alle overheidsprogramma’s duidelijk, laat staan dat helder is of de inzet van overheidsmiddelen de beoogde effecten realiseert. Beleidsevaluaties en in het bijzonder effectonderzoek zijn daarvoor onmisbare instrumenten.

Deze bijdrage gaat over de rol in het begrotingsproces van evaluaties in het algemeen en effectevaluaties in het bijzonder. Gestart wordt met het aangeven van het belang van beleidsevaluaties voor de verantwoording van belastinggeld en het leren van ervaringen uit het verleden. Vervolgens wordt aangegeven hoe het gesteld is met de kwaliteit van het evaluatieonderzoek en wat de barrières zijn om hier verbetering in aan te brengen. Tot slot worden de acties beschreven om hierin verandering te brengen.

Beleidsevaluatie en VBTB

Met de groeiende aandacht voor de verantwoording over het gevoerde overheidsbeleid wordt de laatste jaren de vraag naar effecten van beleid steeds nadrukkelijker gesteld. Dat is een internationale trend, die zich in Nederland heeft gemanifesteerd in het VBTB-project: Van Beleidsbegroting naar Beleidsverantwoording. In het

* Drs. André de Jong en drs. Dick Kabel zijn werkzaam bij het ministerie van Financiën. Dit artikel is een bewerking van de inleiding die de eerste auteur verzorgde op het congres *(Quasi-)experimental evaluations and Dutch society; trends and developments in criminal justice and prevention policies, social integration and education*.

kader hiervan is een begroting nieuwe stijl ontwikkeld waarbij het startpunt niet het geld is dat wordt uitgegeven, maar het beleid en wat men daarmee wil bereiken. De VBTB-operatie was primair bedoeld om het voor de Tweede Kamer eenvoudiger te maken de begroting en het jaarverslag met elkaar te vergelijken en vervolgens het beleid van het kabinet te beoordelen. Inmiddels is VBTB uitgegroeid tot een instrument dat de doelmatigheid en doeltreffendheid van beleid helpt te bevorderen.

Opvallend bij de aanvang van het VBTB-project – en van vergelijkbare projecten in het buitenland – zijn de grote verwachtingen van het definiëren en operationaliseren van prestatie-indicatoren. Op zich blijkt het in een groot aantal gevallen nog wel mogelijk om de doelstellingen van overheidsbeleid te kwantificeren: de overheid streeft naar meer werkgelegenheid, meer veiligheid op straat, meer verkeersveiligheid enzovoort. Het voordeel van dit type kwantificeerbare indicatoren is dat kan worden vastgesteld of een doelstelling is gerealiseerd. Een risico is echter dat de indicator het beleid gaat bepalen. Onbedoelde effecten kunnen zijn dat de werkelijkheid te eenvoudig wordt voorgesteld of dat er gestuurd wordt op prestaties die eigenlijk niet bedoeld zijn. Zo zegt het aantal uitgeschreven boetes weinig over de veiligheid in het land. Daardoor ontstaat het risico van een perverse werking van prestatiemeting.

Een andere beperking van de meeste indicatoren is dat deze weinig zeggen over de *bijdrage* van de overheid aan het bereiken van die doelstellingen (ook wel de netto-effecten genaamd). Daarvoor is evaluatieonderzoek noodzakelijk. In evaluatieonderzoek gaat het om lastige vragen. In de eerste plaats moet worden onderzocht of de doelstelling is bereikt *dankzij* het gevoerde beleid (en niet door de conjunctuur enzovoort). Die lastige vraag kan alleen worden beantwoord als aannemelijk kan worden gemaakt wat er zou zijn gebeurd zonder – of met ander – beleid. Of zoals de Amerikanen en Engelsen zeggen: er is een *counterfactual* nodig. Dat kan bijvoorbeeld een controlegroep zijn die bij voorkeur alleen verschilt van de experimentele groep wat betreft de uitgevoerde interventie. Hierbij heeft een controlegroep die geen enkele interventie ondervindt de voorkeur boven een andere interventie.

Om beleid goed te kunnen beoordelen, dienen niet alleen de beoogde effecten maar ook de neveneffecten van beleid in beeld gebracht te worden, die al dan niet gewenst zijn. Bij de evaluatie van huursubsidie bijvoorbeeld, gaat het niet alleen om het effect op de

betaalbaarheid van het wonen, maar ook bijvoorbeeld om het effect op de armoedeval.

Goed evaluatieonderzoek is dus van vitaal belang voor de verantwoording van belastinggeld, maar ook om te leren van ervaringen.

Wat werkt, wat werkt niet en hoe kan het beter?

Beoordeling huidige situatie

De afgelopen decennia heeft de hoeveelheid beleidsevaluaties in Nederland een grote vlucht genomen. Bijna dertig jaar geleden constateerde Hoogerwerf in zijn oratie *Heeft beleid effect*, dat er zo goed als helemaal niets aan effectonderzoek werd gedaan. In 1992 constateerde Leeuw (1992) nog dat er te weinig werd geëvalueerd. Dat beeld is inmiddels drastisch veranderd: tegenwoordig registreert het Evaluatie Overzicht Rijksoverheid (EOR) ongeveer 250 evaluaties per jaar.

Er wordt dus veel geëvalueerd, maar hoe zit het met de kwaliteit van de evaluaties en wat wordt er van die evaluaties geleerd? Om een indruk te krijgen van de kwaliteit van het evaluatieonderzoek is in het kader van de VBTB-evaluatie een bescheiden steekproef van twintig evaluatierapporten bekeken. Daaruit blijkt dat het gros van het evaluatieonderzoek kwalitatief van aard is, en naast documentanalyse, grotendeels op interviews en enquêtes met betrokkenen gebaseerd. De evaluatieonderzoeken zijn sterk instrumentgericht. Het merendeel van het evaluatieonderzoek is niet gericht op de effecten van beleid, en geen enkele studie slaagt erin om de effecten in beeld te brengen. Daarvoor is het noodzakelijk – zoals hiervoor aangegeven – dat aangegeven wordt wat er zou zijn gebeurd als het beleid niet zou zijn uitgevoerd. Het ontbreken van de *counterfactual* wordt niet toegelicht en ook niet gerechtvaardigd. Uit gesprekken met de evaluatiedeskundigen die de kwaliteit van het evaluatieonderzoek hebben getoetst kwamen de volgende andere problemen naar voren:

- Onduidelijke beleidsdoelstellingen. Beleid wordt geëvalueerd waarvan het beoogde doel onduidelijk is (of vaag). Op zich kan een onderzoeker dan nog wel proberen de effecten van een instrument in beeld te brengen, maar zonder een goed geformuleerde doelstelling is het lastig om de relevante effecten te meten en kunnen deze niet worden beoordeeld.
- Onvoldoende borging onafhankelijkheid. Deskundigen twijfelen aan voldoende onafhankelijkheid bij de onderzoeken.

- Onvoldoende kwaliteit van de rapportage. De kwaliteit van de onderzochte rapporten maakte het moeilijk om de kwaliteit van het onderzoek te beoordelen en deed ook twifelen aan de kwaliteit van het onderzoek.

Zorgwekkend is ook dat uit de VBTB-evaluatie naar voren komt dat de rol van evaluaties in het beleidsproces – uitzonderingen daargelaten – marginaal is (zie IOFEZ, 2004). Uit interviews blijkt dat beleidsmakers veel belang hechten aan beleidsevaluaties, maar dat de praktijk minder positief is. Evaluaties worden als plichtmatig ervaren. Er zouden nauwelijks lessen worden getrokken uit evaluaties en deze zouden bijna nooit onderwerp van debat zijn. Beleidsevaluaties kunnen een grotere rol spelen in het begrotingsproces als de kwaliteit en relevantie toeneemt. En dat kan alleen als de aandacht van beleidsmakers hiervoor toeneemt.

