

Experimenten in de sociale zekerheid

Steeds vaker voeren het Ministerie van Sociale Zaken en Werkgelegenheid (SZW), UWV en de sociale diensten van gemeenten experimenten uit in de sociale zekerheid. Hiermee is een nieuwe fase aangebroken in de evaluatie van hun beleid, met nieuwe kansen maar ook met nieuwe valkuilen. Hoog tijd dus voor een checklist bij de uitvoering van veldexperimenten.

Na jaren van terughoudendheid lijkt de ban nu toch echt gebroken: het Ministerie van SZW, UWV en gemeenten gaan experimenteren met de uitvoering van sociale zekerheid. Zo maakten het Ministerie van SZW en de sociale diensten van gemeenten eind vorig jaar het voornemen bekend om te experimenteren met het toewijzen van re-integratietrajecten in de Wet werk en bijstand (WWB). SZW stelt de toegevoegde waarde (netto-effectiviteit) van re-integratie op de uitstroom naar werk centraal, en wil daarbij de afzonderlijke rol van de diagnosestelling van cliënten, typen instrumenten en die van uitvoerders vast stellen. Dit betekent een koerswijziging van het ministerie, met meer inspraak van wetenschappers: "Om meer inzicht te krijgen in de causaliteit tussen de inspanningen op het terrein van re-integratie en de als gevolg daarvan bereikte resultaten is, na consultatie met experts uit de wetenschap, besloten dit te onderzoeken met behulp van een experiment" (SZW, 2010).

Ook UWV start dit jaar verschillende experimenten. Zo is recent het plan opgevat om in een aantal kantoren te experimenteren met beloningsvormen voor het personeel. Centraal daarbij staat de directe koppeling van prestaties aan het aantal WW-uitkeringsgerechtigden van het kantoor; in de huidige situatie hangen prestaties namelijk nog af van indirecte prestatie-indicatoren, zoals klanttevredenheid en de tijdigheid van verstrekte uitkeringen. Een tweede initiatief van UWV richt zich op de meerwaarde van digitale vis-à-vis persoonlijke dienstverlening (UWV, 2010). Tot slot is er ook het zogenoemde *Fit-4-Work*-experiment, om innovaties bij de re-integratie van langdurige cliënten te stimuleren (UWV, 2010). *Fit-4-Work* richt zich primair op plaatsing in een baan, maar gebruikt daarbij ook de inzet van zorg, preventie, thuishulp en wijkpakket. Naast UWV zijn de sociale diensten en GGD-instellingen van de vier grootste Nederlandse gemeenten betrokken bij dit experiment.

Naast deze recente initiatieven zijn er ook experimenten die al enige tijd lopen. Zo is er sinds 1

januari 2009 het "Tijdelijk besluit bevordering arbeidsinschakeling alleenstaande ouders WWB", dat veertien experimentgemeenten in staat stelt om alleenstaande ouders met kinderen jonger dan twaalf jaar en die vanuit een uitkering in deeltijd werken een financieel extraatje van maximaal 120 euro per maand te geven (SZW, 2009). Over de resultaten van dit experiment volgt in de zomer van dit jaar een rapportage.

Sociale veldexperimenten

Een aantrekkelijke eigenschap van veldexperimenten is dat vrijwel iedereen – van beleidsmaker tot klantmanager – kan begrijpen hoe het werkt. Neem bijvoorbeeld de effectiviteit van re-integratietrajecten. Deze trajecten zullen bij een experiment willekeurig worden toegewezen aan een *treatment*-groep. Daartegenover staat de controlegroep die van een traject onthouden wordt. Het verschil in uitstroom naar werk geeft een zuivere schatting van het effect van re-integratie: selectie-effecten die normaal zouden optreden zijn dan immers al op voorhand geëlimineerd. Het enige criterium op basis waarvan een re-integratietraject al dan niet wordt toegewezen is de factor toeval, zodat gemiddeld genomen de eigenschappen van deelnemers en niet-deelnemers hetzelfde zullen zijn. Experimenten zijn niet alleen mogelijk op het niveau van individuele cliënten die al dan niet een *treatment* in de vorm van een traject of subsidie ontvangen. Ook valt namelijk te denken aan klantmanagers, kantoren of regio's met verschillende vormen van prestatiebeloning of meer professionele ruimte voor medewerkers; de combinatie van prestatiebeloning en meer beleidsruimte is bijvoorbeeld bij het eerstgenoemde toekomstige experiment van UWV van toepassing.

