

Handboek meetmethoden voor effectiviteit van activerend arbeidsmarktbeleid op persoonsniveau

Een handvat voor beleidsmakers en onderzoekers voor het vaststellen van de (netto-)effectiviteit van activerend arbeidsmarktbeleid

Onderzoek uitgevoerd door TNO en SEO Economisch Onderzoek op verzoek van het Ministerie van Sociale Zaken en Werkgelegenheid

Mei 2008

Voorwoord

Op verzoek van het Ministerie van Sociale Zaken en Werkgelegenheid hebben TNO en SEO Economisch Onderzoek gezamenlijk een “evaluatieprotocol” uitgewerkt, dat methoden aanreikt om de netto-effectiviteit van het activerend arbeidsmarktbeleid beter te kunnen vaststellen. Bij TNO heeft deze activiteit plaatsgevonden binnen het vraaggestuurd programma.

De eerste drie hoofdstukken van het evaluatieprotocol zijn van de hand van TNO, de overige van SEO. Het TNO-deel is uitgevoerd door Edwin de Vos en Wim Zwinkels, met medewerking van Jan Besseling, Peter Brouwer, Mike Klerkx en Daan Ooms. Het SEO-deel is uitgevoerd door Arjan Heyma en Chris van Klaveren.

Ten behoeve van het onderzoek was een begeleidingscommissie ingesteld, bestaande uit medewerkers van het Ministerie van Sociale Zaken en Werkgelegenheid, het Ministerie van Financiën, het Centraal Planbureau, de Raad voor Werk en Inkomen en het UWV. De onderzoekers danken de leden van deze commissie voor de prettige en constructieve samenwerking.

Inhoudsopgave

Voorwoord	i
Samenvatting	iv
1 Inleiding	1
1.1 Aanleiding	1
1.2 Doelstelling	1
1.3 Afbakening	2
1.4 Inhoud handboek	3
1.5 Leeswijzer	4
2 Evaluatie van effecten	5
2.1 Nut en noodzaak van effectiviteitsanalyses	5
2.2 Richtlijnen voor effectiviteitsanalyses vanuit de overheid	5
2.3 Soorten analyses	7
2.4 Effectiviteitsanalyses in de beleidscyclus	8
2.5 Effectiviteitsanalyses in de praktijk	10
3 Doelstellingen van beleid en uitkomstmaten voor evaluatie	12
3.1 Inleiding	12
3.2 Activerend arbeidsmarktbeleid	12
3.3 Over welke instrumenten praten we?	13
3.4 Doelstellingen van het activerend arbeidsmarktbeleid	15
3.5 Operationalisatie van doelstellingen in uitkomstmaten	16
3.6 Macro-economische (neven)effecten	20
3.7 Methoden, specificaties, bruikbare gegevens en rekenregels	21
4 Het meten van netto-effectiviteit	22
5 Evaluatiemethoden	33
5.1 Sociaal experiment	35
5.2 Natuurlijke of quasi experimenten	36
5.3 Het benutten van cross sectie variatie	37
5.4 Benutten van variatie in de tijd	41
5.5 Combinatie van evaluatiemethoden	45
5.6 Welke evaluatiemethode is het meest geschikt?	45
6 Toegepaste specificaties van methoden	49
6.1 Standaard econometrische vergelijkingen	49
6.2 Matching	50
6.3 Duurmodellen	58

7	Gegevens, bewerking en presentatie	65
7.1	Benodigde gegevens voor effectiviteitsanalyse	65
7.2	Bewerking van gegevens	69
7.3	Presentatie van effecten	73
	Bijlage 1: Literatuur	77
	Bijlage 2: Begrippenlijst	81

Samenvatting

Nederland investeert vele miljoenen euro's in de re-integratie en activering van werklozen, gedeeltelijk arbeidsongeschikten en niet-uitkeringsgerechtigden (NUG'ers). Daartoe richten instrumenten zich op de vraagkant (bijvoorbeeld door loonkostensubsidies), de aanbodkant (bijvoorbeeld door scholing) en op het efficiënt bijeenbrengen van vraag en aanbod (bijvoorbeeld door het beschikbaar stellen van arbeidsmarktinformatie voor werkzoekenden en werkgevers).

Dit handboek biedt een handvat voor de evaluatie van activerend arbeidsmarktbeleid. Het geeft de state-of-the-art weer van methoden voor onderzoek naar de effectiviteit van investeringen in activerend arbeidsmarktbeleid. De nadruk ligt daarbij op het vaststellen van de netto-effectiviteit op micro-niveau, dat wil zeggen de mate waarin individuele cliënten aan het werk zijn gekomen dankzij de inzet van beleid. Gebruikers van deze handleiding zijn beleidsambtenaren (Rijksoverheid, gemeenten, UWV, RWI etc.) die evaluaties (laten) verrichten naar het activerend arbeidsmarktbeleid en beleidsonderzoekers. Het handboek kan hen behulpzaam zijn bij onder meer het vertalen van een beleidsvraag naar een onderzoeksvraag, het treffen van voorzieningen voor juiste dataverzameling; een eenduidige hantering van het begrippenkader en het bevorderen van transparantie van methoden en presentatie van resultaten.

Om te beoordelen of de inzet van middelen effectief is geweest, dient gekeken te worden naar de beleidsdoelen. Investerings in activerend arbeidsmarktbeleid kunnen bijvoorbeeld worden gedaan om de arbeidsparticipatie te bevorderen, vacatures te vervullen en werkloosheid te verminderen. De beleidsdoelen dienen daartoe geoperationaliseerd te worden in uitkomstmaten. Veranderingen in de uitkomstmaat vormen een basis voor de vaststelling van effectiviteit.

Effecten van beleidsinterventies kunnen worden gemeten door de groep die te maken heeft met een beleidsinterventie (de proefgroep) te vergelijken met de groep die daar niet mee te maken heeft (de controlegroep). Wanneer beide groepen goed vergelijkbaar zijn, dan kunnen verschillen in de uitkomstmaat (bijvoorbeeld de werkloosheidsduur) tussen beide groepen volledig worden toegeschreven aan de werking van de beleidsinterventie. Dat verschil is de netto-effectiviteit of toegevoegde waarde van de interventie. In de praktijk zijn de proefgroep en controlegroep niet altijd goed vergelijkbaar. Ze kunnen verschillen in samenstelling en in de kans op een beleidsinterventie: er is dan sprake van een selectieve inzet van beleidsinstrumenten. Afhankelijk van de uitkomstmaat, het type beleidsinterventie en de beschikbare gegevens voor analyse, kan door middel van specifieke evaluatiemethoden worden gecorrigeerd voor samenstelling- en selectie-effecten. In het handboek wordt een overzicht gegeven van beschikbare evaluatiemethoden met een handreiking welke evaluatiemethode in welke situatie het beste kan worden ingezet. Vervolgens wordt een overzicht gegeven van de benodigde gegevens voor een effectiviteitsanalyse en hoe met deze gegevens het beste kan worden omgegaan. Ten slotte wordt ingegaan op de manier waarop gemeten netto-effecten dienen te worden gepresenteerd. Belangrijk is om te realiseren dat effecten van beleidsinterventies die zijn vastgesteld met behulp van een effectiviteitsanalyse op een afgebakende onderzoekspopulatie ook alleen geldig zijn voor die afgebakende onderzoekspopulatie.

De hoofdstukken 1 tot en met 4 van het handboek zijn gericht op beleidsmakers die in de opzet en uitwerking van het activerend arbeidsmarktbeleid rekening willen houden met de manier waarop dit beleid kan worden geëvalueerd. De technische uitwerking van zo'n evaluatie vindt plaats in de hoofdstukken 5 tot en met 7 en is met name gericht op onderzoekers.

1 Inleiding

Kennis over de effectiviteit en doelmatigheid van de ingezette instrumenten en middelen bij het aan het werk houden en aan het werk helpen van mensen is cruciaal. Een belangrijk doel is dan ook het ontwikkelen van handvatten voor het meten van effectiviteit in een complexe beleids- en uitvoeringsomgeving.

1.1 Aanleiding

De evaluatie van het activerend arbeidsmarktbeleid is de afgelopen jaren een steeds belangrijker onderwerp geworden. In 2001 deed een interdepartementale werkgroep de aanbeveling om dit soort evaluaties te standaardiseren. Vanuit overwegingen van financiële verantwoording heeft de Rijksoverheid belang bij een dergelijke standaard, bijvoorbeeld in het kader van de VBTB (Van Beleidsbegroting Tot Beleidsverantwoording). Maar ook gemeenten en ketenpartners hebben belang bij standaardisatie van evaluatiemethoden. Zo kan een verantwoord inzicht gegeven worden in de mate waarin middelen doelmatig besteed zijn. En tenslotte hebben ook beleidsonderzoekers belang bij standaardisatie, want zij kunnen resultaten beter onderling vergelijken.

Lange tijd was er geen eenduidige systematiek als het gaat om goede effectiviteits-analyse van activerend arbeidsmarktbeleid. Het is daarom niet duidelijk of verschillen in uitkomsten over de effectiviteit van beleid veroorzaakt worden door verschillen in de gebruikte gegevens, de manier waarop gegevens worden bewerkt, de gehanteerde onderzoeksmethodiek en/of de wijze waarop resultaten gepresenteerd zijn.

Vandaar dit initiatief om een handleiding voor effectstudies te schrijven, met daarin aandacht voor onder meer verschillende uitkomstmaten, de relatie tussen effectiviteit en omgevingsvariabelen en de relatie tussen effectiviteit en onderzoeksmethoden. Gebruikers van deze handleiding zijn beleidsambtenaren die evaluaties (laten) verrichten naar het activerend arbeidsmarktbeleid en beleidsonderzoekers.

1.2 Doelstelling

Het hoofddoel van het onderhavige handboek is het zo eenduidig en transparant mogelijk schetsen van een methodologisch kader voor het vaststellen van de effectiviteit van het activerend arbeidsmarktbeleid.

Het handboek beoogt te voorzien in de verantwoordings- en onderzoeksbehoeften van diverse belanghebbende partijen, zoals de Rijksoverheid, gemeenten, UWV, CWI, RWI, etc. Het handboek is derhalve niet uitsluitend onderzoekstechnisch van aard, maar bevat ook handvatten en biedt een hulpmiddel voor het in de praktijk uitvoeren van evaluatie-onderzoek.

Het handboek kan voor beleidsambtenaren en onderzoekers behulpzaam zijn bij onder meer:

- Het vertalen van een beleidsvraag naar een onderzoeksvraag.
- Het nemen van maatregelen voor dataverzameling zodat vooraf rekening kan worden gehouden met de evalueerbaarheid van beleid.
- De communicatie tussen beleidsmakers en onderzoekers, zowel tussen beide groepen als onderling, door middel van een eenduidige hantering van het begrippenkader.
- Ondersteuning van beleidsmakers bij het opstellen van startnotities/offerteverzoeken en het beoordelen van offertes voor onderzoek.

¹ IBO (2001) "Aan de slag", eindrapport van de werkgroep Toekomst van het arbeidsmarktbeleid, Interdepartementaal Beleidsonderzoek, Den Haag.

- Ondersteuning en begeleiding van onderzoekers bij de uitvoering van het onderzoek.
- Het bevorderen van transparantie en controleerbaarheid van resultaten.
- Het uitdragen en mainstreamen van ontwikkelde kennis.

Een verder wegliggend doel is het verbeteren van de effectiviteit op basis van de gegenereerde kennis. Dit betreft zowel het aanpassen van het beleid als het verbeteren van de uitvoering.

Het handboek is geen opgelegde methodiek voor de uitvoering van evaluatieonderzoek. Het doel is vooral om bestaande methodieken bij elkaar te brengen, convergentie te bewerkstelligen en waar mogelijk standaarden te stellen. Voor sommige vraagstukken bestaan er verschillende methodieken naast elkaar. In dit handboek worden aanbevelingen gedaan over de idealiter te volgen werkwijze. Overigens kunnen er in de praktijk argumenten zijn om daarvan af te wijken, bijvoorbeeld omdat benodigde data niet of onvolledig aanwezig zijn. Wij gaan hier in dit handboek verder op in door ook bij beperktere informatie de aangewezen analysemethoden uit te werken. Ook kunnen de korte doorlooptijd van het onderzoek of de beperkt beschikbare middelen een beperking opleggen aan de uitvoerbaarheid van de evaluatie volgens het handboek. In deze gevallen biedt dit handboek een kader om afwegingen te maken en bewust en gefundeerd de keuze te maken om af te wijken van de aanbevolen werkwijze.

1.3 Afbakening

In de loop van de tijd is al het nodige gebeurd rondom de evaluatie van activerend arbeidsmarktbeleid. In hoofdstuk 2 wordt uitgebreid ingegaan op evaluaties van het Nederlandse beleid. Tot op heden is er echter geen standaard gesteld voor de evaluatie van activerend arbeidsmarktbeleid. Op sommige andere beleidsterreinen is hier wel sprake van. Met name de OEI-systematiek, die als standaard geldt bij de evaluatie van infrastructuurprojecten (zie hiervoor hoofdstuk 2). Anders dan de OEI-leidraad besteedt het onderhavige handboek niet zozeer aandacht aan de maatschappelijke kosten en baten van het beleid. Dit komt mede doordat de meetmethoden nog sterk in ontwikkeling zijn. In dit handboek staat dan ook het gebruik van onderzoeksmethoden naar kwantificering van effectiviteit centraal. Daarbij ligt de nadruk op netto-effectiviteit op micro-niveau. Het vaststellen hiervan kan worden gezien als een cruciale stap in het kunnen meten van bijvoorbeeld de kosteneffectiviteit of voor het uitvoeren van een (maatschappelijke) kosten-baten analyse.

Het handboek is geen protocol voor alle interventies die op het terrein van arbeidsmarktbeleid worden genomen. Het is toegespitst op het activerend arbeidsmarktbeleid. Dit omvat interventies die bedoeld zijn om de werkloosheid te bestrijden, de uitstroom naar werk te bevorderen en de aansluiting tussen vraag en aanbod op de arbeidsmarkt te verbeteren.