Evalueren van beleid is een lastig vak

Een belangrijk obstakel bij de evaluatie van beleid is dat beleidsdoelstellingen vaak lastig evalueerbaar zijn. Dat kan te maken hebben met de aard van het beleidsterrein – hoe stel je bijvoorbeeld de effecten vast van het Nederlandse buitenlandse beleid of van de Nederlandse defensie? – maar ook met de wijze waarop beleidsdoelstellingen totstandkomen. Beleid is het resultaat van een politiek proces en bij tegengestelde opvattingen zal de formulering van beleidsdoelstellingen het resultaat zijn van een politiek compromis. De politieke context waarin beleid totstandkomt, kan er verder toe leiden dat beleid voortdurend in verandering is, en dat er geen tijd is om af te wachten wat de resultaten zijn van het gevoerde beleid. Ook zal het als lastig worden ervaren als de conclusies van evaluatieonderzoek geen ondersteuning bieden voor het gevoerde beleid. Een complicerende factor van meer praktische aard is dat vaak pas over evaluatie wordt nagedacht tegen de tijd dat het beleid geëvalueerd moet worden. En dan kan het te laat zijn om de daarvoor benodigde data te verzamelen. Daarmee komen we op de oorzaken van gebrekkige evaluaties binnen het bereik van de opdrachtgevers en uitvoerders van evaluaties. Bij de opdrachtgevers binnen de departementen is vaak onvoldoende (concentratie van) deskundigheid en wordt de beschikbare capaciteit teveel uitgesmeerd.

Mede daardoor is er nog te weinig besef dat voor het meten van de effecten van beleid op de een of andere manier de *counterfactual* moet worden geconstrueerd. Om toch tot een evaluatie van beleid te komen wordt vaak de toevlucht genomen tot monitoring en procesevaluaties. Een groot probleem is ook dat onderzoek vaak wel wordt uitbesteed, maar dat er weinig garanties zijn ingebouwd dat de degenen die het onderzoek doen tot een onafhankelijk oordeel kunnen komen.

Verbetering van de kwaliteit en de relevantie van evaluatieonderzoek

Gegeven de huidige staat van de beleidsevaluatie wil het kabinet langs drie wegen komen tot betere evaluaties (zie Tweede Kamer, vergaderjaar 2004–2005, 29 949, nr. 1):

- Vooraf dient beleid beter onderbouwd te worden. Dat betekent dat bij de formulering van beleid het antwoord beschikbaar moet zijn op vragen als: wat is het maatschappelijk probleem dat moet worden opgelost? Wat is de verantwoordelijkheid van de overheid daarbij? Welke instrumenten worden ingezet? Hoe dragen die bij aan de oplossing van het probleem? Welke effecten mogen van beleid worden verwacht? Hoe kan dat worden gemeten? Wat moet daarvoor gebeuren?
- Achteraf dienen beleidsterreinen integraal te worden doorgelicht. Dat dient bij voorkeur te geschieden op het niveau van beleidsdoelstellingen zoals die zijn terug te vinden in de begroting, en niet op het niveau van instrumenten. Bij beleidsdoorlichtingen is het de bedoeling dat een aantal simpele, doch fundamentele vragen wordt beantwoord. Naast de vragen die – als het goed is – al zijn beantwoord bij de onderbouwing van beleid – dient inzicht te worden gegeven in de uitvoering en de effectiviteit van beleid. Hoe is het beleid uitgevoerd en wat waren daarvan de effecten? Beleidsdoorlichtingen kunnen door de beleidsmakers worden opgesteld, maar cruciaal is dat op de een of andere wijze een kritische blik wordt georganiseerd om te voorkomen dat beleidsevaluatie beperkt blijft tot een activiteit waarbij de slager zijn eigen vlees keurt.
- Naast de beleidsdoorlichtingen – en idealiter als onderdeel hiervan – dienen daar waar het mogelijk en nuttig is evaluaties naar de effecten van beleid te worden verricht. De beste manier om de

effecten van beleid te identificeren is met (quasi-)experimenten. Langs die weg kan namelijk zeker worden gesteld dat de waargenomen effecten van beleid ook inderdaad het gevolg zijn van beleid en niet van andere factoren.

Dit artikel gaat verder over de laatste vorm van evaluatieonderzoek. Het voorgaande laat zien dat het stimuleren van evaluatieonderzoek naar de effecten van beleid onderdeel is van een bredere strategie.

Mogelijkheden en beperkingen van effectevaluaties

De kracht van onderzoek naar effecten van beleid – kortweg: effect-evaluaties – is dat kennis ontstaat over wat wel werkt en wat niet. Met effectonderzoek kan kennis worden verkregen over de effectiviteit van concrete maatregelen en van goed af te bakenen onderdelen van beleid. Dat houdt tegelijkertijd een beperking in. Het is namelijk lastig om voor bijvoorbeeld het hele onderwijsstelsel te komen tot een geloofwaardige *counterfactual*. Dat zou inzicht vereisen in de onderwijsresultaten in de situatie dat de overheid alles zou regelen, of juist niets. Wel is het mogelijk om de onderdelen te evalueren die cruciaal zijn voor de werking van het stelsel. De kunst is om die onderdelen goed te selecteren om te komen tot kennis die ook relevant is. Een kwetsbare eigenschap van effectonderzoek is verder dat het onderzoek niet duidelijk maakt waarom een instrument bepaalde resultaten oplevert. Als dat niet op de een of andere manier aannemelijk kan worden gemaakt, is het risico groot dat het onderzoek als niet relevant terzijde wordt geschoven (zie Leeuw, 2005). Een andere belangrijke eigenschap van effectonderzoek is dat de uitkomsten tijd-, plaats-, en persoonsgebonden zijn. Hetzelfde programma voor een andere jaargang deelnemers kan andere resultaten opleveren, net als de resultaten van hetzelfde programma op een andere manier geanalyseerd tot andere resultaten kan leiden (zie Fay, 1996). Daarom is het noodzakelijk om de uitkomsten van evaluaties te generaliseren met behulp van metastudies. De activiteiten van de *Campbell collaboration* kunnen in dit verband van grote waarde zijn. Twee laatste beperkingen zijn dat wél naar het effect, maar slechts zelden naar de verhouding tussen kosten en baten wordt gekeken, laat staan naar het meest doelmatige instrument en dat uitsluitend de effecten voor de deelnemers in beeld

worden gebracht. Het laatste is een belangrijke beperking als de indirecte effecten (neveneffecten) van groot belang zijn. Zo dienen bijvoorbeeld arbeidsmarktprogramma's gecorrigeerd te worden voor substitutie, verdringing en indirecte effecten via de belastingen en de loonvorming.

Deze kanttekeningen zijn bedoeld om te benadrukken dat niet alle heil kan komen van effectevaluaties, hoe perfect dan ook uitgevoerd, maar dat deze onderdeel moeten vormen van een bredere evaluatiestrategie, waarin ook zorg wordt gedragen dat lessen van beleidsevaluatie worden benut.

Beschikbaarheid en effectevaluaties

Het valt op dat we voor effectevaluaties grotendeels zijn aangewezen op de VS, Canada, en het Verenigd Koninkrijk, terwijl de overheidsuitgaven in die landen vergeleken met continentaal Europa juist laag zijn. In de Angelsaksische landen lijkt er met andere woorden een grotere bereidheid te bestaan om kritisch te kijken naar de effecten van overheidsbeleid. Een schoolvoorbeeld daarvan is het arbeidsmarktbeleid. In Nederland wordt daar relatief veel geld aan besteed, terwijl Nederland voor de kennis van de effectiviteit van de gehanteerde instrumenten grotendeels op het buitenland was aangewezen (zie IBO-werkgroep Arbeidsmarktbeleid, 2000). Inmiddels worden door het Ministerie van Sociale Zaken stappen gezet om hier verandering in te brengen.

Maar ook op andere terreinen worden initiatieven genomen om tot meer effectevaluaties te komen. Zo sprak recentelijk de minister van Onderwijs, Cultuur en Wetenschappen tijdens een Oeso-congres de volgende woorden: 'Let practice prove policy right and wrong. Of course, we need to look for evidence that our measures are really working. But it's just as important to learn from things that go wrong. For this we need good experiments as part of the policy process. They are essential if we are to put taxpayers' money to good use.'