Wordt geabstraheerd van de kosten die ermee gemoeid zijn, dan is een experiment vrijwel altijd te verkiezen boven quasi-experimentele evaluatiemethoden die achteraf corrigeren voor selectie-effecten (Blundell en Costas Dias, 2009). Bij die methoden zijn namelijk betwistbare veronderstellingen nodig over de wijze waarop selectie bij de toewijzing van treatments zich manifesteert. Gezien deze voordelen van experimenten is het dus merkwaardig dat deze in de sociale zekerheid in Nederland zo lang onbenut zijn gebleven (Van der Klaauw, 2010). Een argument tegen het gebruik van experimenten is altijd geweest dat een willekeurige toewijzing niet rechtvaardig zou zijn naar cliënten of uitvoerders in de controlegroep. Verschillen in beleid tussen regio's, gemeenten of zelfs op het niveau van klantmanagers zijn er echter

PIERRE KONING

Programmaleider bij het
Centraal Planbureau

ook en deze worden wel breed geaccepteerd. Sterker nog: beleidsmakers beschouwen de ruimte die de WWB biedt om tot maatwerk te komen als belangrijke voorwaarde voor succes.

Drie valkuilen

Het starten van een experiment is in beginsel eenvoudig, maar het ook op deugdelijke wijze uitvoeren vergt meer toewijding. Dit geldt zowel voor beleidsmakers als voor uitvoerders in het veld. Voor hen is het dus van belang voldoende voorbereiding te treffen, en zich daarbij vooral ervan bewust te zijn wat er mis kan gaan. Ervaringen met experimenten leren namelijk dat selectie bij de toewijzing van treatments hardnekkig is, zo hardnekkig dat het de willekeurige, gerandomiseerde, toewijzing aan de start en ook de daadwerkelijke uitvoering kunnen doorkruisen. Het is daarom van belang dat uitvoerders een scherp oog ontwikkelen voor zulke versturende effecten. Drie hiervan komen het meest voor: ten eerste experimenten die eigenlijk *pilots* zijn; ten tweede vrijwillige deelname aan experimenten en ten derde experimenten zonder draagvlak.

Experiment of pilot?

Het mag geen verbazing wekken dat klantmanagers en uitvoerders van kantoren een experiment als een goede mogelijkheid beschouwen voor het ontwikkelen van nieuw beleid, zoals een nieuwe manier om hun klanten te benaderen, of om de kwaliteit van hun dienstverlening te verbeteren. Vaak bestaat daarbij al het voornemen en de hoop dat dit beleid niet te lang voorbehouden blijft aan een kleine groep cliënten of uitvoerders, maar snel breed zal worden uitgerold. Vanuit managementperspectief valt een dergelijke aanpak goed te verdedigen: kinderziekten komen snel boven water, nog voordat ze schade zouden kunnen aanrichten op grote schaal. Bovendien houden medewerkers vaak van een duidelijke koers van het management. Met andere woorden: experimenten moeten niet te vrijblijvend zijn en daadwerkelijk uitmonden in nieuw beleid. Toch is het de vraag of de onderzoeksopzet zich bij een latere, bredere toepassing nog wel als een experiment kwalificeert: niet-deelnemende klantmanagers of kantoren in de controlegroep kunnen zich namelijk ook voorbereiden op het nieuwe beleid. Dit effect zal vooral sterk zijn als zij geregeld contact met elkaar hebben, iets wat zeker niet ondenkbaar is bij organisaties met een hoofdkantoor (UWV) of bij gezamenlijke platformen voor overleg, zoals die voor de gezamenlijke sociale diensten (Divosa). Evenzo geldt dat individuele cliënten erop kunnen anticiperen dat ze na enige tijd ook in aanmerking voor een subsidie of traject komen. In dat geval sijpelt het effect van de treatment dus door in de controlegroep en levert het experiment geen

