

CPB Document

No 48

Januari 2004

**Lerend beleid: het versterken van beleid door
experimenteren en evalueren**

Maarten Cornet, Dinand Webbink

Centraal Planbureau
Van Stolkweg 14
Postbus 80510
2508 GM Den Haag

Telefoon (070) 338 33 80
Telefax (070) 338 33 50
Internet www.cpb.nl

ISBN 90-5833-156-3

Inhoud

Ten geleide	5
Samenvatting	7
1 Inleiding	9
2 Het wat en waarom van beleidsexperimenten	11
2.1 Waarom: het beleidsevaluatieprobleem	11
2.2 Wat: gecontroleerde en natuurlijke beleidsexperimenten	12
3 Gecontroleerde beleidsexperimenten	13
3.1 Het gecontroleerde experiment	13
3.2 Voordelen van gecontroleerde beleidsexperimenten	14
3.3 Nadelen van gecontroleerde experimenten	14
3.4 Amerikaanse ervaringen met gecontroleerde beleidsexperimenten	16
4 Natuurlijke beleidsexperimenten	19
5 Gecontroleerde en natuurlijke experimenten en niet-experimentele analyse vergeleken	21
6 Toepassing van beleidsexperimenten in Nederland	23
6.1 Gecontroleerde experimenten bij ‘slepende beleidskwesties’	23
6.2 Kansen voor evaluatie door handige vormgeving van beleid	24
6.3 Van pilots naar gecontroleerde experimenten	25
7 Conclusies	27
Referenties	29
Abstract	31

Ten geleide

De belangstelling voor ‘evidence based policy’ neemt de laatste jaren sterk toe. Gecontroleerde en natuurlijke experimenten met beleidsinstrumenten, met de daaraan verbonden evaluaties, kunnen een stevige empirische onderbouwing leveren voor bestaande en voor nieuwe beleidsinstrumenten. Dit rapport geeft een overzicht van nationale en internationale ervaringen met beleidsexperimenten en gaat in op mogelijke toepassingen in Nederland.

Het rapport is geschreven door Maarten Cornet en Dinand Webbink op verzoek van het Ministerie van Economische Zaken, ter voorbereiding van een bundel waarin een interdisciplinair perspectief op marktordening wordt geboden.

F.J.H. Don, directeur

Samenvatting

De effecten van beleid zijn vaak niet bekend. Dit geldt zowel voor al lang lopend beleid als voor nieuw beleid. De kosten van het uitvoeren en continueren van ineffectief beleid kunnen zeer hoog zijn. Experimenten kunnen overtuigend bewijs leveren over de effecten van beleid. De baten van experimenten kunnen daarom zeer hoog zijn. De mogelijkheden om kennis te vergaren met behulp van beleidsexperimenten worden in Nederland nog weinig benut.

Twee typen experimenten worden onderscheiden: gecontroleerde en natuurlijke experimenten. Bij een gecontroleerd experiment creëert een onderzoeker door toeval een experimentele groep en een controlegroep. De experimentele groep ondergaat een bepaalde interventie; de controlegroep ondergaat die interventie niet. Het verschil in resultaten tussen beide groepen kan alleen worden toegeschreven aan de interventie. Bij een natuurlijk experiment wordt de experimentele en de controlegroep gevormd door een toevallige situatie in de werkelijkheid, bijvoorbeeld door elementen uit regelgeving of populatieveranderingen.

Gecontroleerde experimenten stuiten in Nederland vaak op ethische bezwaren, voortkomend uit de ongelijke behandeling van personen. Daar staat tegenover dat alle toekomstige generaties kunnen profiteren van de kennis verworven in een gecontroleerd experiment. Het alternatief 'niet experimenteren' draagt het risico in zich dat ineffectief of zelfs schadelijk beleid wordt voortgezet. Gecontroleerde experimenten zijn al lange tijd bij de ontwikkeling van nieuwe geneesmiddelen gebruikelijk.

Gecontroleerde experimenten kunnen op veel terreinen worden toegepast. Ze kunnen vooral voordelen opleveren bij 'slepende kwesties': dit zijn beleidsissues waar al lange tijd discussie bestaat over de effecten van beleidsinstrumenten. Voorbeelden hiervan zijn: financiële prikkels ter bevordering van de deelname aan technische studies, prikkels voor de reïntegratie van werklozen, R&D-subsidies of regels rond marktwerking en innovatie.

Natuurlijke experimenten zijn een relatief goedkope bron van kennis over de effecten van beleid. Bijkomend voordeel is dat de resultaten beschikbaar komen op relatief korte termijn. Natuurlijke experimenten komen echter voort uit toevallige omstandigheden die zich niet op alle terreinen voordoen.

De vormgeving van beleid kan een eerste stap zijn voor de evaluatie van beleid. Tot dusver blijven verschillende mogelijkheden onbenut. In de eerste plaats zou meer gebruik gemaakt kunnen worden van het aselect toekennen van subsidies, bijvoorbeeld in situaties waarin de vraag naar subsidies groter is dan het budget. Door subsidies aselect toe te kennen wordt een experimentele en een controlegroep gevormd als basis voor een toekomstige evaluatie. Een andere mogelijkheid is het bewust inbouwen van discontinuïteiten in subsidieregelingen. In Nederland is bijvoorbeeld de afgelopen jaren extra geld voor personeel en computers toegekend aan scholen met minstens 70 % achterstandsleerlingen. Scholen net beneden deze grens kregen geen extra subsidie. Een dergelijke knip in de regeling maakt een evaluatie mogelijk. Door

scholen net rond de knip te vergelijken wordt een echt experiment nagebootst. Door vaker in regelingen zulke discontinuïteiten in te bouwen ontstaan mogelijkheden om de effecten van beleid vast te stellen.

In Nederland worden regelmatig 'pilots' uitgevoerd. 'Pilots' hebben meestal geen controlegroep en verschillen daarmee van beleidsexperimenten. Het ontbreken van een controlegroep maakt het ook lastig om de effecten vast te stellen. Door bij 'pilots' systematisch aandacht te besteden aan het zoeken naar een aselechte controlegroep kan meer overtuigende 'evidence' worden verkregen.