We onderscheiden onder meer de volgende instrumenten²:

- Vaardigheidstraining voor het zoeken naar werk (sollicitatietraining, cursus effectief onderhandelen, empowerment e.d.).
- Scholing.
- Sancties.
- Uitstroompremies.
- Instroomsubsidies en tijdelijke loonkostensubsidies voor reguliere banen.
- Directe werkgelegenheidscreatie, dat wil zeggen gesubsidieerde banen waaraan de eis van additionaliteit is verbonden.
- Subsidies voor werklozen die een eigen bedrijf willen starten.

² Overigens is deze lijst niet uitputtend. Zie voor een uitvoerige beschrijving van instrumenten het volgende hoofdstuk, paragraaf 2.1 en De Koning (2006).

- Het registreren en beschikbaar stellen van arbeidsmarktinformatie over vacatures en werkzoekenden, als mede het faciliteren van het zoekproces.
- Actieve bemiddeling, bijvoorbeeld het selecteren van geschikte vacatures voor werkzoekenden en het selecteren van geschikte werkzoekenden voor werkgevers die op zoek zijn naar personeel. Ook sancties (of de dreiging ermee) vallen hieronder.
- (Intensieve) begeleiding van werkzoekenden bij het zoeken naar werk: regelmatige voortgangsgesprekken, monitoring van de zoekactiviteiten, e.d.

Deze instrumenten van activerend arbeidsmarktbeleid bespreken we nader in hoofdstuk 3³. Soms worden meerdere instrumenten tegelijk of na elkaar aangeboden en door meerdere ketenpartners uitgevoerd. Zo kan een Workfirst-aanpak bestaan uit diagnosestelling door een keuringsbedrijf, intensieve begeleiding van werkzoekenden door re-integratiebedrijf A én de regie daarop bij de gemeente. Het handboek is bedoeld om de stand van zaken weer te geven van de meetmethoden voor effectiviteit van activerend arbeidsmarktbeleid op persoonsniveau, maar het is niet bedoeld voor alle soorten studies betreffende deze instrumenten. Het beperkt zich tot studies met gebruik van microdata, waarbij gegevens van individuen worden gebruikt om effecten vast te stellen op microniveau die kunnen worden uitgesplitst naar doelgroepen (zie hoofdstuk 4).

In dit handboek wordt ook kort aangegeven hoe activerend arbeidsmarktbeleid op macro-niveau geëvalueerd kan worden op basis van microgegevens. Het precies vaststellen van macro-effecten op de economie en de arbeidsmarkt (o.a. loonvorming) valt echter buiten het bereik van dit handboek. Dergelijke analyses vragen namelijk om toepassing van andere onderzoeks- en analysestechnieken, zoals het gebruik van arbeidsmarktmodellen op macroniveau⁴.

1.4 Inhoud handboek

De inhoud van het handboek is in het volgende schema per hoofdstuk beknopt weergegeven.

Tabel 1.1: Inhoud van het handboek per hoofdstuk

Hoofdstuk	Beschrijvingen: inhoud (per paragraaf)
1. Inleiding	<ul style="list-style-type: none"> • Aanleiding: het belang van het evalueren van beleid • Doelstelling handboek • Afbakening handboek • Beschrijving inhoud handboek
2. Evaluatie van effecten	<ul style="list-style-type: none"> • Het nut en de noodzaak van het evalueren van effectiviteit van activerend arbeidsmarktbeleid voor diverse stakeholders • Rol effectiviteitsanalyses in de beleidscyclus • (Wettelijke) voorschriften en leidraden rond het evalueren van beleid, zoals VBTB, RPE 2006 en OEI-leidraad • Onderscheid naar typen effectiviteit (bruto-effectiviteit, netto-effectiviteit, kosteneffectiviteit, etc.)
3. Doelstellingen van beleid en uitkomstmaten voor evaluatie	<ul style="list-style-type: none"> • Het doel van arbeidsmarktbeleid (waaronder re-integratiebeleid) • Onderscheid naar typen arbeidsmarktbeleid • Re-integratie-instrumenten als onderdeel van het arbeidsmarktbeleid en hun doelstelling • Relatie tussen uitkomstmaten, effectiviteit en de doelstellingen van beleid

³ Workfirst wordt hier niet apart genoemd want bestaat in de regel uit meerdere instrumenten, waaronder indicatiestelling, werkbegeleiding en/of sollicitatietraining.

⁴ Zie bijvoorbeeld het rapport Reinventing the Welfare State van het CPB (De Mooij, 2006).

4. Het meten van netto-effectiviteit	<ul style="list-style-type: none"> • Relatie tussen soorten effectiviteitsmeting en doelstelling van beleid
5. Evaluatiemethoden	<ul style="list-style-type: none"> • Overzicht van de belangrijkste evaluatiemethoden voor het meten van effecten
6. Toegepaste specificaties van methoden	<ul style="list-style-type: none"> • Mogelijkheden van empirische specificaties binnen evaluatiemethoden met een afweging tussen complexiteit en toegevoegde waarde van de onderzoeksopbrengsten
7. Gegevens, bewerking en presentatie	<ul style="list-style-type: none"> • Data eisen afhankelijk van evaluatiemethode en uitkomstmaat. Bewerking van onderzoeksgegevens. Presentatie van onderzoeksresultaten
Bijlage	<ol style="list-style-type: none"> 1. Literatuur 2. Begrippenlijst

1.5 Leeswijzer

De hoofdstukken 1 tot en met 4 van het handboek zijn gericht op beleidsmakers die in de opzet en uitwerking van het activerend arbeidsmarktbeleid rekening willen houden met de manier waarop dit beleid kan worden geëvalueerd. De technische uitwerking van zo'n evaluatie vindt plaats in de hoofdstukken 5 tot en met 7 en is met name gericht op onderzoekers. In hoofdstuk 5 wordt een overzicht gegeven van de belangrijkste methoden voor het meten van effectiviteit. Daar wordt ook aangegeven in welke situaties welke methode het meest geschikt is. Hoofdstuk 6 besteedt uitvoerig aandacht aan verschillende mogelijkheden voor empirische specificaties binnen de meest gebruikte methoden, afhankelijk van de doelpopulatie, de uitkomstmaat en de beschikbare gegevens voor de effectiviteitsanalyse. Hoofdstuk 7 ten slotte gaat in op de benodigde gegevens voor het kunnen uitvoeren van de verschillende evaluatiemethoden, de manier waarop die gegevens dienen te worden bewerkt en de manier waarop onderzoeksresultaten kunnen worden gepresenteerd.

2 Evaluatie van effecten

Effectiviteitsanalyses kunnen een belangrijke rol spelen in de beleidscyclus. Dit geldt zowel vooraf bij het maken van beleidskeuzes, als achteraf bij het vaststellen of de beoogde effecten bewerkstelligd zijn tegen de laagste kosten. Vanuit de Rijksoverheid en gemeenten zijn inmiddels handreikingen en leidraden ontwikkeld voor het uitvoeren van effectiviteitsanalyses in specifieke domeinen. Dit handboek is bedoeld voor de evaluatie van activerend arbeidsmarktbeleid.

2.1 Nut en noodzaak van effectiviteitsanalyses

De Rijksoverheid, UWV/CWI en gemeenten investeren vele miljoenen euro's in de re-integratie en activering van werklozen, gedeeltelijk arbeidsongeschikten en niet-uitkeringsgerechtigden (NUG'ers). Het thema activering en re-integratie⁵, gekoppeld aan het concept van een 'activerende sociale zekerheid', blijft ook de komende jaren actueel. Daarnaast wordt geïnvesteerd in preventie van uitval uit de arbeidsmarkt en anticipatie op een veranderende arbeidsmarkt, ofschoon dit veelal een taak is van sociale partners en individuele werkgevers en werknemers.

De interventies zijn in de eerste plaats nodig om werkloosheid en kwalitatieve verschillen tussen vraag en aanbod van arbeid te beperken. Daarbij wordt steeds belangrijker dat 'niemand aan de kant mag blijven staan'. Omdat daarmee publieke middelen gemoeid zijn, dienen gemeenten, CWI en UWV dit dan wel op de meest kosteneffectieve wijze uit te voeren.

Burgers spreken via parlement en gemeenteraden de bestuurders nadrukkelijk aan op de resultaten van beleid. Ook de Rekenkamers hebben steeds vaker belangstelling in onderwerpen als effectiviteit, doelmatigheid, kwaliteit en noodzaak van beleid.

De overheid streeft naar een zo optimaal mogelijke inzet van publieke middelen, zodat de beoogde effecten zoveel mogelijk worden bereikt en tegen zo laag mogelijke kosten. Belangrijke informatie voor de beoordeling van beleid is de verhouding tussen de maatschappelijke opbrengsten en de maatschappelijke kosten⁶. Voor een kosteneffectief beleid dienen de kosten van verstrekte subsidies, scholing of voorzieningen om mensen aan het werk te houden of te krijgen op termijn lager te zijn dan de maatschappelijke baten. Baten zijn bijvoorbeeld de bespaarde uitkeringsgelden en de extra opbrengsten die de staat ontvangt uit de omzet- en loonbelasting. Naast deze financiële baten kunnen er ook nog andere maatschappelijke baten zijn, zoals het welzijn van burgers en de sociale cohesie. Wanneer de baten groter zijn dan de kosten is beleid kosteneffectief. Door beperkte budgetten moet een keuze gemaakt worden tussen verschillende beleidsopties die op zichzelf effectief kunnen zijn. Hiervoor is het van belang de effectiviteit van beleidsopties te bepalen. Wanneer de effectiviteit bekend is kan nagegaan worden welke beleidskeuzes de grootste baten opleveren voor het gegeven budget.

2.2 Richtlijnen voor effectiviteitsanalyses vanuit de overheid

Vanuit de Nederlandse overheid is er nadrukkelijk aandacht voor de effectiviteit van beleid. In 1999 verscheen de nota 'Van Beleidsbegroting Tot Beleidsverantwoording' (VBTB) van het Ministerie van Financiën. Deze nota bevatte voorstellen voor een begroting en verantwoording nieuwe stijl. Centraal stond de wens om te komen tot een duidelijker koppeling tussen input (personeel, materieel, geld), throughput (het productieproces), output (in de vorm van geleverde producten en diensten) en outcome (tot welke resultaten of effecten heeft het productieproces

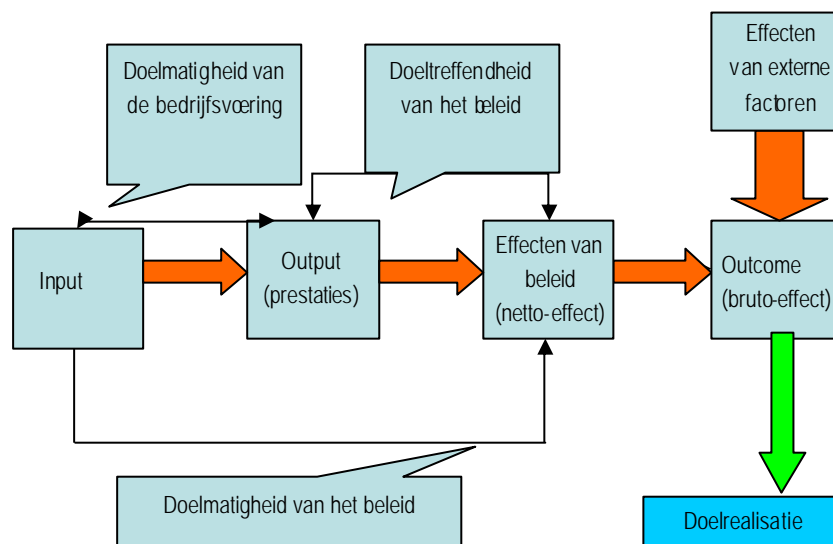
⁵ Met 'activering', 'terugkeer naar werk', 'werkhervatting' en 'het vinden van werk' worden in dit handboek zowel het vinden van een nieuwe functie bij een nieuwe werkgever, hervatting van de oude functie bij de oude werkgever, een nieuwe functie bij de oude werkgever als werkhervatting als zelfstandige bedoeld.

⁶ Teulings in: Proceedings of the conference 'Measuring effects: Roads to success in rehabilitation' How to measure re-employment policies in The Netherlands?' In De Vos ea, 2004.

geleid?). De VBTB moet als instrument dienen om de doelmatigheid en doeltreffendheid van beleid op landelijk niveau te bevorderen. Invoering van de VBTB moedigde in die hoedanigheid ook aan tot het uitvoeren van effectiviteitsanalyses. Overigens bleek uit een evaluatie dat de VBTB in de praktijk nog moeilijk uitvoerbaar was voor diverse ministeries (IOFEZ, 2004).

Het Ministerie van Financiën heeft in 2001 de Regeling Prestatiegegevens en Evaluatieonderzoek (RPE) opgesteld. Deze RPE is, naar aanleiding van de VBTB-evaluatie, herzien en omgedoopt tot de Regeling Periodiek Evaluatieonderzoek en Beleidsinformatie 2006 (RPE 2006)⁷. De RPE 2006 bepaalt hoe, binnen de rijksoverheid, moet worden omgegaan met beleidsinformatie en evaluatieonderzoek. In lijn met de VBTB gaat het erom het beleid vooraf beter te onderbouwen en achteraf beter te evalueren. Figuur 2.1 geeft het begrippenkader weer, zoals dat in de RPE 2006 wordt gehanteerd.

Figuur 2.1: Begrippenkader RPE 2006



Bron: RPE 2006

Leidraden in specifieke domeinen

Dit handboek is bedoeld als richtlijn voor de effectiviteitsanalyse van activerend arbeidsmarktbeleid. In deze zin is het handboek vergelijkbaar met andere richtlijnen zoals de VBTB en de RPE. De VBTB en de RPE bevatten generieke richtlijnen met betrekking tot effectiviteitsanalyses. Deze richtlijnen zijn bedoeld voor evaluaties van beleid op het niveau van de Rijksoverheid. In specifieke domeinen worden daarnaast ook leidraden ontwikkeld voor effectstudies. Het bekendste voorbeeld in Nederland is de OEI-leidraad.