Waar pilots vooral richting moeten geven over hoe beleid effect kan hebben, is het bij experimenten de vraag of beleid effect heeft gehad

betrouwbare resultaten op. De managementterm pilot zou daarom meer op zijn plaats zijn dan die van experiment: het is de nadrukkelijke intentie om na verloop van tijd de treatment op te schalen, van pilot naar algemeen niveau.

Dat pilots en experimenten verschillende uitgangspunten kennen, is gelegen in het feit dat experimenten niet tot nieuw beleid hoeven te leiden. Daar waar pilots vooral richting moeten geven over hóé beleid effect kan hebben, is het bij experimenten de vraag óf beleid effect heeft gehad. Dit verschil lijkt misschien subtiel, maar kan grote gevolgen hebben. Dergelijke verschillen in uitgangspunten vormden in 2007 voor de Raad van State dan ook een reden tot afwijzing van eerdere plannen tot extra voorzieningen voor alleenstaande ouders in de WWB: "Indien daartoe aanleiding wordt gezien, zal aan het indienen van een definitief wetsvoorstel een experiment

voorafgaan. Zo'n experiment heeft dan als oogmerk het opdoen van ervaringen en inzichten die kunnen worden betrokken in het definitieve wetsvoorstel. In dit geval is echter sprake van de omgekeerde weg: het wetgevingsproces is geheel voltooid, en vervolgens wordt een experiment voorgesteld" (Raad van State, 2007).

Vrijwillige deelname

Een beproefde aanpak bij experimenten in de WWB is dat gemeenten zich zelf aan kunnen melden voor deelname aan een experiment. Dit is bijvoorbeeld het geval bij het experiment voor alleenstaande ouders in de WWB. Evenzo zal UWV geneigd zijn met enkele regiokantoren in overleg te treden, om zo te kunnen bepalen welke daarvan bereid zijn tot experimenten en ook de, administratieve, belasting kunnen dragen die hiermee gemoeid is. Het moge duidelijk zijn dat een dergelijke aanpak draagvlak creëert en het voor het ministerie of hoofdkantoor prettiger werken is. Dit geldt vooral bij het verzamelen van betrouwbare onderzoeksresultaten.

De keerzijde van vrijwillige deelname laat zich echter raden. Het valt moeilijk te ontkennen dat een vrijwillige keuze niet leidt tot een selectieve groep deelnemers. Bij uitvoerders mag verwacht worden dat de meest enthousiaste, vaak al goed presterenden onder hen zich zullen voegen onder de deelnemers. Bij cliënten in de WW of in de WWB kan het beide kanten op werken: kansrijke of juist kansarme cliënten kunnen zich melden of zelfs een duwtje in de rug krijgen van klantmanagers. In beide gevallen is een directe vergelijking van uitkomsten in de treatment- en controlegroepen dan weinig zinvol. Voor de echte puriteinen onder de onderzoekers is de term experiment dus niet van toepassing als deelname vrijwillig is.

Toch betekent dit niet dat vrijwilligheid bij de deelname aan experimenten de meting van effecten geheel onmogelijk maakt. In tegenstelling tot bij pilots is het met wat extra informatie toch mogelijk tot uitspraken te komen. Door de uitkomsten van treatment- en controlegroepen ook voorafgaand aan het experiment te meten, een zogenaemde nulmeting, is het namelijk mogelijk te corrigeren voor a priori verschillen in beide groepen. In die zin kan gesproken worden van een *difference-in-difference*-schatting. Hiermee kan worden geschat of het verschil tussen de treatment- en controlegemeenten toe- of afneemt als gevolg van het experiment. Deze methode is dus ook goed toepasbaar op het WWB-experiment. Maar daarnaast is ook een variant op dit thema mogelijk, namelijk door te kijken naar het verschil tussen de arbeidsdeelname van alleenstaande ouders in de WWB met kinderen onder de twaalf jaar en alleenstaande ouders in de WWB met kinderen tussen de twaalf en achttien jaar. Is dit verschil toegenomen in de experimentgemeenten, dan heeft het beleid voor de ouders met kinderen onder de twaalf jaar dus effect gehad. Ter controle kan worden bezien of het verschil gelijk gebleven is in de controlegemeenten. De conclusie is dus dat het experiment op slimme wijze dient te worden aangevuld met extra variatie in uitkomsten, tussen groepen of over de tijd. Door de beschikbaarheid van administratieve gegevens van het CBS is het goed mogelijk om experimenten met deze informatie te verrijken.