1 Inleiding

Wat zijn de effecten van beleid? Deze voor de hand liggende vraag is vaak moeilijk te beantwoorden. Dat geldt zowel voor lang lopend beleid als voor nieuw uit te voeren beleid. Zo wordt op verschillende beleidsterreinen al geruime tijd veel geld uitgegeven zonder dat bekend is wat de effecten zijn. Een voorbeeld daarvan is het beleid gericht op het bestrijden van achterstanden in het onderwijs. De Rekenkamer concludeerde op basis van een overzicht van 35 evaluatiestudies dat er geen relatie kan worden gelegd tussen de ontwikkeling in prestaties van leerlingen met onderwijsachterstanden en het gevoerde beleid.¹ Andere voorbeelden zijn: subsidies en fiscale regelingen voor innovatie, technologie, milieu, energie of export, computers in het onderwijs en scholing van werklozen. Ook zijn recent minder gunstige ervaringen opgedaan met grootschalige beleidswijzigingen, bijvoorbeeld de liberalisering van de taximarkt of de invoering van het studiehuis. Natuurlijk zijn deze wijzigingen niet ‘over één nacht ijs’ ingevoerd, maar een stevige empirische onderbouwing van dit beleid ontbrak.

Tegen deze achtergrond is het begrijpelijk dat de belangstelling voor ‘evidence based policy’ de laatste jaren sterk toeneemt. De invoering van het VBTB-proces kan als een illustratie daarvan worden gezien. Het recente rapport van de Rekenkamer over de uitvoering van beleid onderstreept dat er op het terrein van evaluatie van beleid nog een wereld te winnen is.² Daarnaast zijn de afgelopen jaren in Nederland diverse rapporten verschenen waarin wordt opgeroepen tot experimenteren met beleid.³ Zo pleit het CPB in een recente studie voor het uitvoeren van beleidsexperimenten om kennis over de effecten van kennisbeleid te verwerven.⁴

Kennis van de effecten van beleid versterkt het beleid. Beleid dat niet werkt of zelfs schadelijk is, kan worden gestopt. Beleid dat werkt kan breed worden ingevoerd. Omdat de maatschappelijke kosten van ineffectief beleid zeer hoog kunnen zijn, kunnen de maatschappelijke baten van kennis van de effecten van beleid zeer hoog zijn.

Beleidsexperimenten kunnen overtuigende kennis leveren over de effecten van beleid. De maatschappelijke baten van beleidsexperimenten zijn dus in potentie hoog.

In deze bijdrage gaan we in op de voor- en nadelen van experimenteren met beleid. Paragraaf 2 bespreekt het beleidsevaluatieprobleem: waarom experimenteren? Deze paragraaf definieert ook twee soorten experimenten: het gecontroleerde beleidsexperiment en het natuurlijke beleidsexperiment. Paragraaf 3 bespreekt de ervaringen met gecontroleerde beleidsexperimenten; paragraaf 4 die met natuurlijke beleidsexperimenten. Paragraaf 5 zet de voor- en nadelen van gecontroleerde beleidsexperimenten, natuurlijke beleidsexperimenten en niet-experimentele evaluatiemethoden tegen elkaar af. Paragraaf 6 beschrijft de kansen voor toepassing van experimenten in de Nederlandse context. Paragraaf 7 sluit af met conclusies.

¹ Algemene Rekenkamer (2001).

² Algemene Rekenkamer (2003).

³ Zie bijvoorbeeld WRR (2002)

⁴ CPB (2002).

2 Het wat en waarom van beleidsexperimenten

2.1 Waarom: het beleidsevaluatieprobleem

Het vaststellen van effecten van beleid is vaak lastig omdat veel factoren van invloed kunnen zijn. Vergelijkingen in de tijd, voor en na invoering van de beleidsmaatregel, zijn vaak onvoldoende omdat vele factoren kunnen veranderen. Maar ook vergelijking tussen groepen die wel en niet onder de maatregel vallen, geeft meestal niet direct een betrouwbare schatting van het beleidseffect. Onderzoekers zullen veelal slechts rekening kunnen houden met een beperkt aantal factoren. Het gevolg is dat factoren die niet worden waargenomen, de resultaten kunnen vertekenen. Een voorbeeld van dit veel voorkomende evaluatieprobleem is het vaststellen van de effecten van training voor werklozen. Om het effect van de training vast te stellen dient de groep deelnemers te worden vergeleken met een groep niet-deelnemers. Daarbij dient rekening gehouden te worden 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 motivatie of bepaalde vaardigheden, is dat moeilijk, 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.

In formele zin ziet dit evaluatieprobleem er als volgt uit:

$$Y_i = \alpha + \beta X_i + \delta T + \varepsilon_i$$

met Y is de uitkomstvariabele, bijvoorbeeld loon of duur van de werkloosheid, X zijn persoonlijke kenmerken van deelnemer i die door de onderzoeker worden waargenomen, T is het al of niet ondergaan van de interventie, bijvoorbeeld het deelnemen aan de training, ε zijn niet waargenomen factoren en α , β en δ te schatten parameters. De meest interessante parameter is δ , die geeft namelijk het effect weer van de interventie. Het belangrijkste probleem bij veel evaluaties is dat niet-waargenomen kenmerken kunnen samenhangen met de interventie: $(\text{cov}(\varepsilon_i, T) \neq 0)$. Bijvoorbeeld, deelnemers aan een training zijn sterker gemotiveerd dan niet-deelnemers. De onderzoeker neemt dit echter niet waar. In dat geval zal een schatting van δ behalve het effect van de training ook het effect van de motivatie oppakken.

2.2 Wat: gecontroleerde en natuurlijke beleidsexperimenten

In principe kan dit evaluatieprobleem worden opgelost door rekening te houden met meer factoren (X uitbreiden). 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 de econometrie zijn verschillende verfijnde methoden voorgesteld om het selectieprobleem bij evaluaties aan te pakken, denk bijvoorbeeld aan het zogenoemde Heckman-model⁵ en het propensity score matching.⁶ Het probleem met deze methoden is echter dat additionele aannamen nodig zijn. Het blijft moeilijk om hard te maken dat het gevonden effect daadwerkelijk het resultaat is van de interventie en niet het gevolg is van niet-waargenomen kenmerken.