Het Onderzoeksprogramma Effecten Infrastructuur (OEI) is op initiatief van het Ministerie van Verkeer en Waterstaat en het Ministerie van Economische Zaken uitgevoerd door het Centraal Planbureau, in samenwerking met diverse andere (economische) onderzoeksbureaus. Dit heeft in 2000 geleid tot een leidraad voor het uitvoeren van effectstudies bij grote infrastructuurprojecten (Eijgenraam et al., 2000). Bij dergelijke projecten gaat het om substantiële investeringen, waarvan de baten zich over tientallen jaren uitstrekken. De OEI-leidraad maakt gebruik van de maatschappelijke kosten-baten analyse als evaluatiemethode. De OEI-leidraad is overigens vooral gericht op ex ante evaluatieonderzoek.

⁷ De RPE wordt op zijn beurt geïntegreerd in het Rijksbegrotingsvoorschrift (RBV).

2.3 Soorten analyses

Begrippen als effectiviteit en doelmatigheid leggen een relatie tussen de input en de effecten van beleid. Deze effecten dienen aan te sluiten bij de doelstellingen van beleid. Doen ze dit niet, dan spreken we van neveneffecten.

We onderscheiden drie methoden waarmee de effectiviteit van beleidsinterventies of van afzonderlijke beleidsinstrumenten kan worden gemeten: de effectiviteitsanalyse, de kosteneffectiviteitsanalyse en de kosten-batenanalyse.

Met een *effectiviteitsanalyse* kan worden achterhaald of de gestelde doelen ook daadwerkelijk worden gehaald dankzij de interventie. Vinden mensen bijvoorbeeld vaker een baan *doordat* zij een re-integratietraject hebben gevolgd? Ook kan worden achterhaald welke instrumenten het beste voor wie werken. Zie bijvoorbeeld Hekelaar et al. (2006), Lauwerys et al. (2000) en Rubbrecht et al. (2005).

De effecten die we in de praktijk waarnemen, zijn zogenaamde bruto-effecten. Figuur 2.1 illustreert dat deze bruto-effecten tot stand komen door zowel het beleid als door externe factoren. Waar het om gaat is welke toegevoegde waarde de beleidsinstrumenten die we onderzoeken concreet hebben bij het bereiken van deze effecten. Dit noemen we de netto-effectiviteit.

Deze netto-effectiviteit kunnen we dus niet direct waarnemen. In de praktijk van het evaluatieonderzoek gaan we daarom uit van de bruto-effecten, die we waarnemen, waarna we door een correctie voor externe factoren het netto-effect van het beleid bepalen.

Als voorbeeld noemen we de effectiviteit van re-integratie. Het effect van re-integratie wordt gemeten als een verandering van de kans op werkhervatting. De bruto-effectiviteit wordt vastgesteld als het percentage werkhervattingen bij cliënten die re-integratie-instrumenten hebben ontvangen. Onbekend is dan nog of die cliënten zonder de inzet van deze re-integratie-instrumenten ook of wellicht zelfs sneller een baan zouden hebben gevonden. Door te corrigeren voor de kans op werkhervatting zonder de inzet van instrumenten, kan de netto-effectiviteit van de re-integratie-instrumenten worden afgeleid. Stel bijvoorbeeld dat een re-integratiebedrijf 100 klanten krijgt voor re-integratie en er na verloop van tijd 40 uitstromen naar werk. Wanneer het re-integratiebedrijf alle klanten met rust gelaten zou hebben en er zouden er toch 30 uitstromen naar werk, is het netto-effect van de ingezette trajecten 10 klanten.

Bovenstaand voorbeeld geeft aan dat netto- en bruto-effecten veelal niet aan elkaar gelijk zijn en in sommige gevallen zelfs van teken kunnen verschillen. De netto-effectiviteit kan positief zijn terwijl de bruto-effectiviteit negatief is, vice versa. Vooraf is niet duidelijk wat de relatie is tussen bruto- en netto-effecten, omdat dit afhangt van de selectieve inzet van beleidsinstrumenten. Het maakt voor de effectmeting nogal wat uit of de groep die een interventie krijgt veel kansrijker of juist kansarmer is dan de groep die geen interventie krijgt, waarmee vergeleken wordt. Is er een groot verschil dan is het verschil in bruto-effect tussen beide groepen een onbetrouwbare benadering van de effectiviteit van de interventie. In hoofdstuk 5 wordt teruggekomen op de methodologische consequenties hiervan.

Een *kosteneffectiviteitsanalyse* gaat een stap verder. De effecten van het beleid (instrument) worden in een kosteneffectiviteitsanalyse afgezet tegen de gemaakte kosten. Met een kosteneffectiviteitsanalyse kan de inzet van verschillende instrumenten worden vergeleken op de verhouding tussen kosten en effectiviteit. Het kan bijvoorbeeld doelmatiger zijn om van een bepaald budget goedkope trajecten in te kopen dan dure trajecten die weliswaar effectiever zijn maar waarmee per uitgegeven Euro minder mensen aan een baan geholpen worden. De effecten worden in een kosteneffectiviteitsanalyse echter niet in geldelijke baten uitgedrukt: er kan dus geen uitspraak worden gedaan over het saldo van kosten en baten in financiële termen.

In een *kosten-batenanalyse* wordt wel een uitspraak gedaan over het saldo van kosten en baten. Zowel de input als de effecten worden in geld uitgedrukt. Hiermee kan worden nagegaan in hoeverre de opbrengsten de kosten van de interventie overtreffen. Kosten-baten analyses zijn er in diverse soorten en hangen af van het perspectief (de aanbesteder van re-integratie, de uitvoerder, de cliënt of de maatschappij).

De kosten zijn hierbij de uitgaven die voor de uitvoering van het beleid worden gedaan. Dit zijn allereerst de directe uitgaven voor instrumenten die worden ingezet. Andere kosten zijn bijvoorbeeld personeelskosten voor de uitvoering en/of controle van het beleid. Deze elementen zijn redelijk eenvoudig in geld uit te drukken. Bij de baten van beleid is dit niet altijd het geval. Activerend arbeidsmarktbeleid kan maatschappelijke baten hebben die moeilijk in geld zijn uit te drukken. Voorbeelden zijn minder criminaliteit en een schonere openbare ruimte. De kosten-batenanalyse is in enkele studies in het sociale domein toegepast: zie Groot et al. (2003), Veld in 't et al. (2005) en Kok (2007). Ook Social Return on Investment (SROI) is in feite een kosten-batenanalyse, waarbij effecten in geld worden uitgedrukt. Voor een handleiding en voorbeeldcases van SROI zie Scholten (2003) en Scholten (2005).

Het in geld uitdrukken van dit soort baten valt buiten het kader van dit handboek. De toepassing van onderzoeksmethoden op het terrein van activerend arbeidsmarktbeleid is nog sterk in ontwikkeling. Belangrijk is te realiseren dat zowel een kosteneffectiviteitsanalyse als een kosten-batenanalyse gebaseerd is op een effectiviteitsanalyse. De kwaliteit van de effectiviteitsanalyse stuurt in grote mate de kwaliteit van beide andere analyses. Daarom concentreren we ons in dit handboek op de manier waarop die effectiviteitsanalyse het beste kan worden uitgevoerd.

2.4 Effectiviteits analyses in de beleidscyclus

Om goede beleidsbeslissingen te kunnen nemen, is inzicht in de effecten van verschillende beleidskeuzen van groot belang. Aan de hand van effectiviteitsanalyses vooraf (ex ante) kan bepaald worden welk beleid gevoerd moet worden om de beoogde effecten te bewerkstelligen. Met behulp van effectiviteitsanalyses achteraf (ex post) kan worden vastgesteld of het beleid ook daadwerkelijk tot deze effecten heeft geleid.

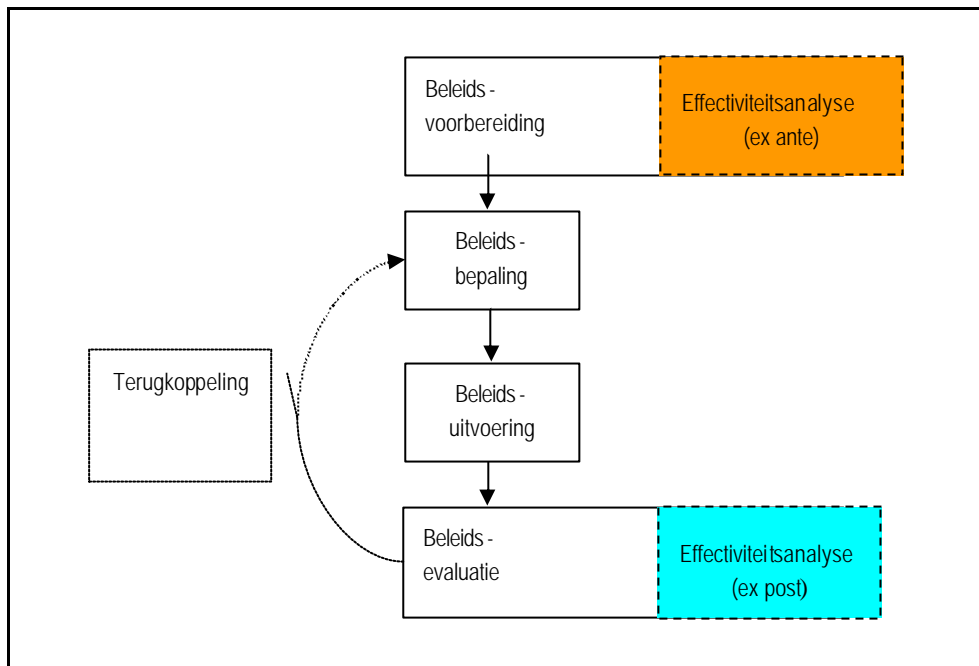
De uitkomsten van de effectiviteitsanalyses kunnen vervolgens weer gebruikt worden om het beleid aan te passen. Op deze manier spelen effectiviteitsanalyses een belangrijke rol in de verschillende fasen van de beleidscyclus (zie figuur 2.2).

Wanneer een effectiviteitsanalyse uitvoeren?

In de praktijk is het niet altijd eenvoudig om het juiste moment te bepalen voor het uitvoeren van een effectiviteitsanalyse. Een ex ante effectiviteitsanalyse wordt voorafgaand aan de beleidsbepaling uitgevoerd, zodat de resultaten gebruikt kunnen worden bij het maken van de beleidskeuze. Om een ex ante evaluatie tijdig uit te voeren, dient dus vooraf duidelijk te zijn wanneer de beleidsbeslissingen genomen (moeten) worden. Ex ante effectiviteitsanalyses kunnen bijvoorbeeld noodzakelijk zijn om financiering te krijgen voor beleid of voor een project.

Aangezien de uitvoering van beleid een continu proces is en vaak tussentijds bijgesteld wordt, is het moment voor een ex post effectiviteitsanalyse minder eenduidig vast te stellen. In feite zijn ex post evaluaties dan ook veelal (tussentijdse) evaluaties van het beleid over de afgelopen periode. Dit kan het einde van een kalenderjaar zijn, of een ander moment waarop er verantwoording afgelegd moet worden over het gevoerde beleid.

Figuur 2.2: Effectiviteitsanalyses in de beleidscyclus



Ex ante effectiviteitsanalyse

Dit handboek is gericht op kwantitatieve methoden om de effectiviteit van beleid vast te stellen. Voor een goede toepassing van deze methoden zijn empirische gegevens nodig. Bij ex ante evaluaties zijn deze empirische gegevens per definitie niet aanwezig, omdat het over toekomstig uit te voeren beleid gaat. Toch zijn er wel manieren om zinvolle ex ante evaluaties uit te voeren, met gebruik van empirische gegevens.

Ten eerste kan gebruik worden gemaakt van ervaringen met en resultaten van eerder uitgevoerd beleid. Er kan gekeken worden naar het beleid in voorgaande jaren of beleid dat elders uitgevoerd is.

Dit elders uitgevoerde beleid kan bijvoorbeeld ontleend worden aan ervaringen in andere landen. Gemeenten kunnen ook gegevens ontleenen aan vergelijkbare gemeenten. In feite vormen op deze manier de resultaten van ex post effectiviteitsanalyses van eerder uitgevoerd beleid de input voor de ex ante effectiviteitsanalyse.

Ten tweede kunnen, vóór daadwerkelijke invoering van nieuw beleid, verschillende beleidsalternatieven experimenteel ingezet worden. Door experimenten uit te voeren worden op kleine, gecontroleerde, schaal empirische gegevens verzameld, voor gebruik in de ex ante evaluatie. Op het terrein van activerend arbeidsmarktbeleid zijn slechts weinig ex ante evaluaties uitgevoerd. Voorbeelden zijn enkele studies naar samenwerking tussen bedrijfsverenigingen en arbeidsbureaus via één-loket en één-dossier modellen en evaluaties van het gebruik van WW-experimenten⁸.

Ook met gebruik van empirische gegevens uit het verleden, zoals hierboven beschreven, blijven ex ante evaluaties betrekking hebben op een toekomstige situatie. Er moet dus op worden gelet dat de gehanteerde gegevens een betrouwbare en bruikbare inschatting zijn voor resultaten in de toekomst.

⁸ Doodeman et al. (1996), Van den Berg et al. (2003), Gorter en Kalb (1996), Neset al. (2002).

Een belangrijk resultaat van ex ante effectiviteitsanalyse kan zijn dat gericht gegevens worden verzameld die noodzakelijk zijn voor een juiste ex post effectiviteitsanalyse.