Geen treatment

In hun overzichtsartikel noemen Blundell en Costas Dias (2009) de uitval van deelnemers als belangrijkste gevaar bij experimenten. In de literatuur wordt de term *treatment* daarom soms ook wel vervangen door de term *intention-to-treat*. Dit wil zeggen dat bij een gerandomiseerde toewijzing van treatments deelnemers toch kunnen besluiten niet deel te nemen. Het negeren van deze uitval door de 'echte' treatmentgroep te schonen voor uitval is doorgaans niet vrij van complicaties, aangezien de beslissing om niet deel te nemen meestal selectief is. In dat geval is de onderzoeker weer goeddeels terug bij af: quasi-experimentele methoden zijn nodig om voor zulke selectie-effecten te corrigeren.

Vooraf wanneer het gaat om experimenten bij de uitvoering van sociale zekerheid is het gevaar van uitval groot. Sterker nog: klantmanagers kunnen formeel meedoen aan een experiment, maar het voorgenomen beleid negeren zodat uitval gewoonweg niet zichtbaar is. Als effecten van het experiment uitblijven, dan is het dus de vraag wat nu werkelijk is gebeurd: is het experiment eigenlijk wel uitgevoerd, of heeft het experiment zelf geen effect? Deze problemen spelen vooral wanneer het experiment de beleidsruimte van klantmanagers aan banden legt, zoals bij richtlijnen voor het uitvaardigen van sancties aan werklozen of richtlijnen voor de toewijzing van trajecten. Voor wat betreft het laatste vormen de ervaringen met een groot-schalig sociaal experiment in Zwitserland met computersystemen voor klantmanagers om hun cliënten passende re-integratietrajecten toe te wijzen een goede illustratie (Koning, 2008; Behnke *et al.*, 2007). Deze systemen werden niet bindend opgelegd, maar als hulpmiddel voor klantmanagers. Het effect van de systemen was niet meetbaar, aangezien klantmanagers het systeem vrijwel altijd terzijde legden. Kennelijk ervoeren klantmanagers het systeem als een inbreuk op hun beroepspeer, en soms zelfs als een toekomstige bedreiging voor hun baan. De vraag is dus hoe bindend je maatregelen aan uitvoerders wilt opleggen om tot een zinvol experiment te komen.

Het WWB-experiment

Kijken we naar de recente ervaringen in Nederland, dan valt tot dusver waarschijnlijk het meest te leren van het ver gevorderde WWB-experiment voor alleenstaande ouders. Een half jaar na de start van het experiment boden de onderzoekers hierbij een eerste tussenrapportage aan de Kamer aan (Knoef *et al.*, 2009). In eerste instantie richtte de aandacht zich daarbij vooral op moeilijkheden bij de aanlevering van gegevens van gemeenten over het beleid dat zij voeren voor de doelgroep. Inmiddels zijn deze problemen goeddeels verholpen en staan de beoogde effecten van het experiment centraal: meer arbeidsuren, meer inkomsten uit arbeid en meer volledige uitstroom uit de WWB.