In de recente evaluatieliteratuur neemt daarom de belangstelling toe voor experimentele benaderingen, in de vorm van gecontroleerde experimenten en vooral in de vorm van natuurlijke experimenten.

Definitie van het gecontroleerde beleidsexperiment

De onderzoeker creëert door toeval een experimentele en een controlegroep. De experimentele groep ondergaat een bepaalde beleidsinterventie ('treatment'). Evaluatie van de beleidsinterventie is het meten van het verschil in resultaten tussen beide groepen. Dit verschil kan alleen worden toegeschreven aan de interventie. Het verschil is het effect van de beleidsinterventie.

Definitie van het natuurlijke beleidsexperiment

Een toevallige situatie in de werkelijkheid creëert een experimentele en een controlegroep. Alleen de experimentele groep heeft een bepaalde interventie ondergaan. Evaluatie van de beleidsinterventie is het meten van het verschil in resultaten tussen beide groepen. Dit verschil kan alleen worden toegeschreven aan de interventie. Het verschil is het effect van de beleidsinterventie.

⁵ Heckman (1979).

⁶ Zie bijvoorbeeld Dehejia en Wahba (2002).

3 Gecontroleerde beleidsexperimenten

3.1 Het gecontroleerde experiment

In de natuurwetenschappen en in de medische wereld wordt het gecontroleerde experiment als de ideale methode beschouwd om kennis te vergaren. In een medisch experiment wordt bijvoorbeeld een groep patiënten door toeval toegewezen aan twee groepen:

- de experimentele groep die het medicijn krijgt;
- de controlegroep die geen medicijn maar een placebo krijgt.

Door deelnemers door toeval toe te wijzen aan een experimentele en een controle groep wordt de kans sterk verkleind dat andere factoren dan de beoogde interventie de resultaten beïnvloeden. Het verschil in gezondheid tussen beide groepen kan daarom met een grote mate van zekerheid worden toegeschreven aan het medicijn. Om de zekerheid nog verder te vergroten wordt vaak het zogenoemde ‘dubbelblind’ experiment toegepast. Deelnemers noch begeleiders weten in dat geval welke groep de experimentele of de controlegroep is.

Ook in de sociale wetenschappen, waaronder de economie, kunnen experimenten een belangrijk instrument zijn om kennis te vergaren. Zo kan een gecontroleerd experiment het hiervoor beschreven selectieprobleem bij de deelname aan training oplossen. Door deelnemers door toeval (bijvoorbeeld loting) toe te wijzen aan de trainingsgroep en aan de controlegroep is er geen reden om aan te nemen dat de groepen systematisch verschillen op niet-waargenomen factoren. Anders gezegd, de kans dat e en T samenhangen, is zeer gering. Daardoor kan het effect van T overtuigend worden vastgesteld.

Pilots zijn geen gecontroleerde experimenten

De random toewijzing van deelnemers aan de experimentele en aan de controlegroep is het fundament van het bewijs. Dit betekent dat er een cruciaal verschil bestaat tussen gecontroleerde experimenten en de zogenoemde ‘pilots’, die in het beleid veel voorkomen. Het bewijs verkregen uit pilots is minder overtuigend omdat niet duidelijk is in hoeverre de resultaten veroorzaakt worden door de interventie. Het is bijvoorbeeld mogelijk dat resultaten van ‘pilots’ niet representatief zijn omdat de deelnemers veel gemotiveerder zijn dan gemiddeld. Ook is het mogelijk dat deelnemers aan een ‘pilot’ spontaan veranderen doordat ze meer ervaring krijgen.

3.2 Voordelen van gecontroleerde beleidsexperimenten

Gecontroleerde experimenten leveren overtuigend bewijs

Het grote voordeel van gecontroleerde experimenten is dat zij overtuigend empirisch bewijs bieden over wat werkt en niet werkt. Uiteindelijk vinden wetenschappers en beleidsmakers bewijs verkregen langs experimentele weg overtuigender dan bewijs dat op andere wijze is verkregen. Bijkomend voordeel is dat dit bewijs vaak helder gecommuniceerd kan worden.

De baten van experimenteren kunnen zeer hoog zijn

Het voeren van beleid waarvan de uitkomsten niet duidelijk zijn, kan hoge kosten opleveren. Dit geldt des te meer als dit beleid jarenlang wordt gecontinueerd. Door te experimenteren kan kennis worden verkregen over wat werkt en niet werkt. Het is duidelijk dat de baten hiervan zeer hoog kunnen zijn.

3.3 Nadelen van gecontroleerde experimenten

Ethiek: experimenteren met mensen

Een veel genoemd nadeel van experimenteren is het ethische aspect. Een gecontroleerd experiment geeft per definitie ongelijke behandeling van mensen. Een groep mensen wordt immers mogelijk een succesvolle interventie onthouden. Deze ongelijkheid bij het experimenteren kan de politieke haalbaarheid onder druk zetten. Daar kan tegen ingebracht worden dat in de medische wetenschap experimenten met random toewijzing geheel geaccepteerd zijn. Waarom zou het dan in de economische wetenschap ethisch niet verantwoord zijn? Ook is soms op voorhand niet zeker of een bepaalde interventie de beoogde voordelen daadwerkelijk genereert. Als het effect met een grote mate van zekerheid is vastgesteld, kunnen komende generaties daarvan profiteren, hetzij door een bepaalde interventie grootschalig toe te passen hetzij door het niet continueren van bestaand beleid. Het alternatief 'niet experimenteren' draagt het risico in zich dat ineffectief of schadelijk beleid wordt voortgezet. Het bezwaar van ongelijke behandeling kan mogelijk ook worden verminderd door de experimentele groep 'extra middelen' te geven en de controlegroep te geven waar ze normaliter 'recht' op hadden. Verder kunnen beleidsmakers deelname aan het experiment vrijwillig maken; gevolg is wel dat niet zozeer het effect van het beleidsinstrument onderzocht wordt, als wel het effect van het aanbieden van het beleidsinstrument. Tot slot kunnen beleidsmakers eventuele politieke bezwaren omzeilen door te bevorderen dat derden gecontroleerde beleidsexperimenten opzetten, bijvoorbeeld onderzoekers verbonden aan universiteiten, in plaats van zulke experimenten zelf te initiëren.