Ex post effectiviteitsanalyse

Een ex post effectiviteitsanalyse stelt met kwantitatieve methoden de effectiviteit van het uitgevoerde beleid vast. Hiervoor is het noodzakelijk om tijdens de beleidsuitvoering de benodigde empirische gegevens vast te leggen en te verzamelen. Al bij de start van de uitvoering van het beleid dient daarom vastgesteld te worden welke effecten moeten worden gemeten en welke gegevens hiervoor nodig zijn.

2.5 Effectiviteitsanalyses in de praktijk

In Nederland is het onderzoek naar de effectiviteit van activerend arbeidsmarktbeleid voor zowel de rijksoverheid als gemeenten de laatste jaren sterk in ontwikkeling. Onder meer Heyma en De Graaf-Zijl (2005), Zwinkels (2005), Heyma, Zwinkels en Van Seters (2003), Van der Heul, e.a. (2003) en De Graaf-Zijl et al. (2006) gaan in op de effecten van re-integratie voor arbeidsgehandicapten en/of werklozen. Er is onderzoek gedaan naar de netto-effectiviteit van re-integratietrajecten in Rotterdam (Hekelaar et al., 2006) en Amsterdam (Kok et al., 2006). Voor het gemeentelijke domein zijn ook landelijke analyses verricht (Kok et al., 2006; De Graaf-Zijl et al., 2006, Groot et al., 2007).

Daarnaast is er onderzoek naar de maatschappelijke kosten en baten van gesubsidieerde arbeid en sociale activering (Brouwer et al., 2006) en de maatschappelijke kosten en baten van re-integratie (Kok et al. 2006). De onderzoeksmethoden daarvoor en de aannames die daarbij gemaakt worden zijn nog sterk in ontwikkeling en aan discussie onderhevig (CPB, 2007). Daarbij worden methoden en instrumenten ontwikkeld om beter zicht te krijgen op de maatschappelijke baten.

Voorbeelden gebruik effectiviteitsanalyses: profiling en benchmarking

Beleidsmakers kunnen effectiviteitsanalyses op verschillende manieren gebruiken in het beleidsproces. Twee voorbeelden die we hier noemen zijn profiling en benchmarking.

Profiling

Bij *profiling* gaat het om de vraag welk beleidsinstrument in welk geval het meest geschikt is. Bij de inzet van re-integratie-instrumenten is dan de vraag of de juiste persoon het juiste traject aangeboden krijgt. We stellen dan niet alleen de gemiddelde effectiviteit vast van verschillende re-integratiediensten, maar differentiëren dit. Op grond van objectieve en persoonlijke kenmerken van cliënten zonder baan stellen we vast wat de kans is op werkhervatting op verschillende momenten na instroom in de WW of bijstand. Daardoor is het mogelijk vast te stellen bij welke werkzoekende welke trajecten moeten worden ingezet en of dat, gezien de verwachte werkhervattingskansen, een optimale allocatie van middelen is en of *de juiste klant het juiste traject ontvangt*. Met deze wetenschap kan (beperkt) budget zo worden ingezet dat over het geheel meer positieve beleidseffecten behaald worden.

Benchmarking

De prestaties op het gebied van re-integratie zijn om te zetten in een *benchmark*. Bij effectiviteitsanalyse van re-integratie bestaat namelijk de mogelijkheid om te kijken naar verschillen in de toegevoegde waarde van uitvoerders (re-integratiebedrijven) of van aanbesteders (gemeenten/UWV).

Inzicht in bijvoorbeeld de prestaties van re-integratiebedrijven via een openbare benchmark kan de effectiviteit op drie manieren vergroten. Allereerst beschikken opdrachtgevers over een instrument om het meest geschikte re-integratiebedrijf te selecteren. Daarnaast loont het voor re-integratiebedrijven om te investeren in effectiviteit, aangezien deze prestaties directer

zichtbaar worden. Ten derde kan informatie over effectiviteit gebruikt worden om de aanpak te verbeteren. In Australië bleek dat na invoering van het Benchmarksysteem met Star Ratings (een sterrensysteem dat de effectiviteit van de re-integratiedienstverlening weerspiegelt) de plaatsingscijfers van re-integratiebedrijven spectaculair waren gestegen. Ook in Nederlandse studies komt naar voren dat sommige re-integratiebedrijven vier maal zo succesvol zijn dan hun slechtste concurrenten⁹.

Relatie tussen proces en effect

Diverse studies wijzen op een relatie tussen proces en effect. Proceskenmerken blijken soms van grote invloed te zijn op het effect¹⁰. De effectiviteit van een interventie is bijvoorbeeld niet een absoluut gegeven, maar hangt sterk af van het moment van inzet.

Uit een studie voor de RWI blijkt dat hoe eerder een interventie bij bijstandsgerechtigden wordt ingezet hoe groter de kans op een geslaagd traject is¹¹. Daarnaast zijn er onderdelen van het beleidsproces die een tijdelijk negatief effect op het beoogde doel kunnen hebben. Een voorbeeld hiervan is een scholingsinstrument dat wordt ingezet om de baankans voor bijstandsgerechtigden te vergroten. Hierbij kan het voorkomen dat tijdens de scholingsperiode de kans op het vinden van werk afneemt en de uitkeringsduur toeneemt, omdat mensen ‘vast’ zitten in de scholing. Het positieve effect van het afronden van scholing is hierbij meestal wel groter dan het negatieve effect tijdens scholing.

Bij beleidsevaluaties is het daarom van belang het hele proces te evalueren en proceskenmerken te kwantificeren en mee te nemen in de effectanalyse. Daarnaast kan een kwalitatieve procesevaluatie een onderbouwing vormen van de resultaten van de effectiviteitsanalyse.

⁹ Zwinkels (2005).

¹⁰ Zie onder meer het onderzoek over sturingsinformatie voor lokaal re-integratiebeleid van het RWI (Beek et al., 2008).

¹¹ Groot (2007).

3 Doelstellingen van beleid en uitkomstmaten voor evaluatie

3.1 Inleiding

De meerwaarde van het handboek voor bestaand beleid ligt in het vaststellen van een vergelijkbare methode om de effecten van beleid op een juiste wijze te meten. Voor toekomstig beleid geldt dat een handboek het mogelijk maakt bij de vormgeving van het beleid vooraf rekening te houden met de evalueerbaarheid. In het vorige hoofdstuk bespraken we dat een belangrijke voorwaarde voor evalueerbaarheid is dat het doel van het activerend arbeidsmarktbeleid helder is, omdat de te bepalen effecten afhangen van de doelstellingen. In dit hoofdstuk gaan we daarom in op de doelstelling van het activerend arbeidsmarktbeleid en de daarvan afgeleide uitkomstmaten. Deze uitkomstmaten kunnen vervolgens gebruikt worden bij het vaststellen van de netto-effectiviteit. We spreken hier bewust van uitkomstmaat en niet van effectmaat. Het kwantitatieve effect (of zo men wil de uiteindelijke effectmaat) wordt bepaald aan de hand van (verschillen in) de uitkomstmaten. Bijvoorbeeld: werkloosheidsduur is een uitkomstmaat en de effectmaat is een verkorting van de werkloosheidsduur met 25%.

Zoals in hoofdstuk 2 is vermeld, beperken we ons in dit handboek tot de effectiviteitsanalyse. De kosten-effectiviteit en de kosten-baten analyse verlopen voor elke uitkomstmaat volgens een vast stramien waarbij de effectiviteitsanalyse als basis gebruikt wordt. In het geval van een kosteneffectiviteitsanalyse worden de kosten van de uitvoering van beleid bij de analyse betrokken en in het geval van de kosten-baten analyse worden de baten zoveel mogelijk in geld uitgedrukt.

3.2 Activerend arbeidsmarktbeleid

In brede zin omvat activerend arbeidsmarktbeleid alle sociaal-economische regelingen die tot doel hebben de werking van de arbeidsmarkt te verbeteren. Het handboek spitst zich toe op het activerend arbeidsmarktbeleid bestaande uit (wettelijke) regelingen en de uitvoering daarvan met als doel de werkloosheid te bestrijden, de uitstroom naar werk te bevorderen en de aansluiting tussen vraag en aanbod op de arbeidsmarkt te verbeteren. Daartoe richten de regelingen zich op de vraagkant (denk aan loonkostensubsidies), de aanbodkant (denk aan scholing) en op het efficiënt bijeenbrengen van vraag en aanbod. Een voorbeeld van dit laatste is matching van werkzoekenden en vacatures en het beschikbaar stellen van arbeidsmarktinformatie voor werkzoekenden en werkgevers. Ook het bevorderen van ketensamenwerking bij de uitvoering kan een doel zijn van activerend arbeidsmarktbeleid.

Een verbeterde werking van de arbeidsmarkt resulteert op haar beurt in een aantal maatschappelijke gewenste uitkomsten zoals een verhoging van de arbeidsparticipatie en arbeidsproductiviteit en, in het verlengde daarvan, een geringere aanspraak op uitkeringen van gemeenten en UWV.

Soms richt het activerend beleid zich niet direct op betaalde arbeid, maar is het doel met name maatschappelijke participatie in enigerlei vorm.

Activerend beleid, re-integratiebeleid en flankerend beleid

Daar waar instrumenten van activerend arbeidsmarktbeleid zich richten op een transitie van de sociaal-economische positie van personen met een sociale zekerheidsuitkeringen en van de zogeheten NUG'ers (niet-uitkeringsgerechtigden) in termen van inactiviteit naar reguliere arbeid, spreken we van re-integratiebeleid. Re-integratiebeleid is dus een onderdeel van activerend arbeidsmarktbeleid. Een voorbeeld van activerend arbeidsmarktbeleid zonder directe re-integratiedoelstelling is het faciliteren van dagbestedingsactiviteiten gefinancierd vanuit de AWBZ. Dergelijke voorzieningen richten zich weliswaar op een verbetering van de sociale

positie van de deelnemer maar niet in termen van het vinden van regulier werk. In die zin behoren ze dus niet tot re-integratiebeleid.

Het activerend arbeidsmarktbeleid wordt aangevuld door het zogenaamde flankerend beleid (soms ook genoemd passief arbeidsmarktbeleid, bijvoorbeeld door de OESO, zie onderstaand kader). Voorbeelden van flankerend beleid zijn onder meer kinderopvangregelingen en schuldhelpverlening. Schuldsanering kan bedoeld zijn om de toeleiding naar werk te ondersteunen, door te bewerkstelligen dat schuldeisers niet alle looninkomsten opeisen. Vanuit deze optiek bezien bevat het passief flankerend beleid dus ook een activerend element. Flankerend beleid vergemakkelijkt de overstap naar regulier werk.

OECD: For analytical and policy purposes, the OECD splits measures into so-called “active” and “passive” measures. The former comprise a wide range of policies aimed at improving the access of the unemployed to the labour market and jobs, jobrelated skills and the functioning of the labour market while the latter relate to spending on income transfers, namely unemployment benefits and early retirement pensions.

Uit: What works and for whom: a review of OECD countries' experiences with active labour market policies (Martin en Grubb, 2001)

3.3 Over welke instrumenten praten we?

Zoals hierboven gesteld is het doel van activerend arbeidsmarktbeleid om door middel van interventies de werking van de arbeidsmarkt te verbeteren. De negen instrumenten kunnen zowel door overheidsorganisaties als door private instellingen als re-integratiebedrijven uitgevoerd worden.

De volgende instrumenten vormen onder meer onderdeel van die interventies:

- Instrumenten die zich richten op de aanbodkant:
 1. Vaardigheidstraining voor het zoekproces (sollicitatietraining, cursus effectief onderhandelen, empowerment, e.d.).
 2. Scholing.
 3. Financiële prikkels zoals sancties en uitstrooppremies.

Het doel van deze instrumenten is het verbeteren van de werking van de arbeidsmarkt onder meer door middel van het vergroten van de employability en de motivatie van de werkzoekende. Door trainingen en scholing wordt de kwaliteit van het aanbod verbeterd en beter afgestemd op de kennis- en kundebehoefte vanuit de vraagkant van de arbeidsmarkt. Door middel van uitstrooppremies wordt de werkzoekende financieel geprikkeld om een baan te aanvaarden. Sancties op uitkeringen in verband met inactief zoekgedrag kunnen eveneens een belangrijk instrument zijn. Vaak is de cliënt naast genoemde instrumenten onderhevig aan (de dreiging van) een sanctie.

- Instrumenten die zich richten op de vraagkant:
 4. Financiële prikkels zoals sancties en instroomsubsidies en tijdelijke loonkostensubsidies voor reguliere banen.
 5. Directe werkgelegenheidscreatie (gesubsidieerde (additionele) banen, stages, vrijwilligerswerk, etc.).
 6. Subsidies voor starten eigen bedrijf.

Deze instrumenten creëren een vraag naar arbeid die zonder overheidsinterventies door de reguliere arbeidsmarkt niet tot stand zou zijn gebracht. Daardoor worden werkzoekenden in staat gesteld om op een gesubsidieerde werkplek voor de markt relevante werkervaring op te doen.

Doorstroom op basis van de opgedane werkervaring naar een niet-gesubsidieerde baan staat daarbij vaak voorop. Regulier werk hoeft dan niet per sé het eindstation te zijn, althans voor diegenen die vanwege een functionele beperking minder verdien capaciteiten hebben. Voorbeelden daarvan zijn onder meer beschutte werkplekken binnen de sociale werkvoorziening voor werkzoekenden met een SW-indicatie. Bij dergelijke werkplekken ligt een grote nadruk op het belang van sociale activering en maatschappelijke participatie.

Subsidies voor het starten van een eigen bedrijf richten zich op de aanbodzijde maar kunnen direct (voor de ondernemer) en indirect (als het bedrijf gaat uitbreiden) ook zorgen voor werkgelegenheid.

- Instrumenten die zich richten op het matchen van vraag en aanbod:
 7. Registreren en beschikbaar stellen van arbeidsmarktinformatie over vacatures en werkzoekenden, faciliteren van het zoekproces.
 8. Actieve bemiddeling.