En geleidelijk komen er ook op onderwijsterrein voorbeelden beschikbaar van beleidsrelevante effectevaluaties. Een mooi voorbeeld is een effectevaluatie waaruit blijkt dat een extra subsidie (belastingaftrek) voor werkgevers bij scholing van oudere werknemers een positief effect heeft. Maar het effect wordt veroorzaakt door uitstel van scholing en niet door de subsidiemaatregel (Leuven en Oosterbeek, 2004).

Dit soort studies smaakt naar meer. Maar ook op andere terreinen zijn er hoopvolle signalen, bijvoorbeeld op justitieterrein. Om recidive te voorkomen van gedetineerden bekostigt het Ministerie van Justitie gedragsinterventies die erop zijn gericht om het gedrag zodanig te beïnvloeden dat recidive wordt voorkomen. Gedragsinterventies kunnen worden aangeboden als voorwaarde of als onderdeel van de strafrechtelijke sanctie of maatregel. Om de kwaliteit van gedragsinterventies te kunnen vaststellen, heeft het Ministerie van Justitie de Erkenningscommissie Gedragsinterventies Justitie ingesteld. Deze commissie beoordeelt of gedragsinterventies (kunnen) leiden tot vermindering of het voorkomen van recidive. Het is de bedoeling dat negatief beoordeelde gedragsinterventies niet meer worden uitgevoerd. Justitie zal in de toekomst alleen positief beoordeelde gedragsinterventies financieren.

Maar ook op heel andere terreinen zijn er inspirerende voorbeelden. Zo zijn in Nederland door de SWOV (Stichting Wetenschappelijk Onderzoek Verkeersveiligheid) de effecten van rotondes onderzocht op de verkeersveiligheid. Deze bleken aanzienlijk te zijn: 50% minder ongelukken en 75% minder gewonden, hetgeen ertoe heeft geleid dat steeds meer kruispunten worden omgezet in rotondes (circa 1500 in 2004).

Hoe tot meer effectevaluaties te komen?

Het simpelweg voorschrijven van meer effectevaluaties lijkt in ieder geval weinig zinvol om tot meer en kwalitatief betere effectevaluaties te komen. Een dekkende programmering van effectonderzoek is immers niet mogelijk en het is a priori niet duidelijk waar effectevaluaties wel mogelijk zijn en welke effectevaluaties veel toegevoegde waarde hebben. Wel is duidelijk dat je daarvoor onderzoekers nodig hebt die dit type onderzoek kunnen uitvoeren en beleidsmakers die het beleid kennen. En dat niet alleen: onderzoekers en beleidsmaker moeten elkaar ook nog weten te vinden.

Gelukkig beschikt Nederland over onderzoekers die dit type onderzoek kunnen en willen doen, dus daar bestaat vooralsnog geen bottleneck. Het lijkt daarom het beste om de aandacht in eerste instantie te richten op het stimuleren van de vraag naar dit type onderzoek. De aanbesteding, de uitvoering en het begrijpen van goede effectevaluaties vereist immers de nodige kennis, tijd en

geld. Dat maakt het waarschijnlijk dat effectevaluaties alleen van de grond komen als beleidsmakers gemotiveerd raken om meer effectonderzoek te vragen. Dit is een belangrijke reden om in de begrotingen en jaarverslagen – waarin overzichten zijn opgenomen van uitgevoerd evaluatieonderzoek naar de doelmatigheid en doeltreffendheid van beleid – de effectevaluaties als afzonderlijke categorie zichtbaar te maken. Op die manier worden de uitgevoerde effectevaluaties in het zonnetje gezet en kunnen ze als inspiratie dienen voor andere beleidsterreinen. Om dat te ondersteunen is het de bedoeling dat de onderzoeksrapporten via internet eenvoudig zijn te downloaden. Tegelijkertijd maken de overzichten duidelijk op welke terreinen er veel gebeurt en waar er nog weinig gebeurt. Het aandacht vestigen op de goede voorbeelden is met andere woorden een eerste weg om het uitvoeren van effectevaluaties te bevorderen. Een tweede weg om effectevaluaties te stimuleren zijn de eerder genoemde beleidsdoorlichtingen die in principe voor elk beleids-terrein zullen worden gemaakt. Een beleidsdoorlichting zal onder andere antwoord moeten geven op de vraag wat er bekend is over de effecten van de gehanteerde beleidsinstrumenten. Daartoe kunnen de in Nederland en in het buitenland uitgevoerde effectonderzoeken worden benut. Tegelijkertijd kunnen beleidsdoorlichtingen duidelijk maken waar de lacunes zitten. Beleidsdoorlichtingen bieden bovendien de mogelijkheid om beleidsmakers en onderzoekers bij elkaar te brengen om zo tot een planning te komen van effectonderzoek. Er zijn ongetwijfeld nog meer mogelijkheden om de totstandkoming van effectonderzoek te stimuleren. Zo lijkt het de moeite waard om in het buitenland te kijken naar voorbeelden van goed onderzoek, maar ook naar hoe de totstandkoming van effectonderzoek en de benutting daarvan worden gestimuleerd.

Tot slot

Uit het voorgaande komt een niet al te rooskleurig beeld naar voren van de staat van de beleidsevaluatie. Beleidsevaluaties kunnen een grotere rol spelen in het begrotingsproces als de kwaliteit en relevantie toeneemt. En dat kan alleen als de aandacht van beleidsmakers hiervoor toeneemt. Het zal een proces van lange adem zijn om deze vicieuze cirkel te doorbreken. Dat is noodzakelijk want er is geen alternatief voor evaluatieonderzoek voor een overheid die verantwoording wil afleggen over het gevoerde beleid en die wil leren van de ervaringen.

Literatuur

IOFEZ

Evaluatie van VBTB

Den Haag, Tweede Kamer der
Staten-Generaal, 29 949. nr. 1,
2004

IBO-werkgroep Arbeidsmarkt- beleid

Aan de slag

Den Haag, Ministerie van
Financiën, 2000

Fay, R.G.

*Enhancing the effectiveness of
active labour market policies;
evidence from programme
evaluations in oecd countries*

Oeso, Labour market and social
policy occasional papers no. 18,
1996

Leeuw, F.L.

*Productie en effectiviteit van
overheidsbeleid*

Utrecht, 1992

Leeuw, F.L.

*Trends and developments in
program evaluation in general
and in criminal justice programs
in particular*

European journal of criminal
policy and research, 26, 2005

Leuven, E., H. Oosterbeek

*Evaluating the effect of tax
deductions on training*

Journal of labor economics,
22e jrg., nr. 2, 2004



Onderstaande internetsites hebben betrekking op het thema van dit nummer van Justitiële verkenningen. De selectie werd gemaakt door de tijdschriftredactie en is uiteraard niet uitputtend.

www.campbellcollaboration.org

Misschien wel hét startpunt voor evaluatieonderzoekers en andere geïnteresseerden. De website biedt toegang tot de Campbell Library, die bestaat uit twee databases: C2-Spectr omvat bijna 12.000 ingangen tot beschrijvingen van experimentele evaluaties op het terrein van sociologie, psychologie, onderwijs en criminologie; in C2-Ripe zijn de door de zogeheten C2 Coordinating Groups goedgekeurde onderzoeken te vinden, naast protocols, besprekingen en samenvattingen. Uitgebreide informatie over de richtlijnen waaraan goed evaluatieonderzoek volgens de Campbell Collaboration dient te voldoen, een congresagenda en links.

www.wodc.nl

Vindplaats voor tal van evaluatiestudies op justitieel terrein.

www.cpb.nl/nl/onderwerpen/18.html

Op de site van het Centraal Planbureau zijn evaluatiestudies te vinden op allerlei maatschappelijke terreinen. Onder andere over het effect van overheidsinterventies in het onderwijs.

www.uu.nl/uupublish/onderzoek/onderzoekcentra/ercomer

Site van European Research Centre on Migration and Ethnic Relations (ERCOMER), ondergebracht bij de Faculteit Sociale Wetenschappen van de Universiteit Utrecht. Het centrum verricht veel vergelijkend onderzoek op het terrein van migratie en integratie.

www2.fmg.uva.nl/imes

Site van het Amsterdamse Institute of Migration and Ethnic Studies (IMES). Het instituut verricht onder meer onderzoek naar de theorie en praktijk van beleidsinterventies in migratieprocessen. Beschrijvingen van lopend onderzoek, een overzicht van publicaties en een uitgebreide pagina met links.

www.policyhub.gov.uk

Site van de Britse overheid met zeer veel praktische informatie voor zowel beleidsmakers als beleidsevaluatoren. Uitgebreide aandacht voor vragen over het meten van effecten van beleid, het vinden van bewijzen voor effecten, evaluatieuitgangspunten en -technieken enzovoort.

www.wa.gov/wsipp/crime

De site van het Washington State Institute for Public Policy biedt een uitgebreide pagina met samenvattingen van (evaluatie)studies op het terrein van criminal justice: het effect van straffen en maatregelen bij volwassenen, jeugdigen en seksueelgeweldplegers.