Kosten en tijdsduur

Een nadeel van experimenteren is dat het opzetten, uitvoeren, volgen en evalueren van een goed experiment niet eenvoudig is en aanzienlijke kosten, inclusief de kosten van de maatregel zelf, met zich mee kan brengen. Bovendien zijn de resultaten soms pas na enkele jaren bekend. Experimenteren kan een kwestie van lange adem zijn, terwijl de tijdshorizon van veel politici vaak beperkt is tot een kabinetsperiode. Daar staat echter tegenover dat de baten zeer hoog kunnen zijn.

De beperkte duur van experimenten

Experimenten hebben een beperkte duur. Het directe gevolg kan zijn dat effecten die na het beëindigen van het experiment optreden niet worden waargenomen. Ook kunnen deelnemers anders reageren op een éénmalige interventie dan op een permanente interventie. Overigens kan bij de opzet van een experiment gekozen worden voor variatie in de duur van de interventie.

Sommige vragen kunnen niet goed beantwoord worden met experimenten

Experimenten zijn vooral geschikt om antwoord te geven op vragen op microniveau, bijvoorbeeld over het effect van een verlaging van de inkomstenbelasting op het arbeidsaanbod. Hoe deze veranderingen doorwerken op macroniveau, dat wil zeggen welke algemene evenwichtseffecten zich voordoen, kan met een experiment moeilijk worden vastgesteld. Ook is het in sommige gevallen niet goed mogelijk of zinvol om te experimenteren. In de praktijk lijkt bijvoorbeeld het door toeval toewijzen van een verhoging van het minimumloon aan een bepaalde groep bedrijven niet haalbaar.

Opschaling van kleinschalige experimenten

Een recent kritiekpunt is dat de resultaten van experimenten geen goed beeld geven van hetgeen gaat gebeuren als de interventie op grotere schaal wordt toegepast. Het is bijvoorbeeld mogelijk dat de doelpopulatie verandert door het opschalen van het experiment. Als een training voor personen in de bijstand nationaal beleid zou worden is het mogelijk dat bepaalde personen, in het vooruitzicht van de training, al bij voorbaat de bijstand verlaten. Het gevolg daarvan kan zijn dat de populatie van het experiment afwijkt van de populatie bij het opschalen van het experiment. Overigens geldt dit probleem evenzeer voor alle niet-experimentele technieken.

Niet alle personen in de experimentele groep doen mee aan de interventie

Bij veel experimenten zijn niet alle personen in de experimentele groep bereid om de interventie, bijvoorbeeld deelname aan een training, te ondergaan. De consequentie daarvan is dat het experiment niet het effect van deelname aan de training meet maar het effect van het aanbieden van de training. Om het effect van deelname aan de training vast te stellen zijn dan vaak additionele aannamen nodig. Dit lijkt echter niet een echt groot probleem te zijn omdat het beleid vaak juist geïnteresseerd is in het effect van het aanbieden van een interventie.

Spill-overs: de experimentele groep beïnvloedt de controlegroep

Om tal van redenen kunnen er spill-overs zijn tussen de experimentele en de controlegroep. Zo werd bij de evaluatie van de extra scholingsaftrek voor werknemers van 40 jaar en ouder gevonden dat de controlegroep van 39 jarigen de deelname aan scholing uitstelde.⁷ Een ander voorbeeld betreft de preventie van HIV-besmetting.⁸ De besmetting in de controlegroep kan in dat geval afhangen van het deel van de bevolking dat wordt behandeld. Bekend is ook dat spill-overs bij R&D-subsidies de resultaten van de controlegroep kunnen vertekenen. Hier kan zelfs een 'Catch-22' situatie optreden, omdat het genereren van spillovers vaak juist het beleidsdoel is. Wanneer de output van de experimentele groep en de controle groep weinig verschilt, kan dat betekenen dat het beleid niet succesvol is en weinig innovatie genereert. Het kan echter ook betekenen dat het beleid zeer succesvol is en veel innovatie genereert met grote uitstralingseffecten.⁹ Het probleem van spill-over-effecten speelt overigens evenzeer bij niet-experimentele benaderingen. Ook zijn er technieken ontwikkeld om het directe effect van het experiment te isoleren van het spill-over-effect.¹⁰

Externe validiteit

In hoeverre is een experimenteel resultaat gemeten op tijdstip t en locatie A ook elders geldig? En is een experimenteel resultaat te extrapoleren wanneer de beleidsintensiteit vergroot of verkleind wordt? De externe validiteit van de resultaten hangt af van de vergelijkbaarheid van de situaties. In het algemeen zullen resultaten uit andere landen of eerdere perioden niet zonder meer geldig zijn voor het Nederland van nu of straks. Dit probleem geldt evenzeer voor resultaten die op niet-experimentele wijze zijn verkregen.

3.4 Amerikaanse ervaringen met gecontroleerde beleidsexperimenten

Er zijn wereldwijd ontelbare voorbeelden van beleidswijzigingen en van evaluatie van beleidswijzigingen. Er zijn echter maar weinig voorbeelden van gecontroleerde sociaal-economische beleidsexperimenten. Greenberg e.a. (1999) kwamen voor de VS voor de periode 1962 tot 1996 tot een aantal van 143 gecontroleerde experimenten waarin totaal 293 beleidsinterventies getest zijn. Het aantal jaarlijks gestarte experimenten blijkt daarbij wel een stijgende trend te vertonen. Een aantal zaken vallen op:

- De meeste experimenten zijn gericht op sociaal-economisch zwakkere groepen en beogen te leren over de inkomens- en werkgelegenheidseffecten van onderwijsbeleid en (on-the-job) training, van overheidshulp bij het zoeken en behouden van werk, en van financiële en materiële bijstand. Steeds vaker worden de experimenteel gevonden effecten ingebed in een

⁷ Leuven en Oosterbeek (2003).