Deze laatste groep van instrumenten verbetert de arbeidsmarktwerking door het verbeteren van de informatievoorziening en het versnellen van de aansluiting tussen vraag en aanbod. Een snellere aansluiting verkort onder meer de uitkeringsduur waardoor de (maatschappelijke) kosten dalen.

- Instrumenten die zich richten op het aanbod van arbeid en op het matchingsproces:
 9. (Intensieve) begeleiding van werkzoekenden.

In de literatuur wordt intensieve begeleiding van werkzoekenden ook meegenomen als onderdeel van activerend arbeidsmarktbeleid. Een voorbeeld van intensieve begeleiding is een jobcoach die de werkzoekende ondersteunt door middel van mentale ondersteuning, het verzorgen van vaardigheidstrainingen, het attenderen op geschikte vacatures, etc.

Omdat dergelijke instrumenten zich richten op zowel het arbeidsaanbod als op het matchen van vraag en aanbod, worden ze als aparte categorie genoemd.

Re-integratieondersteuning kan uit meerdere interventies bestaan. De inzet van één interventie (instrument) sluit de inzet van een ander niet uit. In de praktijk worden dan ook - vanaf het moment dat een cliënt wordt opgepakt totdat deze aan de slag gaat - verschillende interventies per cliënt ingezet. Men spreekt dan van re-integratieondersteuning. Soms gebeurt het toepassen van verschillende interventies opeenvolgend en door verschillende ketenpartners. Re-integratieondersteuning wordt bijvoorbeeld ingezet indien er over een langere periode meerdere obstakels overbrugd moeten worden om de werkzoekende succesvol te re-integreren.

Een combinatie van interventies kan een ander effect teweeg brengen dan de inzet van de interventies afzonderlijk. Voor de effectiviteit is van belang welke interventies onderdeel uitmaken van de re-integratieondersteuning en met welke volgorde de interventies worden ingezet. Daarbij is ook van belang of het gaat om een door de opdrachtgever bewust gekozen combinatie van interventies of één die toevallig tot stand komt (bijvoorbeeld doordat verschillende uitvoerders onafhankelijk van elkaar interventies inzetten).

Afbakening

We richten ons in dit handboek op het vaststellen van de effectiviteit van het activerend arbeidsmarktbeleid voor zover deze betrekking heeft op werklozen die ondersteuning nodig hebben bij het vinden van regulier werk. De facilitering van werklozen die deze hulp niet nodig hebben, valt buiten het bereik van het handboek. Denk hierbij bijvoorbeeld aan de operationele doelstelling voor CWI in de vorm van basisdienstverlening, via onder andere de website www.werk.nl.

Verder zijn voor uitvoeringsinstellingen als CWI en UWV diverse prestatie-indicatoren benoemd. Deze prestatie-indicatoren hebben betrekking op kenmerken van uitvoeringsprocessen, zoals tijdigheids- en volledigheidscijfers voor overdracht van dossiers en waarderingscijfers van werkgevers, werkzoekenden en intermediairs. Soms raken ze ook aan effectiviteit, bijvoorbeeld bij preventie- en uitstroomcijfers van het CWI. Deze indicatoren vallen echter buiten de horizon van dit handboek. Ze dienen immers om de uitvoeringsorganisaties te monitoren en zijn niet geschikt om te dienen als uitkomstmaten voor de besteding van middelen aan activerend arbeidsmarktbeleid.

3.4 Doelstellingen van het activerend arbeidsmarktbeleid

De doelstellingen van het activerend arbeidsmarktbeleid kunnen vanuit vier invalshoeken worden benaderd. Vandaaruit zijn er vijf hoofddoelstellingen te formuleren, welke we in paragraaf 3.5 (tabel 3.1) nader uitwerken:

1. De SZW-begroting 2008: dit betreft de artikelen 42 (arbeidsparticipatie) en 47 (re-integratie). De doelstelling van art. 42 is “zorgdragen voor een toename van de netto-arbeidsparticipatie”, van art. 47 “uitstroom naar regulier werk van die uitkeringsgerechtigden en werklozen die dat niet op eigen kracht kunnen” (SZW, 2007). Deze kunnen worden vertaald als verhogen arbeidsparticipatie (I) en terugdringen (langdurige) werkloosheid (II).
2. Beleidsmatige/politieke overwegingen. Het bevorderen van maatschappelijke participatie (III) en het creëren van gelijke kansen op de arbeidsmarkt (IV) worden gezien als belangrijke ankerpunten in het sociaal-economisch/maatschappelijk beleid.
3. (Macro-)economische overwegingen. Met het activerend arbeidsmarktbeleid wordt ook beoogd de werking van de arbeidsmarkt en het opleidingsniveau van de (beroeps)bevolking te verbeteren omdat dit bijdraagt aan het bevorderen van de arbeidsproductiviteit (V).
4. Financiële overwegingen. Het activerend arbeidsmarktbeleid kan ook tot doel hebben de kosten van sociale zekerheid terug te dringen. Het gaat dan voornamelijk om verlagen van uitkeringskosten, onder meer ook het I-deel van de WWB voor gemeenten. De achtergrond daarbij is veelal het terugdringen van (langdurige) werkloosheid (II). We zullen financiële uitkomstmaten dus daar meenemen.

Concreet kunnen de volgende hoofddoelstellingen van beleid geformuleerd worden:

- I. Verhogen arbeidsparticipatie
- II. Terugdringen (langdurige) werkloosheid
- III. Bevorderen maatschappelijke participatie
- IV. Creëren gelijke kansen op de arbeidsmarkt
- V. Bevorderen arbeidsproductiviteit

Deze vijf hoofddoelstellingen van het activerend arbeidsmarktbeleid zijn in kolom 1 van tabel 3.1 weergegeven. Per doelstelling kan een aantal subdoelstellingen onderscheiden worden. De subdoelstellingen zijn veelal de routes waarlangs de hoofddoelstellingen (kunnen) worden gerealiseerd en gaan soms al in de richting van uitkomstmaten.

Verschillende doelstellingen en uitkomstmaten: belang van indicatiestelling

De doelstelling van het beleid kan van groep tot groep en van individu tot individu verschillen. Sterker nog, deze doelstelling kan per individu over de tijd veranderen. Het vaststellen van de verkeerde doelstelling voor een bepaalde persoon kan leiden tot een suboptimale, dat wil zeggen minder kosteneffectieve inzet van middelen. Voor een doelmatige besteding zijn een juiste indicatiestelling en een juiste vaststelling van het (netto-)resultaat daarom van groot belang.

Scheidslijn tussen doelstelling en uitkomstmaat

De scheidslijn tussen doelstelling en uitkomstmaat is niet altijd even duidelijk te trekken. Er is sprake van een glijdende schaal. Dit is onder meer het gevolg van het steeds ‘smarter’ formuleren van doelstellingen. Dit wordt geïllustreerd door de algemene doelstelling van het re-integratie-artikel (eerst artikel 23, later artikel 47) in de begroting van SZW. Het accent is daarbij steeds meer op het resultaat in plaats van de inzet van re-integratie komen te liggen (zie box)¹².

Box beleidsdoelstellingen

Algemene doelstelling begroting 2002 (beleidsartikel 2)

“Het bevorderen van duurzame arbeidsparticipatie door: het realiseren van een sluitende keten van re-integratie voor nieuwe en langdurig werklozen; het wegwerken van kwalitatieve achterstanden van minder kansrijke werklozen en werkende werkzoekenden; het voor niet-actieven financieel aantrekkelijk maken zich op de arbeidsmarkt aan te bieden; het bevorderen van op minder kansrijke groepen gericht personeelsbeleid door werkgevers.” (SZW, 2001)

Algemene doelstelling begroting 2005 (beleidsartikel 23)

“Inschakelen in reguliere arbeid en voorkomen van uitval uit arbeid van hen die dat niet op eigen kracht kunnen.” (SZW, 2004)

Algemene doelstelling begroting 2008 (beleidsartikel 47)

“Uitstroom naar regulier werk van die uitkeringsgerechtigden en werklozen die dat niet op eigen kracht kunnen.” (SZW, 2007)

Van belang is steeds dat de beleidsmaker of onderzoeker duidelijk vaststelt wat de doelstelling (inclusief eventuele subdoelstellingen) is van het beleid en daar de juiste uitkomstmaten bij zoekt voor evaluatie.

3.5 Operationalisatie van doelstellingen in uitkomstmaten

In tabel 3.1 is een vertaling gemaakt van (sub)doelstellingen naar uitkomstmaten. Sommige uitkomstmaten komen bij meerdere doelstellingen voor. Bijvoorbeeld *Deelname aan reguliere arbeid* is een uitkomstmaat voor de doelstelling *Verhogen arbeidsparticipatie* en *Terugdringen werkloosheid*.

Hieronder bespreken we de diverse uitkomstmaten. Bij de diverse effectiviteitsanalyses worden ze verder geoperationaliseerd (zie hoofdstuk 6).

(1) (Duurzame) deelname aan reguliere arbeid

Verhogen van de arbeidsparticipatie kan geoperationaliseerd worden door de uitkomstmaat *Deelname aan reguliere arbeid*. Dit kan in de vorm van werknemerschap, inclusief tijdelijke en uitzendarbeid, of in de vorm van zelfstandig ondernemerschap. Over gesubsidieerde arbeid kan men twisten of dit een tussen- of einddoel is. Dit zal ook verschillen per individu.

Duurzaamheid kan verder geoperationaliseerd worden in de vorm van de minimale duur van een arbeidscontract bij plaatsing (verwachte duurzaamheid) of op basis van de gerealiseerde arbeidsduur (bijvoorbeeld 6 maanden, al dan niet bij eenzelfde werkgever). De eerstgenoemde heeft het voordeel van het eerder beschikbaar komen van de resultaten voor verdere analyses, de tweede is mogelijk beleidsmatig te prefereren.

De status ‘werkloos’ en ‘regulier werkzaam’ hoeven elkaar niet per definitie uit te sluiten. Een werkloze kan bijvoorbeeld gedeeltelijk uitstromen uit de uitkering. Hij neemt dan deel aan

¹² Zie Beleidsdoorlichting re-integratie (SZW, 2008).

regulier werk en behoudt toch nog recht op een (gedeeltelijke) uitkering. De oplossing voor dit probleem is het incorporeren ervan in de uitkomstmaat. Zo kunnen gemeenten of UWV bijvoorbeeld eisen stellen aan de re-integratie in de vorm van volledigheid door deze te stellen op ‘volledig uit de uitkering’ of ‘uit de uitkering voor een minimaal percentage van de hoogte van de uitkering’.

(2) *Vervulling van vacatures*

Naast de aanbodkant van arbeid kan men bij het de doelstelling *Verhogen van arbeidsparticipatie* ook kijken naar de vraagkant. Het kan hier gaan om het aantal openstaande vacatures, het aandeel langdurige openstaande vacatures, de duur dat vacatures openstaan of de kans op vervulling van vacatures.

Tabel 3.1: Overzicht van (sub)doelstellingen met bijbehorende uitkomstmaten

Beleidsdoelstellingen en subdoelen	Uitkomstmaat
(I) VERHOGEN ARBEIDSPARTICIPATIE	(1) (Duurzame) deelname reguliere arbeid (2) Vervulling van vacatures
(I.a) Verkleinen afstand tot regulier werk	(3) Stapje op re-integratieladder
(I.b) Vergroten arbeidsaanbod	(4) Arbeidsaanbod in personen of fte
(I.c) Vergroten kwaliteit arbeidsaanbod	(5) Arbeidsaanbod naar competenties
(II) TERUGDRINGEN (LANGDURIGE) WERKLOOSHEID	(6) Werkloosheidsduur (1) (Duurzame) deelname reguliere arbeid (7) Baanduur (8) (Hernieuwde) instroom in SUWI-keten
(II.a) Verkleinen afstand tot regulier werk	(3) Stapje op re-integratieladder
(II.b) Vergroten kwaliteit arbeidsaanbod	(5) Arbeidsaanbod naar competenties
(II.c) Terugdringen kosten sociale zekerheid	(9) Uitkeringsduur (10) Uitkeringslasten
(III) BEVORDEREN MAATSCHAPPELIJKE PARTICIPATIE	(11) Deelname maatschappelijke activiteiten
(IV) CREËREN GELIJKE KANSEN OP ARBEIDSMARKT	Verschillen tussen groepen in: <ul style="list-style-type: none"> • (1) Arbeidsparticipatie • (12) Kwaliteit van de baan • (13) Beloning
(V) BEVORDEREN ARBEIDSPRODUCTIVITEIT	(14) Arbeidsproductiviteit
(V.a) Vergroten kwaliteit arbeidsaanbod	(5) Arbeidsaanbod naar competenties
(V.b) Verbeteren allocatie factor arbeid	(15) Kwaliteit van de match (14) Arbeidsproductiviteit (7) Baanduur

(3) Stapje op de re-integratieladder

De afstand tot de arbeidsmarkt is niet voor iedere werkloze in één traject te overbruggen. De weg naar werk wordt als het ware opgedeeld in stukjes. In dit kader wordt vaak gesproken over 'een stapje op de re-integratieladder zetten'. Door re-integratie-activiteiten doen mensen ervaring op en kunnen zij zich ontwikkelen, waardoor zij dichterbij de arbeidsmarkt komen te staan en er op termijn mogelijkheden worden benut om betaald werk te verrichten. Het opdelen van het re-integratietraject in stapjes is ook voor gemeenten vaak de praktijk omdat het de voordelen van maatwerk biedt: voor elk stapje kan de beste aanbieder gekozen worden, of gemeenten kunnen de uitvoering op onderdelen in eigen hand houden.

Het vaststellen van een uitkomstmaat die hier goed rekening mee houdt, is in de praktijk niet altijd helder. Dit komt tot uitdrukking in het feit dat er diverse re-integratieladders gehanteerd worden. Zie figuur 3.1 voor een 'grootste gemene deler' van re-integratieladders.