In hoeverre is het WWB-experiment gevoelig geweest voor de drie genoemde valkuilen? Om met de eerste valkuil te beginnen: gezien de eerdere ervaringen met de Raad van State was het bij aanvang van het experiment dus niet langer de intentie om de extra regelingen voor alleenstaande ouders na afloop van het experiment breed uit te rollen. Het gaat dus om een experiment en niet om een pilot. Wel zijn sinds de installering van het nieuwe kabinet de omstandigheden gewijzigd. Het nieuwe regeerakkoord meldt namelijk dat de vrijstelling voor het werken in deeltijd voor alleenstaande ouders in de WWB zal worden verruimd en dat dit beleid, wellicht na enige aanpassingen, overal zal gaan gelden, ongeacht of het experiment effectief zal blijken of niet. Toch lijkt de schade van deze aankondiging voor het experiment beperkt. Tegen de tijd dat de vrijlating algemeen beleid wordt, zal het experiment namelijk voltooid zijn. De kans dat niet-participerende gemeenten voortijdig gelden hebben vrijgemaakt om al voorbereid te zijn op de nieuwe wet is dan ook minimaal.

De tweede valkuil, selectie bij de deelname van gemeenten, is een groter punt van zorg bij het experiment met alleenstaande ouders in de WWB. Eind 2008 informeerde het Ministerie van SZW de gemeenten over het experiment en kregen zij de mogelijkheid zich aan te melden. Daarbij was het doel om te

De beoogde effecten van het experiment: meer arbeidsuren, meer inkomsten uit arbeid en meer volledige uitstroom uit de WWB

komen tot maximaal dertig deelnemende gemeenten, gelijkmatig verdeeld over drie verschillende grootteklassen van 10.000 tot 50.000 inwoners, van 50.000 tot 150.000 inwoners, en van meer dan 150.000 inwoners. In eerste instantie gaven 27 gemeenten zich op voor het experiment. SZW achtte verdere selectie dus niet nodig. Maar vervolgens zagen sommige gemeenten alsnog af van deelname. Zes gemeenten vielen af in de beginfase, vier in het eerste half jaar, en uiteindelijk is het aantal deelnemende gemeen-

ten gereduceerd tot veertien. De redenen om af te zien van deelname waren ten eerste een gebrek aan capaciteit voor het opzetten van het experiment en het leveren van de data. Ten tweede ontstonden er andere prioriteiten door de economische crisis. Ten derde speelde de overtuiging een rol dat het experiment te weinig toe zou voegen aan het bestaande beleid voor de doelgroep door de gemeente zelf (Knoef *et al.*, 2009).

Deze sterke selectie van gemeenten suggereert dat vooral gemeenten zijn gaan meedoen die een grotere bijstandsproblematiek hebben. Maar het kan ook andersom: gemeenten die hun zaakjes goed op orde hebben, kunnen slim inhaken op het experiment om zo extra gelden aan te boren die zij niet langer uit het gemeentefonds hoeven te putten. Hoe dan ook, de keuze van treatment- en controlegemeenten is moeilijk als willekeurig of representatief te kwalificeren. In deze zin is er dan ook geen sprake van een experiment.

Bij de voorbereiding van het experiment is het risico van selectie niet onopgemerkt gebleven bij SZW. Vandaar dat voor aanvang van het experiment al is nagedacht over extra bronnen van exogene variatie in het gebruik van de instrumenten voor alleenstaande ouders in de WWB. Met exogene variatie doelen we dan op toevallige omstandigheden die ervoor zorgen dat vergelijkbare alleenstaande ouders net wel of net niet toegang krijgen tot extra instrumenten. Te denken valt dan bijvoorbeeld aan de leeftijdsgrens van kinderen van twaalf jaar bij het experiment. Zoals al eerder aangegeven biedt het WWB-experiment twee mogelijke bronnen van informatie om toch tot zuivere schattingen te komen: een nulmeting van het WWB-gebruik van de gemeenten bij de aanvang van het experiment, en micro-informatie over de verdeling van alleenstaande ouders binnen de WWB over kinderen boven en onder de twaalf jaar. De nulmeting stelt de onderzoeker daarbij in staat om de eerder genoemde difference-in-difference-strategie te volgen om het effect van het experiment te meten.