⁸ Philipson (2000).

⁹ Klette e.a. (2000).

¹⁰ Zie bijvoorbeeld Philipson (2000).

(maatschappelijke) kosten-baten-analyse. Slechts zeer weinig experimenten hebben betrekking op de midden- en hogere inkomensgroepen en op bedrijven.

- Deelname aan experimenten was ongeveer even vaak verplicht als vrijwillig.
- Het merendeel van de experimenten in de jaren 60 en 70 betrof nieuwe beleidsinstrumenten. In de jaren 80 en 90 lag de nadruk op experimenten met veranderingen in bestaande beleidsinstrumenten. Slechts zelden hadden experimenten tot doel bestaand beleid te evalueren. De macht van gevestigde belangen en ethische problemen verbonden aan het onthouden van bestaand beleid aan sommige deelnemers zouden dit kunnen verklaren.
- De kosten van experimenten - de kosten van het additionele beleid plus de operationele kosten - dalen in de loop der jaren. Verklaringen zijn het toenemende gebruik van administratieve databestanden boven experiment specifieke enquêtes de afnemende duur van experimenten c.q. sterkere focus op korte-termijn-effecten en de vermindering van het aantal beleidsinterventies per experiment. Met name de twee laatstgenoemde verklaringen kunnen als nadeel hebben dat het experiment minder kennis genereert.
- De opdrachtgever van het experiment is meestal een overheid, soms een niet-gouvernementele organisatie en een enkele keer een bedrijf.

Een aantal Amerikaanse experimenten heeft aantoonbaar bijgedragen aan fundamentele kennis over economisch gedrag en soms zelfs tot belangrijke beleidswijzigingen (zie tabel 3.1).

Mandell e.a. (2001) onderzoeken voor vijf gecontroleerde experimenten hoe en in welke mate de resultaten van gecontroleerde experimenten bijdroegen aan het beleidsproces. Zij concluderen dat de uitkomsten in alle gevallen hebben meegespeeld in beleidsbeslissingen, soms als een overtuigende bevestiging van een heersende mening, en soms met een zeer zwaar gewicht. De beslissing om een beleidsinstrument wel of niet breed in te zetten hangt in het algemeen echter af van vele factoren, waaronder de resultaten van het experiment. Verder inspireren experimenten tot nieuwe, al dan niet incrementele experimenten. Een belangrijke barrière voor de benutting van de resultaten van experimenten lijkt de tijd die noodzakelijkerwijs nodig is voor de opzet, uitvoering en evaluatie. Experimenten blijken vaak geïnitieerd te worden wanneer het beleidsissue 'hot' is; als de resultaten beschikbaar zijn, is de beleidsagenda vaak anders. Dit suggereert dat experimenten vooral van waarde zijn voor beleidsmakers, wanneer zij antwoord geven op slepende kwesties. Een experiment kan natuurlijk ook een onderwerp op de beleidsagenda zetten; voor geen van de vijf gevallen was dat echter het geval. Beleidsmakers blijken de resultaten van experimenten vaak als 'definitief' te beschouwen, ook als er in wetenschappelijke kringen nog flink gebakkeleid wordt. Een intensief communicatiebeleid blijkt nodig te zijn om beleidsmakers te informeren over de inhoud en reikwijdte van de belangrijkste uitkomsten van een experiment, zeker ingeval de omvang van het experiment beperkt is.

Geen enkel gecontroleerd experiment is volkomen gecontroleerd. De feitelijke uitvoering van een experiment blijkt altijd af te wijken van het ideale ontwerp. Krueger (1999) laat zien dat het STAR experiment met klassenverkleining in Tennessee (VS) geplaagd werd door een verdeling van leerlingen over experimentele en controle groepen die niet volledig random was. Wanneer de onderzoeker hiervoor niet corrigeert, kan de analyse een vertekend beeld geven van de werkelijke uitkomst van het experiment. Doet de onderzoeker dat wel, gebruik makend van de geobserveerde afwijkingen van het ideale experimentele ontwerp, dan kan zij niet alleen de werkelijke uitkomst van het experiment rapporteren, maar ook de effecten van onbedoelde, maar wel beleidsrelevante afwijkingen van het ontwerp. Krueger gebruikt bijvoorbeeld het overstappen van leerlingen van controle- naar experimentele groepen om het effect van de stabiliteit van de populatie van een klas op leerprestaties te schatten.

Tabel 3.1 Voorbeelden van experimenten met hun beleidsimpact

Experiment	Resultaat	Impact	Referentie
Health insurance experiment	precieze kennis van de prijsgevoeligheid van de vraag naar zorg	beleidsmakers kunnen de gezondheidseffecten van pakketwijzigingen inschatten	Brook et al 1983 Manning et al 1987
Job Training Partnership experiment	beleid heeft positief effect op participatie en inkomen, voor zover gericht op schoolgaande jongeren	beleidsmakers reduceren fondsen bestemd voor niet-schoolgaande jongeren	Bloom et al 1993
Work-welfare experiments	werkgerichte training en hulp bij het zoeken naar werk helpt	substantiële impact op de Family Support Act van 1998	Manpower Demonstration Research Corp.; Gueron en Pauly 1991

Bron: Burtless (1995), p.69

4 Natuurlijke beleidsexperimenten

Het opzetten van gecontroleerde beleidsexperimenten is soms moeilijk en duur in termen van tijd en geld. Daarom wordt de afgelopen jaren door onderzoekers in toenemende mate gezocht naar natuurlijke beleidsexperimenten. Dat zijn toevallige situaties in de werkelijkheid die vergelijkbaar zijn met een gecontroleerd experiment. Ook bij natuurlijke beleidsexperimenten zijn personen door toeval toegewezen aan een experimentele en een controlegroep. Het verschil met een gecontroleerd experiment is dat de toewijzing in dat geval niet door de onderzoeker is gedaan, maar het gevolg is van toevallige elementen in de regelgeving of door andere politieke of maatschappelijke oorzaken.