Naast het vaststellen van de classificatie van de 'sporten' op de ladder, is ook de vaststelling van de situatie waarin een werkloze zich bevindt soms moeilijk, of alleen na zeer intensieve indicatiestellingen. Bovendien doet de typering niet altijd recht aan complexe praktijksituaties, waar vaak sprake is van meerdere belemmeringen om aan het werk te gaan, zoals gezondheid, schulden, sociale vaardigheden, taal, etc.

Tot slot worden ladders in de praktijk ook nogal eens gehanteerd vanuit de inhoud van het aanbod van het traject (zorg, scholing, bemiddeling).

(4) Arbeidsaanbod in personen of fte

De subdoelstelling van beleid ter bevordering van de arbeidsparticipatie hoeft niet direct gericht te zijn op re-integratie van personen die al naar werk zoeken. Men kan zich ook richten op het vergroten van het arbeidsaanbod, zodat op de arbeidsmarkt het aantal potentiële matches van werkgever-werknemer vergroot wordt. Beleid kan gericht zijn op het aanboren van nieuw aanbod, zoals de stille arbeidsreserve. Daarnaast kan het ook gaan om uitbreiding van het aantal aangeboden arbeidsuren van degenen die al werkzaam zijn.

(5) Arbeidsaanbod naar competenties

Naast het kwantitatieve arbeidsaanbod kan ook het kwalitatieve arbeidsaanbod een uitkomstmaat vormen. Er wordt dan niet alleen naar aantallen maar ook naar competenties gekeken. Veelal worden competenties gemeten door opleidingsniveau en eventueel opleidingsrichting. Er zijn ook andere operationalisaties van competenties, zoals behaalde certificaten, uitgeoefende functies en registratiesystemen als EVC (Erkenning elders Verworven Competenties).

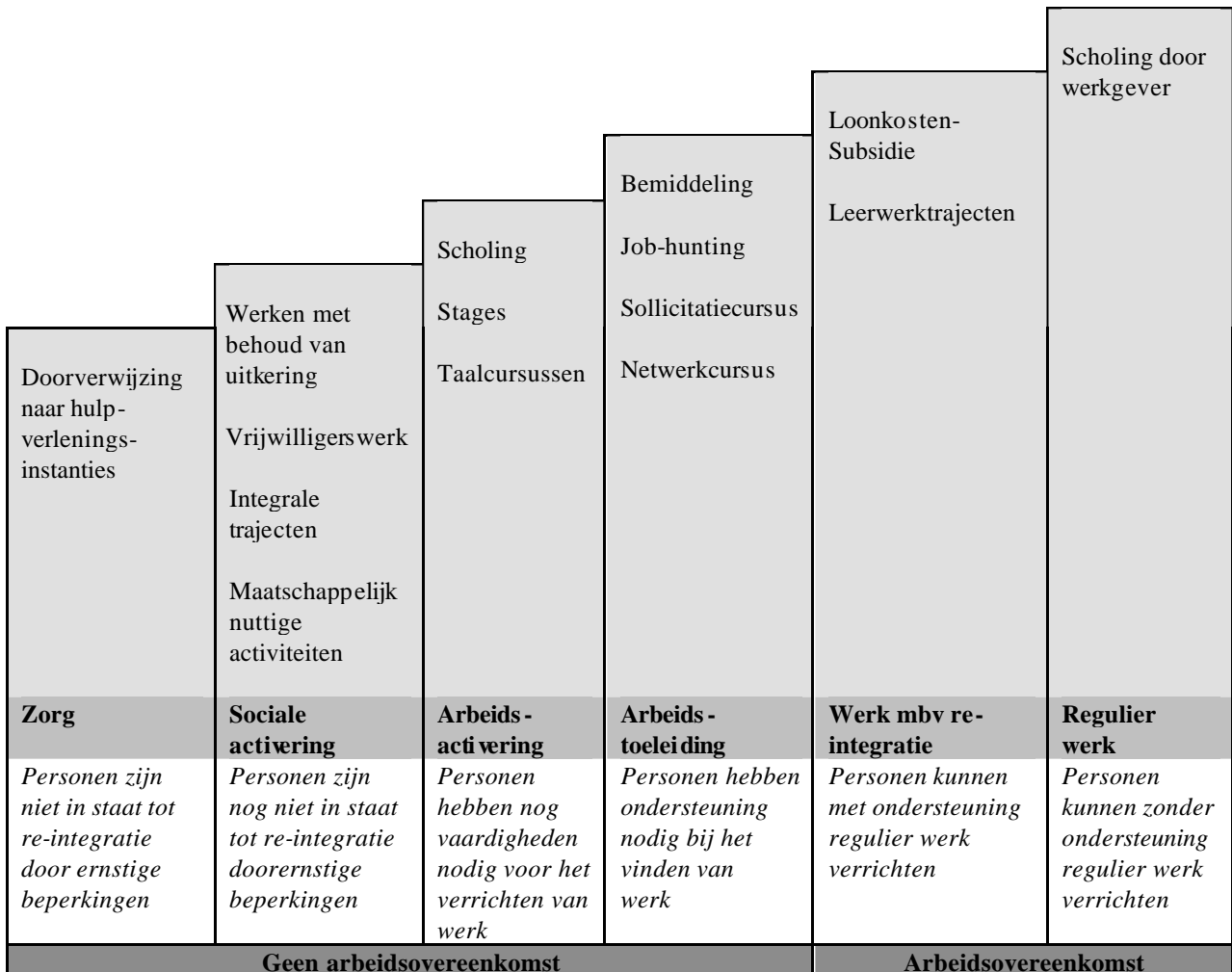
(6) Werkloosheidsduur

Werkloosheidsduur is in principe het communicerende vat van de uitkomstmaat *Deelname aan reguliere arbeid*. Expliciet wordt hier de duur van werkloosheid in de uitkomstmaat opgenomen. Een snellere uitstroom naar werk wordt geprefereerd boven een minder snelle. Zie verder onder *Deelname aan reguliere arbeid* voor het begrip duurzaamheid en de samenloop tussen de status 'werkzaam' en 'werkloos'. Voor beëindiging van werkloosheid kunnen soortgelijke operationalisaties gekozen worden door bijvoorbeeld de werkloosheidsduur pas als beëindigd te beschouwen nadat de persoon die regulier werk gevonden heeft, een bepaalde tijd geparticipeerd heeft op de arbeidsmarkt.

(7) Baanduur

Voor baanduur kan ook duur van de ononderbroken werkzame periode worden gehanteerd. Ook hier ligt het accent op duurzame uitstroom naar werk.

Figuur 3.1: Stapjes op de re-integratieladder



Bron: Beleidsdoorlichting re-integratie (SZW, 2008)

(8) *(Hernieuwde) instroom in SUWI-keten*

Veelal wordt uitstroom uit de uitkering als een succes beschouwd. Relevant is dan dat de persoon geen beroep meer doet op de uitkeringsregeling. Op macro-economisch niveau kan het ook relevant zijn de SUWI-keten als geheel te beschouwen en de (her)instroom te bekijken vanuit ketenperspectief in plaats van uit de afzonderlijke uitkeringsregeling.

(9) *Uitkeringsduur*

Synoniem aan werkloosheidsduur met die kanttekening dat de uitkering ook om andere redenen kan stoppen dan door uitstroom naar werk. Zo is de duur van de WW-uitkering beperkt en kan de bijstandsuitkering stopgezet worden door inkomen van een partner of vermogen.

(10) *Uitkeringslasten*

Naast uitkeringsduur kunnen ook uitkeringslasten een relevante uitkomstmaat zijn. Verschillen met de uitkomstmaat *Uitkeringsduur* worden veroorzaakt door verschillen in hoogte van de uitkering tussen individuen. Het wordt dan aantrekkelijker om te zien of het beleid niet beter gericht kan worden op de groep waar de grootste uitkeringslast bespaard kan worden.

(11) Maatschappelijke participatie in plaats van regulier betaald werk

Voor sommige groepen uitkeringsgerechtigden is deelname aan regulier werk niet weggelegd. Dit is bijvoorbeeld het geval voor mensen met een ernstige fysieke beperking. Het overheidsbeleid is erop gericht om deze groep zo goed mogelijk uit te filteren en geen onderdeel meer te laten vormen van het re-integratiebeleid in enge zin. Denk bijvoorbeeld aan de IVA binnen de WIA. Beleidsmatig kan er echter toch aanleiding zijn om deze groep, hoewel regulier werk nooit gehaald wordt, te laten participeren in de maatschappij.

Uitkomstmaten kunnen wederom geformuleerd worden in termen van maatschappelijke participatie (deelname aan de samenleving, gesubsidieerde arbeid of vrijwilligerswerk). Achterliggende doelstellingen liggen vaak op het terrein van: verminderen maatschappelijke overlast/criminaliteit, verminderen zorgconsumptie, bevorderen zelfwaardering/zelfrespect van het individu (Brouwer, Zwinkels & Van Genabeek, 2006; Kok, Hollanders & Hop, 2006). Uitkomstmaten hebben dan ook betrekking op (één of meerdere van) deze onderdelen.

(12) Kwaliteit van de baan

Bij *Kwaliteit van de baan* gaat het o.a. om arbeidsomstandigheden, contractvorm, aantal uren en werktijden. Ook kan bijvoorbeeld medewerkerstevredenheid als indicator dienen.

(13) Beloning

Bij *Beloning* kan het naast primaire ook om secundaire arbeidsvoorwaarden gaan.

In de vergelijking van verschillen tussen groepen (bijvoorbeeld mannen-vrouwen, ouderen-jongeren, gehandicapten-niet-gehandicapten) dient overigens gecorrigeerd te worden voor 'overige' kenmerken die van invloed zijn op de productiviteit.

(14) Arbeidsproductiviteit

Arbeidsproductiviteit wordt vaak uitgedrukt als toegevoegde waarde per werknemer, maar soms ook gemeten via het loon.

(15) Kwaliteit van de match

Kwaliteit van de match is de mate waarin arbeidsvraag en arbeidsaanbod op elkaar aansluiten, bijvoorbeeld in opleidingsniveau, opleidingsrichting, aantal arbeidsuren etc.

3.6 Macro-economische (neven)effecten

De uitkomstmaten die in dit hoofdstuk beschreven zijn hebben betrekking op voorgenomen beleidseffecten, dat wil zeggen dat het de bedoeling en verwachting is dat beleid leidt tot verandering in de beschreven uitkomstmaten. Daarnaast kunnen er bedoeld of onbedoeld neveneffecten optreden. Deze kunnen verschillend van aard zijn en diverse vormen aannemen:

- **Macro-effecten**

Macro-effecten zijn indirecte effecten die ontstaan door interventies op microniveau die macro-economische consequenties hebben. We noemen twee voorbeelden:

- **Substitutie/verdringing:** de uitkeringsgerechtigde wordt aan werk geholpen maar dit leidt ertoe dat een (andere) werknemer zijn baan verliest of dat andere werkzoekenden geen baan kunnen vinden.
- **Effecten op loonvorming:** re-integratie-interventies kunnen tot vergroting van de participatie leiden en daarnaast ook tot vergroting van het effectief arbeidsaanbod waardoor de loonvoet daalt.

Vanwege dit type effecten is de effectiviteit van beleid op macro-niveau vaak niet hetzelfde als op micro-niveau. Beleid dat op micro-niveau effectief blijkt te zijn, hoeft dat op macro-niveau niet te zijn.

- **Spill-over effecten**
Spill-over effecten zijn effecten op andere markten dan de arbeidsmarkt. Zo zou het kunnen zijn dat de inzet van gesubsidieerde arbeid leidt tot concurrentievervalsing op de goederenmarkt.
- **Herverdelingseffecten**
Naast effecten op de genoemde doelstellingen en uitkomstmaten kan er ook sprake zijn van herverdelingseffecten. Zo kan macro-economisch de totale participatie niet toe- of afnemen door een beleidsinterventie maar kan het het geval zijn dat steeds andere mensen aan het werk komen en werkloos zijn en/of dat de inkomensverdeling wijzigt. Beleidsmatig kan het relevant zijn dat niet een min of meer vaste groep getroffen blijft door werkloosheid maar dat deze ‘rouleert’ onder de beroepsbevolking.

Het vaststellen van macro-effecten gebeurt in Nederland vaak door het Centraal Planbureau (CPB). Dit handboek richt zich op het vaststellen van de effectiviteit op micro-niveau. Voor het vaststellen van macro-effecten zijn macro-economische modellen noodzakelijk. Bespreking van de methoden voor het vaststellen van deze neveneffecten valt buiten het kader van dit handboek.

Voor de neveneffecten geldt daarnaast ook dat het aan de beleidsmaker is om ze al dan niet mee te nemen in de beleidskeuze. Voor een gemeente kan bijvoorbeeld concurrentievervalsing door subsidies op arbeid leiden tot onvrede onder bestaande ondernemers die zonder subsidies werken. Hiervoor kunnen ook uitkomstmaten gedefinieerd worden. We gaan daar in dit handboek niet verder op in.

3.7 Methoden, specificaties, bruikbare gegevens en rekenregels

In de volgende hoofdstukken staat voor met name de onderzoekers beschreven hoe netto-effecten van maatregelen van activerend arbeidsmarktbeleid worden gemeten door een vergelijking te maken tussen de uitkomstmaat mét en zónder de inzet van een beleidsinterventie. Daarbij wordt ingegaan op de belangrijkste evaluatiemethoden (hoofdstuk 5), de mogelijkheden van empirische specificaties (hoofdstuk 6) en de bruikbare gegevens, de rekenregels en de presentatie van gegevens (hoofdstuk 7).

Voor met name beleidsmakers die hun beleid met deze methoden willen (laten) evalueren is een vereenvoudigde samenvatting gemaakt (hoofdstuk 4).