Summaries

Justitiële verkenningen (Judicial explorations) is published nine times a year by the Research and Documentation Centre of the Dutch Ministry of Justice in cooperation with Boom Juridische uitgevers. Each issue focuses on a central theme related to judicial policy. The section *Summaries* contains abstracts of the internationally most relevant articles of each issue. The central theme of this issue (nr. 8, 2005) is *Evaluation research*.

The importance of experimental evaluations in criminology

D.P. Farrington and B. Welsh

More convincingly than any other method, randomized experiments can establish the effect of an intervention on an outcome such as offending. This article reviews randomized experiments in criminology. The authors found 37 published experiments with at least 100 participants in 1957-81 and 85 experiments in 1982-2004. Noteworthy experiments in the two time periods are reviewed in more detail. While most of the earlier experiments produced non-significant findings, the later experiments showed that some interventions were effective (e.g. early developmental prevention, 'hot spots' policing, drug courts) and that others were ineffective (e.g. 'Scared Straight', boot camps and intensive supervision). Most experiments have been carried out in the United States. The authors recommend that other countries, including the Netherlands, should evaluate the effectiveness of criminological interventions using the 'gold standard' experimental design.

Evaluation of complex programmes; a theory driven approach

R. Pawson and C. Klein Haarhuis

In this article, the authors present an evaluation approach that is based on the premise that programmes are theories. Evaluating then becomes a process of testing programme theories. The 'theories-of-change' strategy helps to systematically put existing evidence together in order to reconstruct the 'inner workings' of a policy programme. This way, any failures in the programme can be identified and amended. The utility of this strategy is illustrated by the example of Megan's law, a sex offender notification programme in the United States.

Law enforcement and evaluation; a synthesis of 31 studies*C. Klein Haarhuis, M. van Ooyen-Houben, E. Kleemans and F. Leeuw*

In this article, the authors present the approach and findings of a research synthesis of 31 evaluations in the area of law enforcement conducted under the auspices of the Dutch Ministry of Justice. Two important findings are that (a) most of the enforcement interventions were not directed at natural persons or corporate actors as end-users of 'enforcement' activities, but on intermediary organizations in the government or in its vicinity; (b) the impact of the interventions was largely realized in terms of more production of output, like communication activities, number of criminal investigations etcetera rather than in terms of end effects.

Policies and causal effects; on randomized and natural experiments*H.D. Webbink*

Knowledge of the impact of policy interventions can yield high returns. However, estimating the impact of policy interventions is difficult because of selection of participants into treatment. Ignoring this selection may generate results that are seriously biased. Classical experiments with random assignment to a treatment and control group can solve this selection problem. In practice it can be difficult to realize this type of experiments. Many recent studies exploit so called natural experiments for identifying causal effects of policy interventions. Natural experiments are situations where the forces of nature or government policy have produced environments comparable to a randomised experiment. This paper reviews these new studies, the methods used and recent Dutch applications.

Evaluation research and integration policy*R. Bijl*

Evaluation research in the field of integration of ethnic minorities is gradually increasing. This article aims to stimulate the discussion on how evaluation research should be carried out, given the special characteristics of integration. The author discusses on which aspects of integration policies should aim at and how to assess these. Making proper policy evaluations will become easier because of the increasing availability of quantitative information on the position of ethnic minorities in Dutch society as well as increasing

possibilities to link existing registrations and other types of data of different organisations. Also the author stresses the importance of ex ante evaluations.

The role of evaluations in the budgetary process

A.L.M. de Jong and D.L. Kabel

Evaluation is vital to answers questions like: does the taxpayer get value for money? How can we learn from experiences and make policies more evidence-based in order to improve public sector performance? This article discusses the Dutch experience with evaluations from a broad perspective and presents the main elements of a new approach. In the Netherlands the number of policy evaluations has expanded enormously, but the quality lags behind. Improvement is needed on the following points: more effect evaluations; clearer policy objectives; a better guarantee of impartiality of the evaluations; and, improving the quality of the reporting. Furthermore, the general impression is that policy evaluations are potentially of great value, but that their current role in the policy process – with a few exceptions – is marginal. A strategy will be implemented to improve ex ante underpinning of policies, to refocus evaluations more on policy areas than on instruments and to stimulate impact analyses.

Journal

Het WODC op internet:
www.wodc.nl

Het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) van het ministerie van Justitie is via internet bereikbaar. Het WODC is een kenniscentrum op justitieel terrein dat een wetenschappelijke bijdrage levert aan de ontwikkeling en de evaluatie van justitiebeleid. Op de WODC-site is de volgende informatie te vinden:

1. Zoeken op onderwerp:

- trefwoord ABC (zoeken op standaard trefwoorden);
- onderwerpsgebieden.

2. Publicaties

- jaaroverzichten van alle rapporten (inclusief uitbesteed en *Justitiële verkenningen*);
- uitgebreide samenvattingen en volledige teksten (in PDF) beschikbaar bij vrijwel alle publicaties;
- aanpak criminaliteit (elektronische publicatie van het WODC en het Openbaar Ministerie over interventie- en preventiestrategieën inzake diverse criminaliteitsproblemen).

3. Onderzoek

- overzichten van lopend en afgesloten onderzoek (inclusief uitbesteed onderzoek);

- onderzoeksprogramma;
- monitoren en prognoses;
- procedure uitbesteed onderzoek.

4. Cijfers en prognose

- criminaliteit en rechtshandhaving (teksten en tabellen).

5. Informatiedesk

- bibliotheek;
- bronnengids 'Het criminaliteitsvraagstuk'.
- congressenlijst;
- opmerkelijk op tv;
- veelgestelde vragen;
- ik heb een vraag (vraagformulier informatiedesk).

6. Over het WODC (onder andere nieuwsberichten, organisatiestructuur, medewerkerslijst, vacatures, jaarbericht, commissies en werkgroepen).

Voor vragen over producten en diensten van het WODC kunt u terecht bij de WODC-informatiedesk (zie: rubriek 'Informatiedesk/Ik heb een vraag' op www.wodc.nl).

Hans van Netburg (redacteur WODC-site)

tel.: 070-3 70 69 19

fax: 070-3 70 79 48

e-mail: c.j.van.netburg@minjus.nl

WODC-thesaurus 'Het criminaliteitsvraagstuk' (3e herziene uitgave, 2003)

Het WODC heeft de derde herziene versie van de WODC-thesaurus beschikbaar gemaakt op de website van het WODC. Deze gestructureerde standaard trefwoordenlijst inzake het criminaliteitsvraagstuk bestrijkt de volgende terreinen: criminologie, criminaliteitspreventie, delicten/criminaliteitsvormen, justitiële organisatie, politie, slachtoffers, strafrechtspleging, strafstelsel, gevangeniswezen, reclassering en vreemdelingen. Als extra hulpmiddel is ook een nieuwe uitgave van de geografische thesaurus beschikbaar. De beide thesauri zijn te vinden op de WODC-website (www.wodc.nl) via de knop 'publicaties' en de subknop 'Cahiers', kies vervolgens het jaar 2003. Tevens is de gedrukte versie in een beperkte oplage gratis beschikbaar op aanvraag per e-mail: a.eind@minjus.nl.