Voorbeelden van natuurlijke experimenten

In Nederland zijn recent twee studies uitgevoerd die gebruik maken van toevallige elementen in de regelgeving. De eerste studie evalueert het effect van de extra fiscale aftrek voor scholing van werknemers vanaf 40 jaar.¹¹ Werknemers net beneden de grens van 40 jaar kwamen niet in aanmerking voor de fiscale aftrek, zij vormen de controlegroep. Werknemers vanaf 40 jaar kwamen wel in aanmerking voor de fiscale aftrek, zij vormen de experimentele groep. Door de deelname aan scholing van werknemers rond de grens van 40 jaar te vergelijken wordt een schatting verkregen van het effect van de maatregel.

De tweede studie evalueert een recente maatregel waarmee extra geld voor computers en personeel wordt toegekend aan scholen met minstens 70 % achterstandsleerlingen.¹² Scholen net beneden deze grens kregen geen extra subsidie, scholen vanaf 70 % kregen wel extra subsidie. Door scholen rond de grens van 70 procent te vergelijken (en daarbij rekening te houden met het percentage achterstandsleerlingen) wordt een gecontroleerd experiment nagebootst. Scholen met minstens 70 procent achterstandsleerlingen vormen de experimentele groep, scholen net beneden de grens van 70 procent vormen de controlegroep. Het effect van de extra middelen op de prestaties van leerlingen is geanalyseerd door gebruik te maken van deze zogenoemde regressie-discontinuïteit.

De afgelopen jaren zijn studies die gebruik maken van natuurlijke experimenten op vele beleidsterreinen uitgevoerd, bijvoorbeeld onderwijs, zorg, veiligheid en arbeidsmarkt. Onderstaande tabel geeft ter illustratie een aantal voorbeelden van toepassingen van natuurlijke experimenten.

¹¹ Zie Leuven en Oosterbeek (2003).

¹² Zie Leuven e.a. (2003).

Tabel 4.1 Voorbeelden van studies die gebruik maken 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 Klauw (2001)

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

5 Gecontroleerde en natuurlijke experimenten en niet-experimentele analyse vergeleken

Natuurlijke experimenten versus gecontroleerde experimenten

Een voordeel boven gecontroleerde experimenten is dat natuurlijke experimenten vaak onopgemerkt door deelnemers, begeleiders of door de media plaatsvinden. Bij gecontroleerde experimenten is het mogelijk dat deelnemers zich anders gedragen door het feit dat ze zijn geselecteerd voor een experimentele groep (Hawthorne effecten), of juist doordat ze niet geselecteerd zijn (John Henry effecten, voorbeeld: de leraar van een klas die geen extra middelen is toegewezen, doet extra z'n best om te laten zien dat hij deze pech kan overwinnen (Krueger, 1999)). Dit kan de resultaten vertekenen. Hetzelfde geldt voor experimenten die veel aandacht krijgen in de media. Bij natuurlijke experimenten is dit niet aan de orde. Een ander voordeel van natuurlijke experimenten is dat politieke of ethische bezwaren vaak geen rol blijken te spelen omdat de random toewijzing niet gekozen wordt door de onderzoeker maar het gevolg is van andere factoren. Het nadeel van natuurlijke experimenten is dat de onderzoeker geen greep heeft op de interventie en op de doelgroep. Soms zijn ook aanvullende aannames nodig. Bij een regressie-discontinuïteit wordt bijvoorbeeld verondersteld dat niet waargenomen factoren geen rol spelen rond de discontinuïteit.

Gecontroleerde en natuurlijke experimenteren versus niet-experimenteren

De belangrijkste voor- en nadelen van de verschillende benaderingen zijn in tabel 3.2 samengevat. Gecontroleerde en natuurlijke experimenten kunnen overtuigend bewijs leveren over de effecten van beleid. Bij niet-experimentele analyse bestaat altijd de kans op vertekening door niet-waargenomen factoren. De baten van overtuigend bewijs over de effecten van beleid kunnen zeer hoog zijn. Niet-effectief beleid kan worden gestaakt en effectief beleid kan worden voortgezet. Het uitvoeren van een gecontroleerd experiment kan duur zijn en veel tijd vergen. Evaluaties op basis van natuurlijke experimenten zijn goedkoper en vergen minder tijd. Evaluaties op basis van niet-experimentele analyses zijn veelal ook goedkoper en sneller. Bij gecontroleerde experimenten is het goed mogelijk om het onderzoek te sturen. Bij natuurlijke experimenten is dat niet het geval omdat dit afhangt van toevallige omstandigheden. De sturing bij niet-experimentele analyse is ook afhankelijk van de mogelijkheden om bepaalde interventies waar te nemen en hierover data te verzamelen. De beschikbaarheid van data is het meest zeker bij gecontroleerde experimenten.

Tabel 5.1 Voor- en nadelen van verschillende evaluatiebenaderingen

Experiment	Gecontroleerd experiment	Natuurlijk experiment	Niet-experimentele analyse
Bewijs	overtuigend	overtuigend	kans op vertekening
Baten	hoog	hoog	niet duidelijk
Kosten	laag/hoog	laag	laag/ gemiddeld
Tijdsduur	kort/lang	kort	kort/lang
Sturing	groot	gering	gemiddeld/gering
Beschikbaarheid data	goed	onzeker	onzeker

6 Toepassing van beleidsexperimenten in Nederland

Veel beleidsvragen lenen zich voor experimenten, maar niet alle beleidsterreinen zijn even geschikt. Experimenten zijn vooral geschikt als de focus van de studie smal is, bijvoorbeeld gericht op het vaststellen van het effect van een duidelijk afgebakende interventie op één of enkele doelvariabelen.

Om te bepalen of een experiment moet worden uitgevoerd zou idealiter een afweging gemaakt moeten worden tussen de kosten en baten van een experiment vergeleken met alternatieven. Vaak zal dit neerkomen op het beoordelen van de vraag of de grotere betrouwbaarheid van experimentele resultaten opweegt tegen de extra kosten. Vooral bij 'slepende beleidskwesties' lijkt de waarde van overtuigend bewijs uit een gecontroleerd experiment groot ten opzichte van de kosten van zo'n experiment.