4 Het meten van netto-effectiviteit

Effecten van maatregelen van activerend arbeidsmarktbeleid kunnen worden gemeten door een vergelijking te maken tussen de uitkomstmaat mét de inzet van een beleidsinterventie en de uitkomstmaat zónder de inzet van een beleidsinterventie. Voor een goede meting van netto-effecten dient die vergelijking zo zuiver mogelijk te zijn. In de praktijk hebben we vaak te maken met situaties die niet onmiddellijk vergelijkbaar zijn. Hoofdstuk 5 gaat in op beschikbare evaluatiemethoden die kunnen zorgen voor een zuivere vergelijking van de situatie mét en de situatie zónder beleidsinterventie. Hoofdstuk 6 beschrijft enkele veel voorkomende uitwerkingen van die evaluatiemethoden, afhankelijk van de doelpopulatie, de uitkomstmaat en de beschikbare gegevens voor de effectiviteitsanalyse. Hoofdstuk 7 gaat ten slotte in op de gegevens die nodig zijn voor het kunnen uitvoeren van effectiviteitsanalyses, de manier waarop die gegevens dienen te worden bewerkt en de manier waarop onderzoeksresultaten kunnen worden gepresenteerd. Omdat deze hoofdstukken vrij technisch van aard zijn, zijn ze vooral bedoeld voor mensen die effectiviteitsanalyses voorbereiden of uitvoeren. In dit hoofdstuk worden de belangrijkste onderdelen daarom samengevat.

Counterfactual

Als men het daadwerkelijke netto-effect van maatregelen van activerend arbeidsmarktbeleid op individueel niveau zou willen meten, dan zou dezelfde persoon tegelijkertijd moeten worden vergeleken in een situatie waarin hij of zij wél met deze maatregelen heeft te maken en een situatie waarin hij of zij daar niet mee te maken heeft. Werkzoekenden nemen echter wel of niet deel aan een re-integratietraject maar nooit allebei tegelijk. De niet waargenomen situatie blijft per definitie onbekend: dit staat bekend als de *counterfactual*. Maar dat betekent niet dat het effect van een beleidsinterventie niet zou kunnen worden vastgesteld.

Vergelijking tussen proef- en controlegroep

Effecten van beleidsinterventies kunnen worden gemeten door de groep die te maken heeft met een beleidsinterventie (de proefgroep) te vergelijken met de groep die daar niet mee te maken heeft (de controlegroep). Wanneer beide groepen goed vergelijkbaar zijn, dan kunnen verschillen in de uitkomstmaat (bijvoorbeeld de kans op werkhervatting) tussen beide groepen volledig worden toegeschreven aan de werking van de beleidsinterventie. Dat verschil is de netto-effectiviteit of toegevoegde waarde van de interventie.

Samenstelling- en selectie-effecten

In de praktijk zijn de proefgroep en controlegroep niet altijd goed vergelijkbaar. De groepen kunnen een verschillende samenstelling hebben naar bijvoorbeeld leeftijd en opleidingsniveau, factoren die mede de arbeidsmarktpositie bepalen (samenstellingseffecten). Daarnaast hebben niet alle werklozen dezelfde kans om een re-integratietraject te krijgen en daarmee in de proefgroep terecht te komen: er is sprake van een selectieve inzet van beleidsinstrumenten (selectie-effecten). Waargenomen verschillen in de werkloosheidsduur tussen proefgroep en controlegroep kunnen daarom niet zomaar worden toegeschreven aan de werking van de beleidsinterventie. Afhankelijk van de uitkomstmaat, het type beleidsinterventie en de beschikbare gegevens voor analyse, kan door middel van specifieke evaluatiemethoden wel worden gecorrigeerd voor samenstelling- en selectie-effecten.

Welke evaluatiemethode is het meest geschikt?

Welke evaluatiemethode in welk geval het meest geschikt is hangt af van veel verschillende factoren. De keuze van een evaluatiemethode en vooral de toegepaste invulling ervan is een min of meer subjectieve keuze van degene die de effectiviteitsanalyse uitvoert. Op voorhand is niet voor elke situatie te zeggen wat de beste evaluatiemethode is. Een empirische specificatie dient

uiteindelijk empirisch te worden bepaald én getoetst. Toch kunnen op grond van de bestaande literatuur in algemene termen wel handreikingen worden gedaan over welke evaluatiemethoden in welke situaties goed presteren. Dat is ten slotte ook het doel van dit handboek.

Grofweg hangt de keuze van de optimale evaluatiemethode af van de volgende factoren:

- het moment van evalueren: gedurende het beleidsproces of via een ex-post analyse;
- de beschikbare gegevens: op meerdere momenten of niet;
- het type beleidsinterventie: inzet varieert in de tijd of alleen tussen groepen;
- vergelijkbaarheid van en mate van overlap tussen proef- en controlegroep, en
- de uitkomstmaat: duurzaam, kansmaat of anders.

Belangrijk is dat de inzet van een beleidsinterventie varieert ten opzichte van de te meten uitkomstmaat. Die variatie kan zich voordoen tussen groepen binnen de doelpopulatie, maar ook in de tijd. Variatie zorgt voor verschillen en verschillen zijn nodig om effecten te kunnen identificeren. Als iedereen binnen de doelpopulatie op hetzelfde moment te maken krijgt met een beleidsinterventie, dan is het effect ervan niet te meten.

Om enige richting te geven in het antwoord op de vraag welke evaluatiemethode in welke situatie het meest geschikt is, kan een aantal aanvullende vragen worden gesteld. Via het antwoord op die vragen kan een indicatie worden gegeven welke evaluatiemethode voor de specifieke beleidspraktijk geschikt zou zijn voor het meten van de netto-effectiviteit. Die aanvullende vragen en antwoorden worden samengevat in het schema van Figuur 4.1.

Ten eerste is het de vraag of de effectiviteitsanalyse gedurende de beleidsinterventie kan plaatsvinden of pas achteraf. Als er ruimte is om vooraf de evaluatie te organiseren kan het beste gebruik worden gemaakt van een *experiment*. In een experiment worden de proef- en controlegroep bij voorbaat vergelijkbaar gemaakt door de beleidsinterventie willekeurig in de populatie in te zetten. Bijvoorbeeld door slechts een willekeurig deel van de populatie werklozen actief te bemiddelen naar werk. De andere groep ontvangt dan geen bemiddeling (voor de berekening van absolute netto-effecten) of een alternatieve beleidsinterventie (voor de berekening van relatieve netto-effecten). In de praktijk stuit het willekeurig inzetten van beleidsinterventies vaak op financiële, juridische en ethische bezwaren. De meeste evaluatiemethoden bestaan daarom uit ex-post evaluaties, waarbij de mogelijk selectieve inzet van de beleidsinterventie al heeft plaatsgevonden.

In zeer uitzonderlijke gevallen is achteraf sprake van een spontane experimentele situatie die toevallig is ontstaan en niet van te voren is opgezet. Zo'n situatie wordt aangeduid met een *natuurlijk of quasi experiment*. De doelpopulatie is in dat geval bij toeval willekeurig toegedeeld aan de proef- en controlegroep, denk bijvoorbeeld aan willekeurige administratieve fouten, waardoor een deel van de doelpopulatie niet of pas later te maken krijgt met de beleidsinterventie. Ook kan het voorkomen dat sterk gelijkende groepen wel en niet te maken krijgen met een beleidsinterventie, bijvoorbeeld als gevolg van een leeftijdscriterium in regelgeving. In dat geval kan gebruik worden gemaakt van een *discontinuïteiten benadering*, waarbij een proef- en controlegroep worden samengesteld uit personen die net wel, respectievelijk net niet aan het leeftijdscriterium voldoen. Door vergelijking van deze sterk gelijkende (deel)groepen in de onderzoekspopulatie kan het netto-effect van de beleidsinterventie worden geïdentificeerd. Bestaan dit soort situaties niet, dan zal een ex-post evaluatie op één of andere manier rekening moeten houden met samenstelling- en selectie-effecten voor een goede vergelijking van de uitkomstmaat tussen proef- en controlegroep.

De beste ex-post methode om de proef- en controlegroep vergelijkbaar te maken, is de zogenaamde *matching* methode. Met de matching methode wordt een controlegroep samengesteld die qua kenmerken zoveel mogelijk lijkt op de proefgroep. Een vergelijking tussen de uitkomstmaat tussen beide groepen geeft vervolgens het netto-effect van de inzet van de

beleidsinterventie. Omdat de samenstelling van de controlegroep wordt gebaseerd op waargenomen kenmerken van personen en de situatie waarin zij verkeren, leent deze methode zich bij uitstek voor een netto-effectmeting als er gegevens over veel verschillende kenmerken beschikbaar zijn, die bovendien zowel binnen als buiten de proefgroep voorkomen. In dat geval is de kans groot dat een soortgelijke groep als de proefgroep kan worden samengesteld, waardoor verschillen in samenstelling en selectie worden geëlimineerd.

Er is echter een aantal situaties waarin de matching methode (alleen) geen goede oplossing biedt voor het verschil tussen de proef- en controlegroep. Het eerste voorbeeld daarvan is de situatie waarin de beleidsinterventie variabel in de tijd wordt ingezet, zoals het geval is bij de meeste re-integratietrajecten. Niet iedereen krijgt in de praktijk op hetzelfde moment een re-integratietraject aangeboden, waardoor de proef- en controlegroep niet alleen qua kenmerken, maar ook in de tijd vergelijkbaar moet worden gemaakt. Dit is met name belangrijk wanneer de uitkomstmaat zelf afhankelijk is van de tijd, zoals de werkloosheidsduur, baanduur en zoekduur. De matching methode is daar minder geschikt voor dan bijvoorbeeld een *duurmodel*. Een duurmodel berekent de kans op beëindiging van bijvoorbeeld de werkloosheidsduur, afhankelijk van allerlei invloedsfactoren, waaronder het tijdstip van inzet van de beleidsinterventie. De uitkomstmaat wordt dus afhankelijk van de verstreken duur mét en zonder beleidsinterventie gemodelleerd. Overigens zijn veel kansmaten ook afhankelijk van de (verstreken) tijd, zoals de kans op werkhervatting, de kans op een stapje op de re-integratieladder en de kans op een hernieuwde instroom in de SUWI-keten. Wanneer de inzet van de beleidsinterventie variabel in de tijd plaatsvindt en de uitkomstmaat duurzaam afhankelijk is, dan kan het beste gebruik worden gemaakt van een duurmodel. Voorwaarde is wel dat de beschikbare gegevens het mogelijk maken om variatie in de tijd op individueel niveau te meten (longitudinale gegevens).

Overigens sluiten de matching methode en een duurmodel elkaar niet uit. Zo kan matching worden gebruikt voor het zo goed mogelijk vergelijkbaar maken van de proef- en controlegroep op basis van kenmerken op één moment in de tijd. Vervolgens kan een duurmodel worden gebruikt om de tijdsafhankelijke ontwikkeling tussen de proef- en controlegroep te vergelijken. Deze combinatie van evaluatiemethoden wordt sterk aanbevolen.

Een andere situatie waarin de matching methode geen goede oplossing biedt voor het verschil tussen proef- en controlegroep, is die waarin er relatief weinig overlap is tussen beide groepen. Dat is onder meer het geval wanneer de inzet van een te evalueren beleidsinterventie sterk selectief is. Als voorbeeld kan de inzet van uitstroompremies worden genoemd wanneer deze uitsluitend én talrijk worden ingezet bij ouderen. De samenstelling van de proef- en controlegroep zijn op basis van matching dan niet goed in overeenstemming te krijgen, omdat er een te groot verschil in leeftijd bestaat tussen proef- en controlegroep. Leeftijd is nu eenmaal een belangrijke verklarende factor voor de meeste uitkomstmaten. Matching gaat in dat geval ten koste van een voldoende aantal waarnemingen voor de effectiviteitsanalyse, omdat de proefgroep alleen gematched kan worden met de zeer kleine groep ouderen die géén uitstroompremie hebben ontvangen. In zo'n geval kan de *difference-in-difference* methode uitkomst bieden. Bij de diff-in-diff methode wordt het verschil in uitkomstmaat voor en na de interventie gemeten voor zowel de proefgroep als de controlegroep. Het 'verschil van het verschil' kan vervolgens – na correctie voor samenstelling- en selectie-effecten – worden toegeschreven aan de inzet van uitstroompremies. Als matching het verschil in samenstelling niet kan elimineren, dan biedt de difference-in-difference methode een manier om met dat verschil om te gaan.

Figuur 4.1: Beslisboom voor keuze van de meest geschikte evaluatimethode



Alle ex-post evaluatiemethoden tot dusver gaan uit van de beschikbaarheid van gegevens op meerdere meetmomenten in de tijd, namelijk voorafgaand en volgend op de beleidsinterventie.¹³ Er zijn situaties waarin achteraf niet meer kan worden vastgesteld wat de uitkomstmaat was voorafgaand aan de beleidsinterventie. Er is dan geen informatie beschikbaar over de variatie in de uitkomstmaat in de tijd, zodat evaluatiemethoden die daarvan gebruik maken, zoals een duuranalyse of een difference-in-difference methode, niet toepasbaar zijn. De matching methode blijft in dat geval wel toepasbaar, omdat het een controlegroep samenstelt die sterke gelijkenissen vertoont met de proefgroep. Die sterke gelijkenis maakt het waarschijnlijk dat ook de uitkomstmaat ten tijde van de ontbrekende voormeting vergelijkbaar is. Dit gaat echter alleen op zolang er gegevens over voldoende invloedsfactoren beschikbaar zijn om de controlegroep echt een vergelijkbare samenstelling te geven als de proefgroep. Wanneer het aantal waargenomen invloedsfactoren beperkt is en de overlap in kenmerken gering, zal ook de vergelijking tussen de proefgroep en de geconstrueerde controlegroep niet zuiver zijn. In dat geval kan niet langer gebruik worden gemaakt van een econometrisch model voor de effectiviteitsanalyse, maar zal gebruik moeten worden gemaakt van andere evaluatiemethoden.