Themanummers Justitiële verkenningen

Regelmatig worden, bijvoorbeeld ten behoeve van het onderwijs, grote aantallen exemplaren van *Justitiële verkenningen* nabesteld. Voor de bepaling van de oplage is het wenselijk vroegtijdig op de hoogte te zijn van de mogelijke vraag. Om die reden geven wij, onder voorbehoud, de thema's

die de redactie op dit moment in voorbereiding heeft. Over de inhoud van de nummers kan met de redactie contact worden opgenomen. Ook personen die een bijdrage willen leveren aan een van de aangekondigde nummers wordt verzocht contact op te nemen met de redactie.

- JV1, 2006** februari:
Coffeeshops en cannabis; de toekomst van het Nederlandse softdrugsbeleid
- JV2, 2006** maart/april:
Witwassen
- JV3, 2006** mei:
Demografische ontwikkelingen en justitie

Congressen

- 11 januari 2006
Nationale kennisdag geestelijke gezondheid en verslaving
- 12 januari 2006
Landelijke studiedag mensenhandel & 'moderne slavernij'
- 19 januari 2006
Landelijke Studiedag: Ketenvorming en Bemoeizorg in de OGGZ
- 19 januari 2006
Getuige-deskundigen en deskundigen over getuigen in strafzaken

26 januari 2006
 Technische opsporing en foren-
 sische bewijsvoering II
 26-27 januari 2006
 Uit en thuis met START;
 Risicotaxatie en behandeling in
 de (forensische) psychiatrie
 27 januari 2006
 Jeugdzorg in onderzoek;
 algemeen werkzame factoren
 1 maart 2006
 CIROC-seminar: Georgani-
 seerde misdaad en de Derde
 Wereld.
 Nadere informatie in de
 volgende JV
 11 mei
 WODC-congres 'Demografische
 ontwikkelingen en justitie'
 Nadere informatie in een
 volgende JV
 17 mei 2006
 CIROC seminar: 'Internationale
 smokkelparadijzen'
 Nadere informatie in een
 volgende JV

Hieronder volgen beschrijvingen
 van de congressen.

**Nationale kennisdag geeste-
 lijke gezondheid en verslaving**
 Het Trimbos instituut organi-
 seert ter gelegenheid van haar
 tienjarige bestaan een kennis-
 dag die in het teken staat van
 de belangrijke plaats die kennis
 inneemt bij het verbeteren
 van de geestelijke gezondheid.
 Diverse sprekers uit Nederland,

Australië en de Verenigde Staten
 zullen een bijdrage leveren op
 het gebied van de verbetering
 van preventie, behandeling en
 zorg in de geestelijke gezond-
 heid en verslaving. 's Middags
 zijn er parallelsessies – onder
 andere over 'verslaving' en
 'maatschappelijk onaangepast
 gedrag'. De dag wordt afgesloten
 met een discussie. Sprekers in
 de ochtend zijn onder anderen
 prof. dr. J. Walburg, min. drs. J.F.
 Hoogervorst, prof. dr. A. Beek-
 man, prof. G. Andrews en
 prof. J. Campbell.

Datum: 11 januari 2006 van
 09.30-17.00 uur

Plaats: Muziekgebouw aan het
 IJ, Amsterdam

Informatie: www.trimbos.nl/
 kennisdag.

**Landelijke studiedag mensen-
 handel en 'moderne slavernij'**
 Sinds 1 januari 2005 zijn sek-
 suele uitbuiting, uitbuiting
 in andere arbeidssituaties en
 orgaanverwijdering opgenomen
 in één wetsartikel Mensenhan-
 del. Hierdoor hebben mensen
 in het veld te maken gekregen
 met een nieuwe doelgroep en
 nieuwe ketenpartners.

Tijdens deze studiedag kan
 iedereen die in zijn/haar werk te
 maken kan krijgen met (slacht-
 offers van) mensenhandel en
 uitbuiting zich informeren over
 de toepasbaarheid van artikel

273a. Daarbij komen preventie, signalering, opsporing van daders en opvang van slachtoffers aan bod. Aan de orde komen vragen als: Wat is het verschil tussen het oude Artikel 250a en het nieuwe Artikel 273a? Bieden deze artikelen dezelfde mogelijkheden qua opsporing en vervolging? Hoe kunt u het wetsartikel bijvoorbeeld toepassen in uw werk als opsporingsambtenaar? Wie zijn slachtoffers van mensenhandel en uitbuiting? Op welke signalen van slachtoffers (man, vrouw, jong, oud, allochtoon of autochtoon) moet u als maatschappelijk werker, leraar, medewerker van politie of arbeidsinspectie alert zijn? Bent u zich als maatschappelijk werker of hulpverlener bewust van de mogelijkheden rond het doen van (anonieme) aangifte? Bent u als politiemedewerker op de hoogte van het feit dat ieder buitenlands slachtoffer recht heeft op de B9-regeling en dat u die optie moet aanbieden?

Datum: 12 januari 2006

Plaats: Muis Sacrum te Arnhem.

Informatie: Irene van Baalen, tel.: 071-5148203, irene@leidscongresbureau.nl, www.leidscongresbureau.nl.

Ketenvorming en Bemoeizorg in de OGGZ

De Openbare Geestelijke Gezondheidszorg (OGGZ) richt zich op mensen die de weg naar zorg niet kunnen of willen vinden, maar die wel hulp nodig hebben. Om te voorkomen dat zij op straat of thuis verkommeren en om overlast tegen te gaan, wil het kabinet dat er meer bemoeizorg ingezet wordt, dat er repressief wordt opgetreden en dat de mogelijkheden om dwang toe te passen verruimd worden. Tijdens deze studiedag wordt ingegaan op het belang van een goede ketensamenwerking en de rollen van een aantal partners: de gemeente, de politie en de woningbouwcorporatie. Hierbij gaat specifiek aandacht uit naar de methodiek voor de vorming van een sluitende ketensamenwerking. Tot slot zal tijdens de studiedag aandacht worden geschonken aan één specifieke doelgroep van de OGGZ, die extra zorgwekkend lijkt te worden: zwerfjongeren ouder dan 18 jaar. De studiedag is bedoeld voor alle professionals en hun management in de OGGZ, maatschappelijke opvang, verslavingszorg, reclassering, justitie en politie, welzijnsorganisaties, GGD's, gemeenten, begeleid wonen en woningbouwcorporaties.

Datum: 19 januari 2006
Plaats: Jaarbeurs Utrecht
Informatie: www.leidscongresbureau.nl.

Getuige-deskundigen en deskundigen over getuigen in strafzaken

Recent hebben de ontwikkelingen in de Schiedamse Parkmoord-zaak de publieke belangstelling gewekt voor onder meer het probleem van het redigeren en regisseren van de rapportage over deskundigenonderzoek in het strafproces. Daarbij doen onder meer de taakstelling en de taakopvattingen van deskundigen en juridische procesdeelnemers terzake. Ook is duidelijk geworden dat er nogal wat schort aan de communicatie tussen (forensisch) deskundigen en juristen, inclusief de advocatuur, de staande en de zittende magistratuur. Op deze studiedag komen de rol van getuige-deskundigen en getuigen in de rechtspleging ter sprake. Daarbij ligt het accent op de uitwisseling van ervaringen en inzichten, niet alleen binnen het eigen vakgebied of beroep, maar juist ook over de grenzen daarvan heen. Aanwezig zijn, onder anderen: mr. A.C. Möhring (Directeur Nederlands Forensisch Instituut, Den Haag); prof. dr. R.A.R. Bullens (Algemeen directeur FORA,

Bijzonder hoogleraar aan de Vrije Universiteit Amsterdam); prof. dr. mr. J.F. Nijboer (hoogleraar-directeur van het Seminarium voor Bewijsrecht, Leiden; Raadsheer in het Gerechtshof Amsterdam; Voorzitter Stichting Deskundigen en Rechtspleging); mw. mr. M.J. Dubelaar (onderzoeker, Straf- en Procesrecht Universiteit Leiden); prof. dr. W.A. Wagenaar (hoogleraar Psychologische Functieer en Rechtspsychologie, Universiteit Leiden) en anderen.