6.1 Gecontroleerde experimenten bij 'slepende beleidskwesties'

Bij tal van beleidsmaatregelen zijn gecontroleerde experimenten zeer geschikt om kennis te verwerven over wat werkt en wat niet werkt. Zo kan gedacht worden aan de effectiviteit van subsidies voor bedrijven of projecten op het terrein van milieu, innovatie, arbeidsomstandigheden, regiobevordering, export of scholing. Daarnaast zijn experimenten ook geschikt om de effecten van regelgeving rond het ondernemen te bepalen zoals vestigingseisen of regels voor specifieke beroepsgroepen als makelaars en advocaten. Gecontroleerde experimenten lijken echter bij uitstek geschikt voor het verkrijgen van 'evidence' bij slepende kwesties. Ter illustratie bespreken we hier vier slepende beleidskwesties.

Al vele jaren wordt allerlei beleid beproefd om de deelname aan technische studies te verhogen. Daarbij komt steeds weer aan de orde in hoeverre extra studiebeurzen voor studenten in technische opleidingen kunnen bijdragen aan een grotere deelname. Deze vraag kan worden beantwoord door een gecontroleerd experiment uit te voeren waarin een aselekt gekozen groep eindexamenkandidaten extra studiebeurzen wordt aangeboden. Vervolgens wordt bekeken in hoeverre de deelname aan technische studies in de experimentele groep afwijkt van de deelname in de controle groep.

Een tweede voorbeeld van een slepende kwestie is de armoedeval. In hoeverre kunnen financiële prikkels bijdragen aan reïntegratie van werklozen? In de Verenigde Staten zijn op dit terrein diverse experimenten uitgevoerd. In Nederland, en ook in andere Europese landen, zijn deze tot dusver uitgebleven.

Een derde voorbeeld van een slepende kwestie waar gecontroleerde experimenten belangrijke kennis aan het licht kunnen brengen is de additionaliteit van R&D-subsidieregelingen zoals de Nederlandse WBSO-regeling: hoeveel extra R&D induceert een R&D-subsidie? Deze vraag kunnen we beantwoorden met een gecontroleerd experiment waarbij we een aselekt gekozen groep bedrijven een hoger R&D-subsidietarief aanbieden dan

gebruikelijk. Een concrete implementatie van deze aanpak kan als volgt gaan. De WBSO kent drie verschillende subsidieregimes: naarmate de R&D-activiteit van het bedrijf toeneemt, daalt de marginale subsidievoet. Bij wijze van gecontroleerd experiment kunnen we een aselechte groep bedrijven die net in het volgende, lagere subsidieregime valt, toch de hoge subsidievoet uit het vorige regime aanbieden.

Een vierde slepende kwestie ligt op het gebied van de marktordening. Leidt een verlaging of verhoging van de toetredingsbarrières tot een betere marktwerking en meer innovatie? Experimenten met regels rond ondernemen, zoals vestigingseisen ten aanzien van kwaliteit en milieu, kunnen licht werpen op deze vraag. In een experiment zouden deze regels voor een aselechte groep regio's kunnen vervallen. De experimentele groep en de controle groep kunnen dan vergeleken worden op concurrentie-en innovatie-indicatoren.

6.2 Kansen voor evaluatie door handige vormgeving van beleid

De vormgeving van beleid kan een eerste stap zijn voor het vaststellen van de effecten van beleid. Toevallige elementen in regelgeving kunnen aselechte controlegroepen creëren. Door bewust 'toevallige elementen' in te bouwen in regelingen kan een overtuigende evaluatie worden mogelijk gemaakt. Anders gezegd, door een handige vormgeving van beleid is het mogelijk om na verloop van tijd hard bewijs te verkrijgen over de effecten van dit beleid. In de eerste plaats door in bepaalde gevallen subsidies aselekt toe te kennen, en in de tweede plaats door bewust verschillen (discontinuïteiten) in subsidieregelingen in te bouwen.

6.2.1 Aselekt toekennen van middelen

Bij de toekenning van subsidies zijn er vaak personen of organisaties die geen subsidie ontvangen. Uit deze groepen kan vaak een controlegroep worden gevormd. Een voorbeeld kan worden ontleend aan het Amerikaanse Head Start project. In dit project worden middelen voor voorschoolse educatie beschikbaar gesteld aan bepaalde doelgroepen, met name leerlingen uit arme gezinnen. Recent deed een adviescommissie voor de evaluatie van Head Start de aanbeveling om, in gevallen dat er te weinig middelen beschikbaar zijn voor alle deelnemers, de toewijzing van leerlingen aselekt te laten plaatsvinden. Daarmee wordt voor de toekomst een overtuigende evaluatie van het effect van het programma mogelijk gemaakt.

Een ander voorbeeld ligt op het terrein van tenders. Bij het ontwerp van een tenderregeling kan een echt experiment worden geconstrueerd. Onderzoeksvorstellen die net wel of net niet goed genoeg zijn voor een subsidie kunnen aselekt worden gesubsidieerd.¹³ Naarmate het moeilijker is om objectief vast te stellen welk onderzoeksvorstel van voldoende kwaliteit is en

¹³ Zie ook CPB (1999) en Jaffe (2002).

welke niet, ligt het bovendien meer voor de hand om dit te erkennen en subsidies via toeval toe te kennen.¹⁴

6.2.2 Het bewust inbouwen van discontinuïteiten in regelingen

Een tweede mogelijkheid is om bewust verschillen (discontinuïteiten) in subsidieregelingen in te bouwen, analoog aan het eerder besproken voorbeeld van de extra subsidie voor scholen met minstens 70 procent achterstandsleerlingen. Deze mogelijkheden kunnen bijvoorbeeld liggen bij verschillende subsidies voor bedrijven en projecten¹⁵ of bij de WBSO. Een andere mogelijkheid ligt op het terrein van ICT in het onderwijs. Op dit terrein zijn de afgelopen jaren omvangrijke investeringen gedaan. Over de effecten op de prestaties van leerlingen is tot dusver niets bekend. Deze effecten zijn moeilijk vast te stellen omdat er geen echte controlegroep bestaat. Door in de regelingen enige verschillen in het beschikbaar stellen van middelen in te bouwen ontstaan mogelijkheden om de effecten vast te stellen. Een voorbeeld op dit terrein is te vinden in een recente studie van Goolsbee & Guryan (2002).