Ook de laatste valkuil, die van onvoldoende draagvlak bij de uitvoering, is mogelijk relevant bij het experiment in de WWB. Het probleem ligt daarbij

wel subtieler dan in het extreme geval dat uitvoerders beleidsinstrumenten naast zich neerleggen. Het is veeleer de vraag wat het instrument exact behelst. Het experiment schrijft namelijk voor dat naast de vereiste vrijstelling voor de doelgroep klantmanagers ook kunnen besluiten meer scholing aan te bieden, waarbij ouders een bonus ontvangen van vijftig euro bij voltooiing hiervan. Als alternatief kunnen gemeenten een arbeidspool opzetten voor alleenstaande ouders, als opstap naar reguliere arbeid. Voor zover het gebruik van deze initiatieven door de deelnemende gemeenten gemeld worden, is het wellicht mogelijk om inzicht te krijgen in de afzonderlijke effectiviteit van de instrumenten. Wanneer dan blijkt dat het experiment effect heeft gehad, is het interessant te bezien of het gaat om de werking van financiële prikkels of om de mogelijkheden tot scholing en deelname aan de arbeidspool. Op zich betekent dit een belangrijke verrijking van inzichten voor het gemeentelijke beleid.

Hiermee is echter niet alles gezegd. Onzeker is namelijk of klantmanagers binnen gemeenten zich geheel aan de regeling houden, of zij ten volle op de hoogte zijn van de bonussen, en of zij uitsluitend alleenstaande ouders met kinderen onder de vijf jaar wijzen op de mogelijkheden van een arbeidspool. Het ligt in de opzet van het experiment besloten dat gemeenten flexibel omgaan met het geboden instrumentarium, maar die flexibiliteit kan ook de effectmeting doorkruisen. Bovendien is het niet aannemelijk dat een klantmanager al deze beslissingen vastlegt als het uiteindelijke doel gewoonweg bestaat uit het vinden van een reguliere baan.

Conclusie

Uit het voorgaande mag duidelijk zijn dat het wel-slagen van experimenten afhangt van de toewijding en de alertheid van beleidsmakers en uitvoerders. Beleidsmakers zullen zich moeten bedwingen bij het uitrollen van beleid, hoe goed bedoeld en begrijpelijk hun handelen ook is. Daarnaast moeten uitvoerders zich soms laten dwingen om treatments toe te passen waar ze zelf misschien iets anders over denken. Maar bovenal zullen onderzoekers hun betrokkenheid moeten tonen bij experimenten, er dus dicht bovenop zitten en zich goed verdiepen in de beweegredenen van uitvoerders. Positiever geformuleerd bieden experimenten een uitgelezen mogelijkheid om beleidsmakers, uitvoerders en onderzoekers dichter bij elkaar te brengen.



LITERATUUR

- Behnke, S., M. Frölich en M. Lechner (2007) Targeting labour market programmes: results from a randomized experiment. *IZA Discussion paper*, 3085.
- Blundell, R. en M. Costas Dias (2009) Alternative approaches to evaluations in microeconometrics. *Journal of Human Resources*, 44(3), 565–640.
- Klaauw, B. van der (2010) Arbeidsmarktbeleid. *ESB*, 95(4588), 399.
- Knoef, M., J. Leenheer, J. Nelissen en K. de Vos (2009) *Tussenrapportage 2009. Experiment bevordering arbeidsparticipatie alleenstaande ouders WWB*. Tilburg: CentERdata.
- Koning, P. (2008) Tweedekans-re-integratie. *ESB*, 93(4533), 24–28.
- Ministerie van Sociale Zaken en Werkgelegenheid (2009) *Experiment: extraatje voor alleenstaande ouders die gaan werken*. www.gemeenteloket.minszw.nl, 13 januari.
- Ministerie van Sociale Zaken en Werkgelegenheid (2010) *Toezegging onderzoeksopzet experiment re-integratie*. www.gemeenteloket.minszw.nl, 1 december.
- Raad van State (2007) *Advies over zaaknummer W12.07.0229/III*. www.raadvanstate.nl, 31 augustus.
- UWV (2010) *Kennisagenda 2011*. www.uwv.nl, december.