Datum: 19 januari 2006
Plaats: Congrescentrum De Eenhoorn, Amersfoort
Inlichtingen: Kerckebosch, tel.: 030-6984222.

Technische opsporing en forensische bewijsvoering II

Naast een 'update' met betrekking tot de technische en wetenschappelijke ontwikkelingen in de diverse forensische vakgebieden komen de rol en het belang van de technische opsporing en de daarmee verbonden forensische bewijsvoering aan de orde op deze studiedag. Er is nadrukkelijk aandacht voor de onderlinge relaties tussen technisch opsporingsonderzoek en forensische bewijsvoering. De studiedag is bedoeld voor allen die in de praktijk van het strafrecht werkzaam zijn, zoals (strafrecht)advocaten, (techni-

sche) onderzoekers, (forensische) deskundigen, rechters, officieren van justitie en advocaten-generaal. Daarenboven is de dag van belang voor beleidsambtenaren en wetenschappers (forensisch deskundigen) die met (onderdelen) van de strafrechtsketen te maken hebben. Aanwezig zijn, onder anderen: mr. A.J. Blok (Oud lid van het College van Procureurs-Generaal; Oud decaan Faculteit Rechtsgeleerdheid Universiteit Leiden); prof. dr. J.F. Nijboer (Directeur Seminarium voor Bewijsrecht; Hoogleraar Universiteit Leiden; Senior onderzoeker Nederlands Studiecentrum Criminaliteit en Rechtshandhaving; Raadsheer Gerechtshof Amsterdam; Voorzitter Stichting Deskundigen en Rechtspleging); prof. dr. P. de Knijff (Hoofd Forensisch Laboratorium voor DNA-onderzoek, LUMC); prof. dr. A. P.A. Broeders (hoogleraar criminalistiek, Universiteit Leiden; Chief Scientist Nederlands Forensisch Instituut, Den Haag).

Datum: 26 januari 2006

Plaats: Hotel-Congrescentrum Mitland, Utrecht

Informatie: Kerckebosch, tel.: 030-6984222.

Uit en thuis met START; risicotaxatie en behandeling in de (forensische) psychiatrie

Het taxeren van delictrisico's is met het oog op de maatschappelijke veiligheid van buitengewoon groot belang. Risicotaxatie in de klinische en ambulante praktijk van de forensisch psychiatrische zorg omvat meer. START (Short-Term Assessment of Risk and Treatability) is een methode die dit inzichtelijk maakt. Met de meeste gebruikte risicotaxatie-instrumenten wordt vooral gekeken naar risico's op de middellange en lange termijn. START meet ook de kortetermijnrisico's. De inschatting van de risico's verloopt op een gestructureerde manier. Diverse topics worden beoordeeld op zowel hun risicogehalte als hun preventieve werking. Bovendien biedt START de mogelijkheid om de behandeling optimaal te plannen en te evalueren. Ten slotte verstaat START – in tegenstelling tot de overige risicotaxatie-instrumenten – onder 'gevaar' niet alleen delictgevaar, maar ook gevaar jegens zichzelf. Sprekers: Prof. Christopher D. Webster (Canada) en mevrouw Prof. Corine de Ruiter.

Datum: 26-27 januari 2006

Plaats: GGZ Drenthe, Adventskerk Assen, De Brink Assen

Informatie: J. Veenendaal,

Coördinator symposium commissie FPK, tel.: 0592-334642, www.ggzdrenthe.nl, e-mail: symposiumfpk@ggzdrenthe.nl.

Jeugdzorg in onderzoek; algemeen werkzame factoren

Op initiatief van het ministerie van Volksgezondheid, Welzijn en Sport vindt jaarlijks het congres Jeugdzorg in Onderzoek plaats over (wetenschappelijk) onderzoek in de jeugdzorg. Tijdens deze tweede editie staat onderzoek centraal naar de vraag: Wat werkt in de jeugdzorg? De focus ligt hierbij op algemeen werkzame factoren, ofwel de ingrediënten van de hulp die de kans op effect bij welke doelgroep of hulpvorm dan ook vergroten. Het doel van Jeugdzorg in Onderzoek is het stimuleren van uitwisseling over wetenschappelijke kennis over de jeugdzorg; het verzamelen van good practices hoe onderzoek te doen naar wat werkt in de weerbarstige praktijk en, tot slot, het afstemmen van verschillende onderzoeksprogramma's en initiatieven. De dag geeft een overzicht van het onderzoek naar algemeen werkzame factoren in de jeugdzorg. Ook komt de toepassing en de implementatie van algemeen werkzame factoren in de praktijk aan de orde en hoe hiermee tot kwaliteitsverbetering van de

hulpverlening te komen. Iedere deelnemer kan een bijdrage voor Jeugdzorg in Onderzoek aanmelden. Een bijdrage kan gaan over: het toepassen van onderzoeksresultaten, over de resultaten van een onderzoek naar algemeen werkzame factoren in de jeugdzorg; het implementeren van kennis over algemeen werkzame factoren in de dagelijkse praktijk ter vergroting van het effect van de interventie.

Datum: 27 januari 2006 van 09.00-17.00 uur

Plaats: De Meervaart, Amsterdam

Informatie en aanmelding:
www.jeugdinterventies.nl,
Wieneke Bruinsma,
tel. 030-2306353,
w.bruinsma@nizw.nl.

De WODC-rapporten

Om zo veel mogelijk belanghebbenden te informeren over de onderzoeksresultaten van het WODC zijn alle WODC-rapporten en Cahiers in digitale vorm kosteloos beschikbaar op de website van het WODC (www.wodc.nl). Een complete lijst van WODC-rapporten en -cahiers en uitgebreide samenvattingen van alle sinds 1997 verschenen rapporten zijn daar ook te vinden. Oudere rapporten (vanaf 1999) zullen met terugwerkende kracht op de website beschikbaar komen.

In gedrukte vorm wordt een beperkte oplage van de rapporten in de reeks Onderzoek & Beleid en de Cahiers gratis verspreid. Er is een vaste verzendlijst van ondermeer universitaire en gerechtsbibliotheken, alsmede politieke partijen en media. Functionarissen en instellingen binnen het ministerie van Justitie kunnen op aanvraag afzonderlijke rapporten zonder kosten toegestuurd krijgen. Overige belangstellenden kunnen de rapporten in de reeks Onderzoek en Beleid (O&B) tegen betaling bestellen bij Boom Distributiecentrum te Meppel, tel.: 0522-23 75 55, fax: 0522-25 38 64, e-mail: bdc@bdc.boom.nl.

Hieronder volgen de titelbeschrijvingen van de O&B-rapporten sinds 2003 en van de in 2005 verschenen Cahiers.

Onderzoek en beleid (O&B)

Rovers, G.B., E. de Vries Robbé <i>Interne criminaliteit in de logistieke sector</i> 2005, O&B 235	Broeksteeg, J.L.W., E.M.J. Hardy, S. Klosse, M.G.W.M. Peeters en L.F.M. Verhey <i>Zicht op wetgevingskwaliteit; onderzoek naar de adviezen van de Raad van State inzake wetgevingskwaliteit</i> 2005, O&B 233
Gritter, E., G. Knigge en N.J.M. Kwakman <i>De WED op de helling; een onderzoek naar de wenselijkheid de Wet op de Economische Delicten te herzien</i> 2005, O&B 234	Erp, J.G. van, M.D. van Ewijk <i>Werklast bestuurlijke boete; determinanten van de werkbelasting in de bestuursrechtspleging</i> 2005, O&B 232