6.3 Van pilots naar gecontroleerde experimenten

Op veel beleidsterreinen wordt gebruik gemaakt van 'pilots'. Het probleem daarbij is dat er meestal geen controlegroep wordt samengesteld en zeker geen aselechte controlegroep. De consequentie is dat de resultaten van 'pilots' minder overtuigend zijn. Hier blijven mogelijkheden onbenut. De stap van een pilot naar een gecontroleerd experiment bestaat uit het zoeken naar een aselechte controlegroep. Door hier systematisch aandacht aan te besteden kan belangrijke winst worden geboekt in de richting van 'evidence based policy'.

¹⁴ Bij kleine aantallen kan een aselechte toedeling aan de experimentele en de controlegroep soms toch verschillen tussen de groepen opleveren. Een enigszins gestuurde toedeling kan dan betere resultaten geven.

¹⁵ Zie www.subsidieshop.nl

7 Conclusies

De analyse leidt tot de volgende conclusies:

- De kosten van het voeren en continueren van niet-effectief beleid kunnen zeer hoog zijn.
- Gecontroleerde en natuurlijke beleidsexperimenten kunnen overtuigend bewijs leveren over de effecten van beleid. De baten van beleidsexperimenten kunnen daarom zeer hoog zijn.
- Gecontroleerde experimenten kunnen vooral voordelen opleveren als ze worden toegepast bij ‘slepende kwesties’.
- Natuurlijke experimenten zijn een relatief goedkope bron voor evaluaties. Ook is de termijn waarop resultaten van de evaluatie beschikbaar komen relatief kort.
- De vormgeving van beleid kan een eerste stap zijn voor de evaluatie van beleid. Mogelijkheden die momenteel onbenut lijken te blijven, zijn:
 - a. het aselekt toekennen van middelen bij een te grote vraag naar subsidies;
 - b. het bewust inbouwen van discontinuïteiten in subsidieregelingen.
- Door bij ‘pilots’ systematisch aandacht te besteden aan het zoeken naar een aselechte controlegroep kan meer overtuigende ‘evidence’ worden verkregen.

Referenties

Algemene Rekenkamer, 2001, *Bestrijding van onderwijsachterstanden*, Den Haag.

Algemene Rekenkamer, 2003, *Tussen beleid en uitvoering, lessen uit recent onderzoek*, Den Haag.

Angrist, L. and A.B. Krueger, 2002, Instrumental variables and the search for identification: from supply and demand to natural experiments, *Journal of Economic Perspectives*, vol. 15, no. 5, p. 69-87.

Burtless, G., 1995, The case for randomized field trials in economic and policy research, *Journal of Economic Perspectives*, Spring 1995, p. 63-84.

Brown, M.A., T.R. Curlee and S.R. Elliot, 1995, Evaluating technology innovation programs: the use of comparison groups to identify impacts, *Research Policy*, vol. 24(5), p. 669-684.

CPB, 1999, 'Het vijfde kaderprogramma voor R&D van de Europese Unie', werkdocument 111, juli 1999, Centraal Planbureau, Den Haag.

CPB, 2002, *De pijlers onder de kenniseconomie: opties voor institutionele vernieuwing*, Centraal Planbureau, Den Haag.

Dehejia, R.H. and S. Wahba, 2002, Propensity score matching methods for non-experimental causal studies, *The Review of Economics and Statistics*, vol. 84(1), p. 151-161.

Goolsbee, A. and J. Guryan, 2002, The impact of Internet subsidies in public schools, NBER Working Paper nr. 9090.

Greenberg D., Shroder, M. and M.Onstott, 1999, The social experiment market , *Journal of Economic Perspectives*, Summer 1999, p. 157-172.

Fraker, T. and R. Maynard, 1987, The adequacy of comparison group design for evaluations of employment-related programs, *Journal of Human Resources*, vol. 22, p. 194-227.

Heckman, J.J., 1979, Sample selection bias as a specification error, *Econometrica*, vol. 47, p. 153-161.

- Jaffe, A., 2002, Building programme evaluation into the design of public research-support programmes, *Oxford Review of Economic Policy*, Spring 2002, vol. 18(1), p. 22-34.
- Klette, T., J. Moen and Z. Grilliches,, 2000, Do subsidies to commercial R&D reduce market failures? Microeconomic evaluation studies, *Research Policy*, vol. 29, p. 471-495.
- Leuven, E. and H. Oosterbeek, 2003, Evaluating the effect of tax deductions on training, *Journal of Labor Economics*, forthcoming.
- Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink, 2003, The effect of extra funding for disadvantaged students on achievement, University of Amsterdam, SCHOLAR working paper 39/03.
- Krueger, A. B., 1999, Experimental estimates of education production functions, *Quarterly Journal of Economics*, vol. 114(2), p. 497-532.
- Majone, G., 1989, *Evidence, argument, and persuasion in the policy process*, Yale University Press, New Haven.
- Mandell, Greenberg, Linkz, 2001, The politics of evidence: the use of knowledge from social experiments in U.S. policymaking, working paper presented at the 2002 Annual Meeting of the American Political Science Association.
- Philipson, T., 2000, External treatment effects and program implementation bias, Technical Working Paper Series T0250, NBER.
- WRR, 2002, *Van oude kennis en nieuwe kennis, De gevolgen van ICT voor het kennisbeleid*, Den Haag.

Abstract

The effects of policy interventions are often unknown. Policy experiments offer opportunities to find convincing evidence about the impact of policy. This paper studies the pros and cons of controlled and natural policy experiments in the search for evidence of policy effectiveness. It concludes that controlled policy experiments are appropriate to deliver evidence on policy matters that have dragged on for a long time. Furthermore, it argues that many opportunities to provoke natural experiments in standard policy making exist, and that capture of these opportunities will increase our knowledge about policy effects importantly.