- Lünnemann, K.D.,
M.Y. Bruinsma
Geweld binnen en buiten; aard, omvang en daders van huiselijk en publiek geweld in Nederland
2005, O&B 231
- Wartna, B.S.J., S. el Harbachi,
L.M. van der Knaap
Buiten behandeling; een cijfermatig overzicht van de strafrechtelijke recidive van ex-terbeschikkinggestelden
2005, O&B 230
- Wartna, B.S.J., S. el Harbachi,
A.M. van der Laan
Jong vast; een cijfermatig overzicht van de strafrechtelijke recidive van ex-pupillen van justitiële jeugdinstellingen
2005, O&B 229
- Wartna, B.S.J., N. Tollenaar,
A.A.M. Essers
Door na de gevangenis; een cijfermatig overzicht van de strafrechtelijke recidive onder ex-gedetineerden
2005, O&B 228
- Wartna, B.S.J., N. Tollenaar,
M. Blom
Recidive 1997; een cijfermatig overzicht van de strafrechtelijke recidive van volwassen en jeugdige daders
2005, O&B 227
- Kogel, C.H. de, C. Verwers,
V.E. den Hartogh
'Blijvend delictgevaarlijk'; empirische schattingen en conceptuele verheldering
2005, O&B 226
- Jong, E. de
Alle regels tellen mee; de ontwikkeling van het aantal wetten, Amvb's en ministeriële regelingen in Nederland
2004, O&B225
- Moors, J.A., M.Y.W. von Bergh,
S. Bogaerts, J.W.M.J. van Poppel, A.M. van Kalmthout
Kiezen voor delen? Evaluatie van de eerste fase van de invoering van meerpersoonscelgebruik
2004, O&B 223
- Beijer, A., R.J. Bokhorst,
M. Boone, C.H. Brants,
J.M.W. Lindeman
De Wet bijzondere opsporingsbevoegdheden; eindevaluatie
2004, O&B 222
- Leertouwer, E.C. en
G.L.A.M. Huijbregts
Sanctiecapaciteit 2008
2004, O&B 221
- Leuw, Ed., R.V. Bijl,
A. Daalder
Pedoseksuele delinquentie; een onderzoek naar prevalentie, toedracht en strafrechtelijke interventies
2004, O&B 220
- Velthoven, B.C.J. van,
M.J. ter Voert
Geschildbeslechtingsdelta 2003; over verloop en afloop van (potentieel) juridische problemen van burgers
2004, O&B 219

- Faber, W., A.A.A. van Nunen
Uit onverdachte bron; evaluatie van de keten ongebruikelijke transacties
2004, O&B 218
- Moolenaar, D.E.G., P.P.J. Groen, A.G. Mein, B.S.J. Wartna, M. Blom
Wegenverkeerswet 1994; evaluatie van de wijzigingen 1998
2004, O&B 217
- Wartna, B.S.J., N. Tollenaar
Bekenden van Justitie; een verkennend onderzoek naar de 'veelplegers' in de populatie van vervolgte daders
2004, O&B 216
- Bunt, H.G. van de, C.R.A. van de Schoot (eds.)
Prevention of organised crime; a situational approach
2004, O&B 215
- Dijksterhuis, B.M., M.J.G.
Jacobs, W.M. de Jongste
De competentiegrens van enkelvoudige kamers in strafzaken
2003, O&B 214
- Smit, P.R., F.P. van Tulder, R.F. Meijer, P.P.J. Groen
Het ophelderingspercentage nader beschouwd
2003, O&B 213
- The European Sourcebook of Crime and Criminal Justice Statistics – 2003 (second edition)
2003, O&B 212
- Heide, W. van der, A.Th.J. Eggen (eindred.)
Criminaliteit en rechtshandhaving 2001; ontwikkelingen en samenhangen
2003, O&B 211
- Combrink-Kuiter, C.J.M., E. Niemeijer, J.M.M. Kuppens, M.J. ter Voert
Ruimte voor Mediation
2003, O&B 210
- Eshuis, R.J.J.
Claims bij de rechtbank
2003, O&B 209
- Kogel, C.H. de, C. Verwers
De longstay afdeling van Veldzicht; een evaluatie
2003, O&B 208
- Molenaar, D.E.G., G.L.A.M. Huijbregts
Sanctiecapaciteit 2007; een beleidsneutrale prognose
2003, O&B 207
- Kromhout, M., M. van San
Schimmige werelden; nieuwe etnische groepen en jeugdcriminaliteit
2003, O&B 206
- Tak, P.J.J.
The Dutch Criminal Justice System – Organization and operation – Second revised edition
2003, O&B 205

WODC-Cahiers 2005

Blom, M., A.M. van der Laan, G.L.A.M. Huijbregts
Monitor Jeugd terecht 2005
Den Haag, WODC, 2005,
Cahiers-17

Bijl, R.V., A. Zorlu, A.S. van Rijn, R.P.W. Jennissen, M. Blom
Integratiekaart 2005; de maatschappelijke integratie van migranten in de tijd gevolgd: trend- en cohortanalyses
Den Haag, WODC, 2005,
Cahiers-16

Laclé, Z.D., B. Krop, N.J.H. Huls
Balansverschuiving? Notarissen over ontwikkelingen in de notariële beroepsethiek na vijf jaar Wet op het notarisambt 1999
Den Haag, WODC, 2005,
Cahiers-15

Leertouwer, E.C., F.P. van Tulder, B.J. Diephuis, M. Folke- ringa, R.J.J. Eshuis m.m.v. M. van Gammeren, A. van Son, Y. Visser, N. Dijkhoff
Prognosemodellen justitiële ketens; civiel en bestuur – modelversie 1.0
Den Haag, WODC, 2005,
Cahiers-13

Neve, R.J.M., M.M.J. van Ooyen-Houben, J. Snippe, B. Bieleman, A. Kruize, R.V. Bijl
Samenspannen tegen XTC; tussentijdse evaluatie van de XTC-nota
Den Haag, WODC, 2005,
Cahiers-12

Verwers, C., S. Bogaerts
Strafrechtelijke aanpak van jongvolwassenen (18-24 jaar): een verkenning

Den Haag, WODC, 2005,
Cahiers-11

Baas, N.J.
Wegen naar het rechte pad; strafrechtelijke interventies voor delinquente jongeren, inclusief verplichte nazorg, en naar het effect van interventies die zich voor een strafrechtelijk kader zouden kunnen lenen, alsmede naar interventiecondities die dat effect beïnvloeden

Den Haag, WODC, 2005,
Cahiers-10

Laan, A.M. van der, L.M. van der Knaap, B.S.J. Wartna
Recidivemeting onder jeugd-reclasseringscliënten
Den Haag, WODC, 2005,
Cahiers-9 (in druk; wél op website)

Laan, A.M. van der, L.M. van der Knaap, B.S.J. Wartna
Recidivemeting onder de deelnemers aan de Sociale Vaardigheidstraining voor groepen van het Leger des Heils
Den Haag, WODC, 2005,
Cahiers-8 (in druk; wél op website)

Knaap, L.M. van der, A.M. van der Laan, B.S.J. Wartna
Recidivemeting onder deelnemers van nieuw Positief Initiatief
Den Haag, WODC, 2005,
Cahiers-7 (in druk; wél op website)

**Knaap, L.M. van der, A.M. van
der Laan, B.J.S. Wartna**

*Recidivemeting onder deel-
nemers Binnenste Buiten*
Den Haag, WODC, 2005,
Cahiers-6 (in druk; wél op
website)

**Klein Haarhuis, C.M.,
M.M.J. van Ooyen-Houben,
E.R. Kleemans, F.L. Leeuw**
*Kennis voor beleid; een synthese
van 58 (evaluatie)onderzoeken
op het gebied van rechtshand-
having*

Den Haag, WODC, 2005,
Cahiers-5

Klein Haarhuis, C.M.
*Naar een conceptueel kader voor
de spreiding van rechtspraak;
een verkenning aan de hand van
inzichten uit onderwijs en zorg*
Den Haag, WODC, 2005,
Cahiers-4

Mascini, P., M. Verhoeven
*Literatuurstudie naar de facilite-
ring van de gewelddadige jihad*
Den Haag, WODC, 2005,
Cahiers-3

**Blom, M. (red.), J. Oudhof
(red.), R.V. Bijl (red.),
B.F.M. Bakker (red.)**
*Verdacht van criminaliteit;
allochtonen en autochtonen
nader bekeken*
Den Haag, WODC, CBS, 2005,
Cahiers-2

Register Justitiële verkenningen

Jaargang 2005

- Baggerman, A. en R. Dekker**
Verlichting, revolutie en kindbeeld in Nederland; de periode rond 1800 als keerpunt nr. 5, p. 9
- Beek, D.J. van**
De beveiliging van de maatschappij tegen seksuele delinquenten; een simpele of complexe vraag nr. 1, p. 9
- Berghuis, A.C.**
De bestuurlijke boete en het verlangen naar handhaving nr. 6, p. 30
- Bijleveld, C. en J. Hendriks*
Jeugdige zedendelinquenten; jong geleerd, oud gedaan? nr. 1, p. 95
- Bijl. R.**
Evaluatieonderzoek en het integratiebeleid nr. 9, p. 85
- Bogaerts, S. en S. Vanheule**
Angst voor de grote boze wolf; maatschappelijke reacties op seksueel misbruik nr. 1, p. 48
- Bogaerts, S., A.L. Daalder en A.A.M. Essers**
Interpersoonlijke vaardigheden bij pedoseksuelen; de kortzichtigheid van attitudeonderzoek nr. 1, p. 58
- Bovenkerk, F. en M. Husken**
De culturele erfenis van Don Corleone nr. 4, p. 11
- Brinkgreve, C.**
De terugkeer van het angstaanjagende kind nr. 5, p. 24
- Brouwers, M. en P.R. Smit**
Seksuele delinquentie; de prevalentie door de jaren heen nr. 1, p. 37
- Coolsaet, R.**
Het islamistische terrorisme; percepties wieden en kweekvijvers dreggen nr. 2, p. 9
- Ewijk, M.D. van en E. Niemeijer**
Bestuurlijke boetes; extra werk voor de rechtspraak? nr. 6, p. 73
- Farrington, D.P. en B.C. Welsh**
Het belang van experimentele evaluaties in de criminologie nr. 9, p. 11
- Franke, H.**
Alleen de moerassen zijn vruchtbaar; over de verbanden tussen criminologie en literatuur nr. 4, p. 82

Gemert, F. van <i>Gangsta rap; de controversie voorbij</i>	nr. 4, p. 68
Graaf, G. de <i>De corrupte functionaris, zijn omkoper en hun relatie</i>	nr. 7, p. 49
Grotenhuis, A.L.J. <i>Belastingadviseurs in de knel tussen twee loyaliteiten</i>	nr. 3, p. 99
Hartmann, A.R. <i>Strafbeschikking en bestuurlijke boete; wildgroei in de handhaving?</i>	nr. 6, p. 84
Hoogenboom, B. <i>The Sopranos; narrativiteit als ontregeling</i>	nr. 4, p. 25
Huberts, L.W.J.C. en K. Lasthuizen <i>Tussen corruptie en corruptieparadox; over de omvang van corruptie in Nederland</i>	nr. 7, p. 9
Jong, A.H.M. de en D.L. Kabel <i>De rol van evaluaties in het begrotingsproces</i>	nr. 9, p. 100
Jong, M. de <i>De strijd van de Wereldbank tegen fraude en corruptie</i>	nr. 7, p. 90
Jongbloed, A.W. <i>De gerechtsdeurwaarder; zakenman of functionaris?</i>	nr. 3, p. 64
Klein Haarhuis, C., M. van Ooyen-Houben, E. Kleemans en F. Leeuw <i>Rechtshandhaving geëvalueerd; een synthese van 31 onderzoeken</i>	nr. 9, p. 54
Koops, W. en B. Orobio de Castro <i>Ontwikkelingspsychologen op zoek naar oorzaken van jeugddelinquentie; nieuwe trajecten naar kennis</i>	nr. 5, p. 49
Laclé, Z.D. <i>Notariaat, ethiek en marktwerking</i>	nr. 3, p. 49
Lankhorst, F. en J.M. Nelen <i>Integriteitsproblemen van advocaten en notarissen in relatie tot georganiseerde criminaliteit</i>	nr. 3, p. 32
Leeuwen, M. van <i>Ronselen in Europa voor de Heilige Oorlog</i>	nr. 2, p. 84
Maat, J.H. <i>Buitenlandse corruptie en de aanpak door de Rijksrecherche</i>	nr. 7, p. 64
Malsch, M. <i>Geregisseerde openbaarheid; het O.M. en de zichtbaarheid van transacties</i>	nr. 6, p. 113

Mascini, P. en M. Verhoeven <i>Kan de gewelddadige jihad zonder sympathisanten?</i>	nr. 2, p. 63
Mein, A.G. <i>Wet bestuurlijke boete; recht doen aan gemeentelijke autonomie</i>	nr. 6, p. 35
Michiels, F.C.M.A. <i>De bestuurlijke boete als remedie tegen handhavingstekorten</i>	nr. 6, p. 60
Nelen, J.M. <i>De strafrechtelijke afdoening van ambtelijke corruptie</i>	nr. 7, p. 37
Niemeijer, E. en M. ter Voert <i>Vertrouwen onder druk; vrije juridische beroepen tussen professie en commercie</i>	nr. 3, p. 9
Nijboer, J.A. <i>Beeldvorming over criminele kinderen en het justitiebeleid</i>	nr. 5, p. 35
Nijboer, J.F. <i>Plea bargain en O.M.-afdoening als functionele equivalenten?</i>	nr. 6, p. 128
Nijnatten, C. van <i>Het kind als communicatieve partner</i>	nr. 5, p. 85
Oei, T.I. <i>De terbeschikkingstelling: van een bijzondere tot een gewone maatregel?</i>	nr. 1, p. 23
Pal, J. van der en H.M.G. Slangen <i>Reikwijdte bestuurlijke boete is te beperkt</i>	nr. 6, p. 23
Pawson, R. en C. Klein Haarhuis <i>Evaluatie van complete programma's: een theorie-gestuurde aanpak</i>	nr. 9, p. 42
Roy, O. <i>Euro-islam; de jihad van binnenuit?</i>	nr. 2, p. 28
Schilder, A. en W.H.J.M. Nuijts <i>De accountant; onafhankelijkheid in een meervoudig spanningsveld</i>	nr. 3, p. 77
Schuilenburg, M.B. en A. de Jong <i>De militarisering van de openbare ruimte; over de invloed van videogames op onze werkelijkheid</i>	nr. 4, p. 47
Sikkema, E. <i>Ambtelijke corruptie in de strafwetgeving</i>	nr. 7, p. 26

Sikkema, P. <i>Het kind als consument</i>	nr. 5, p. 77
Sissener, T.K. <i>De vele gezichten van corruptie; een antropologische kijk</i>	nr. 7, p. 76
Stokkom, B.A.M. van <i>Zero tolerance in de praktijk; handhaving van de 'kleine norm' door politie of boa's?</i>	nr. 6, p. 44
Terpstra, J. en T. Havinga <i>Gemeenten, boetes en kleine ergernissen</i>	nr. 6, p. 10
Veen, H.T. van der <i>Drugshandel en politiek geweld</i>	nr. 2, p. 95
Veen, V.C. en C. de Ruiter <i>De effectiviteit van behandelingen bij seksuele delinquenten; een overzicht van de internationale literatuur</i>	nr. 1, p. 75
Webbink, H.D. <i>Causale effecten van beleid; over gecontroleerde en natuurlijke experimenten</i>	nr. 9, p. 72
Weijers, I. <i>'Mondige kinderen'</i>	nr. 5, p. 66
Wijk, A.Ph., R. Loeber, H. Ferwerda, A. Smulders en R. Vermeiren <i>Jeugdige zedendelinquenten en geweldplegers; een vergelijking op grond van de Pittsburgh Youth Study</i>	nr. 1, p. 105
Wijk, R. de <i>Terrorismebestrijding; een strategie van 'hearts and minds'</i>	nr. 2, p. 47
Willems, J.C.M. <i>Het kindbeeld in het Verdrag inzake de Rechten van het Kind</i>	nr. 5, p. 94
Zijl, J. van <i>Het O.M. beschikt tot straf; een strafrechtelijk novum</i>	nr. 6, p. 97