Benodigde gegevens voor de effectiviteitsanalyse

Om de verschillende evaluatiemethoden toe te kunnen passen, zijn gegevens nodig over de uitkomstmaat, de beleidsinterventie en overige factoren, voorzover die verschillen tussen proef- en controlegroep en tevens de uitkomstmaat kunnen beïnvloeden. Tabel 4.1 geeft een overzicht van uitkomstmaten die volgen uit de doelstellingen van het activerend arbeidsmarktbeleid zoals beschreven in Hoofdstuk 3. Voor elke uitkomstmaat geeft de tabel globaal aan welke gegevens nodig zijn om het niveau ervan te kunnen meten. Meer specifiek kan de start van deelname aan reguliere arbeid vanuit een uitkering worden gedefinieerd bij een gelijktijdige beëindiging van de uitkering of al bij een verlaging van de uitkering. Andersom kan de uitkomstmaat ook worden gedefinieerd afhankelijk van de beschikbare gegevens. Zo kan werkhervatting worden gedefinieerd als uitstroom uit de uitkering als gevolg van het aanvaarden van werk, als het moment waarop een baan wordt geaccepteerd bij een werkgever of het moment waarop het arbeidscontract wordt ondertekend. Het is voor een effectiviteitsanalyse van groot belang altijd zeer helder te zijn over de definitie van de uitkomstmaat, zeker wanneer deze een tijdsdimensie heeft, zoals de werkloosheidsduur, baanduur en uitkeringsduur.

Soms zijn meerdere gegevensbronnen nodig om de uitkomstmaat vast te stellen. Wanneer in plaats van de werkloosheidsduur of deelname aan reguliere arbeid de baanzoekduur tot aan werkhervatting de relevante uitkomstmaat is, dan is het van belang om zowel het beginpunt van de baanzoekduur als het eindpunt van de baanzoekduur zorgvuldig te formuleren. Arbeidsgehandicapten zullen bijvoorbeeld pas naar werk gaan zoeken als duidelijk is dat ze (een deel van) het recht op de arbeidsongeschiktheidsuitkering gaan verliezen en dus (deels) werkloos worden. Start van de baanzoekduur is dan bijvoorbeeld de start van een WW-uitkering voor deze groep. Het vinden van een baan kan dan als einde van de baanzoekduur worden beschouwd, of pas wanneer deze baan de volledige verdien capaciteit gebruikt, dan wel duurzaam is (langer duurt dan bijvoorbeeld een half jaar). De precieze definitie zal behalve van de beschikbaarheid van gegevens ook afhangen van beleidsdoelstellingen.

¹³ Relevante gegevens over de situatie voorafgaand aan de inzet van een beleidsinterventie kunnen in veel gevallen ook achteraf worden vastgesteld, bijvoorbeeld op basis van administratieve gegevens. Pas wanneer deze gegevens niet meer zijn te achterhalen, spreken we van een niet beschikbare voormeting.

Tabel 4.1: Globaal overzicht van benodigde gegevens voor het vaststellen van uitkomstmaten

Uitkomstmaat	Benodigde gegevens
(1) (Duurzame) deelname reguliere arbeid	Datum of indicator start arbeidsdeelname; Duurzame arbeidsdeelname vaak gedefinieerd als arbeidsdeelname van minimaal een half jaar
(2) Vervulling van vacatures	Datum of indicator vervulde vacature(s) of aantal vervulde vacatures, eventueel als aandeel van aantal openstaande vacatures
(3) Stapje op re-integratieladder	Trede op de re-integratieladder, indicatiestelling o.i.d.
(4) Arbeidsaanbod in personen of fte	Aantal werkzoekenden of indicatie arbeidsparticipatie
(5) Arbeidsaanbod naar competenties	Aantal werkzoekenden of indicatie arbeidsparticipatie per competentieniveau
(6) Werkloosheidsduur	Datum start en datum einde werkloosheid; Als alternatief geldt de baanzoekduur, waarvan de start kan afhangen van uitkeringsafhankelijkheid (bijvoorbeeld start WW bij arbeidsgehandicapten) en het einde van de start arbeidsdeelname
(7) Baanduur	Datum start en datum einde van baan
(8) (Hernieuwde) instroom in SUWI-keten	Datum of indicator van instroom in SUWI keten
(9) Uitkeringsduur	Datum start en datum einde uitkering; Datum einde uitkering eventueel vervangen door datum start baan
(10) Uitkeringslasten	Zie uitkeringsduur, plus uitkeringsniveau
(11) Deelname maatschappelijke activiteiten	Datum of indicator start relevante activiteiten
(12) Kwaliteit van de baan	Kenmerken van de baan, zoals beroepsniveau, soort contract, aantal uren, werktijden, etc.
(13) Beloning	Bruto of netto loon per maand of per uur
(14) Arbeidsproductiviteit	Toegevoegde waarde per werknemer of loonkosten
(15) Kwaliteit van de match	Kenmerken van de baan in combinatie met kenmerken van de werknemer, zoals gevraagde en aangeboden opleidingsniveau, gevraagde en aangeboden aantal arbeidsuren, etc.

Behalve over de uitkomstmaat is ook informatie nodig over de inzet van een beleidsinterventie waarvan het effect op de uitkomstmaat dient te worden geëvalueerd. In Hoofdstuk 3 is een overzicht gegeven van de verschillende instrumenten waarop dit handboek zich richt. Tabel 4.2 geeft per instrument een overzicht van de informatie die nodig is om vast te kunnen stellen of deze beleidsinterventies zijn ingezet. Bij voorkeur dient het moment waarop de interventie plaatsvindt zo exact mogelijk bekend te zijn. Dit is met name bij een effectiviteitsanalyse met duurmodellen van belang.

Tabel 4.2: Overzicht van benodigde gegevens voor het vaststellen van de inzet van verschillende soorten beleidsinstrumenten

Beleidsinstrument	Benodigde gegevens
Vaardigheidstraining voor het zoekproces	Datum (of indicator) start vaardigheidstraining, eventueel datum afronding vaardigheidstraining
Scholing	Datum (of indicator) start scholing, eventueel datum afronding scholing of behaalde diploma's
Financiële prikkels (zoals sancties, uitstroompremies en loonkostensubsidies)	Datum start én datum einde (of indicator) van financiële prikkel, eventueel met omvang van financiële prikkel
Directe werkgelegenheidscreatie	Datum (of indicator) start werkgelegenheid, eventueel met omvang van het aantal plaatsen
Subsidies voor starten eigen bedrijf	Datum (of indicator) van inzet subsidie of start eigen bedrijf, eventueel met omvang subsidie
Beschikbaar stellen van arbeidsmarktinformatie	Datum (of indicator) start nieuwe dienstverlening
Actieve bemiddeling	Datum (of indicator) start actieve bemiddeling
(Intensieve) begeleiding	Datum (of indicator) start intensieve begeleiding, eventueel met datums (of indicatoren) start individuele componenten

Factoren die doorgaans van invloed zijn op de uitkomstmaten in Tabel 4.1 kunnen worden opgedeeld in persoonskenmerken, kenmerken van de baan, kenmerken van de uitkering en omgevingskenmerken. Tabel 4.3 geeft voor elk van deze groepen een niet uitputtende lijst met factoren die relevant en in meer of minder mate beschikbaar zijn voor een netto-effectiviteitsanalyse van maatregelen van activerend arbeidsmarktbeleid. In zo'n analyse kunnen kenmerken van de huidige baan of uitkering van belang zijn, maar ook die van de laatste baan of uitkering. Voor een analyse van de baanduur zijn bijvoorbeeld vooral kenmerken van de huidige baan van belang, bij een analyse van de werkloosheidsduur vooral kenmerken van de laatste baan.

Er zijn diverse bronnen beschikbaar waaruit de benodigde gegevens voor een effectiviteitsanalyse kunnen worden betrokken. Allereerst kan onderscheid worden gemaakt tussen administratieve bestanden en enquêtegegevens. Het belangrijkste voordeel van enquêtegegevens ten opzichte van administratieve bestanden is dat ze vaak specifiek gericht zijn op het onderwerp van onderzoek. Daardoor kunnen zeer veel relevante gegevens voor een effectevaluatie worden verzameld, waaronder zachte factoren als de motivatie van de werkzoekende om een baan te vinden. Belangrijkste nadeel is echter dat enquêtegegevens ten opzichte van administratieve bestanden vaak een zeer beperkt deel van de onderzoekspopulatie betreffen, waarbij de respons niet altijd representatief is voor de gehele onderzoekspopulatie. Daardoor is de betrouwbaarheid en detaillering van de onderzoeksresultaten vaak beperkt. Ook bevatten enquêtes meer dan administratieve bestanden opinies en percepties van individuen, waardoor de betrouwbaarheid en onderlinge vergelijkbaarheid van de gegevens in het geding kunnen zijn.

Voorbeelden van bekende enquêtes voor arbeidsmarktonderzoek zijn:

- OSA-Arbeidsaanbodpanel (www.uvt.nl/osa/)
- Het Sociaal-economisch Panelonderzoek (www.cbs.nl)

In de praktijk wordt veelal gebruik gemaakt van administratieve bestanden met gegevens over werkzoekenden en beleidsinterventies. De belangrijkste administratieve bestanden voor het meten van de (netto-)effectiviteit van maatregelen van activerend arbeidsmarktbeleid zijn aanwezig bij de volgende organisaties:

- UWV: met name de WW- en AO-uitkeringsadministratie, registratie van Verzekerde Personen en Dienstverbanden en registratie van de re-integratiedienstverlening aan WW-

gerechtigden en arbeidsgehandicapten. De meeste van deze bestanden worden beschikbaar gesteld via het CBS

- CWI: inschrijvingsgegevens van niet-werkende werkzoekenden en vacaturegegevens
- Gemeenten: bijstandsadministratie en gegevens re-integratiedienstverlening bijstandsgerechtigden en niet-uitkeringsgerechtigden
- CBS: voor een overzicht van beschikbare bestanden met administratieve gegevens en grootschalige enquêtegegevens, zie de catalogus microbestanden op www.cbs.nl

Tabel 4.3: Overzicht van factoren die van invloed kunnen zijn op de uitkomstmaten in Tabel 4.1

Persoonskenmerken	Baankenmerken (huidige of laatste baan)
<ul style="list-style-type: none"> • Opleidingsniveau • Opleidingsrichting • Leeftijd of geboortedatum • Geslacht • Arbeidservaring (arbeidsjaren, werkloosheidsperiodes, aantal vervulde banen, etc.) • Overige (verworven) competenties • Uitkeringsverleden (aantal malen in uitkering, type uitkering, duur uitkering, etc.) • Mate van beschikbaarheid (o.a. soort dienst, gewenste aantal arbeidsuren) • Arbeidsmotivatie • Perceptie eigen arbeidsmogelijkheden • Gezondheid of gezondheidsklachten • Gezondheidsbeleving • Arbeidsongeschiktheidspercentage • Afstand tot arbeid (fase-indeling of soortgelijke indicator) • Etniciteit of land van herkomst (van ouders) • Burgerlijke staat • Huishoudsituatie • Aantal kinderen • Leeftijd van kinderen • Regio 	<ul style="list-style-type: none"> • Loon en looncomponenten • Arbeidsduur • Arbeidstijden • Type arbeidscontract (tijdelijk of vast, flexibele arbeid) • Beroep(sniveau) • Bedrijfstak • Bedrijfs grootte • Indicator seizoensarbeid • Gevraagde opleidingsniveau • Andere vereiste of gewenste competenties
	<hr/> <p>Kenmerken van (huidige of laatste) uitkering</p> <hr/> <ul style="list-style-type: none"> • Uitkeringsniveau • Uitkeringsomvang (als percentage van volledige uitkering) • Uitkeringssoort • Potentiële uitkeringsduur (recht op uitkering, bijvoorbeeld bij aanvang van uitkeringsperiode)
<hr/> <p>Omgevingskenmerken</p> <hr/> <ul style="list-style-type: none"> • Moment van evaluatie (jaar, kwartaal) • Conjuncturele omstandigheden naar tijd en regio. Voorbeelden van gangbare indicatoren voor de economische situatie die relevant zijn voor de regionale arbeidsmarkt zijn: <ul style="list-style-type: none"> ○ Omvang werkloosheid gerelateerd aan omvang werkgelegenheid, bij voorkeur voor verschillende segmenten van de arbeidsmarkt (b.v. opleidingsniveau) ○ Aantal vacatures gerelateerd aan omvang werkgelegenheid, bij voorkeur voor verschillende segmenten van de arbeidsmarkt (b.v. beroepsniveau) ○ Verhouding omvang werkloosheid en aantal vacatures (U/V-ratio), bij voorkeur voor verschillende segmenten van de arbeidsmarkt ○ Economische groei, uitgedrukt in toename toegevoegde waarde van productie ○ Arbeidsproductiviteit als de omvang van de toegevoegde waarde per arbeidsplaats <hr/>	

Bewerking van gegevens

Beschikbare gegevens zijn niet onmiddellijk bruikbaar voor een effectiviteitsanalyse. Veel gegevens dienen eerst te worden bewerkt, zodat ze betekenisvol kunnen worden ingezet. Een eerste belangrijke stap in de bewerking van gegevens voor een effectiviteitsanalyse is de selectie van de relevante onderzoekspopulatie. Daarbij gaat het in de eerste plaats om de doelgroep waarvoor de effectiviteit van een beleidsinterventie wordt berekend. Belangrijk is om te realiseren dat effecten van beleidsinterventies die zijn vastgesteld met behulp van een effectiviteitsanalyse op een afgebakende onderzoekspopulatie ook alleen geldig zijn voor die afgebakende onderzoekspopulatie.

Voor de bepaling van duurperioden, zoals werkloosheidsduren en baanduren, zijn goede datums van het begin en eind van de toestand waarvoor een duur wordt bepaald van groot belang. Deze datums komen vaak uit administratieve bestanden, waarin deze gegevens niet altijd exact worden geregistreerd. Om schijnnaauwkeurigheid te voorkomen, kan het beste gebruik worden gemaakt van maandgegevens. Als kan worden gekozen, dan wordt aanbevolen om alleen begindatums te gebruiken in een effectiviteitsanalyse, omdat die over het algemeen nauwkeuriger worden geregistreerd dan einddatums. Het is telkens van belang de uitkomstmaat goed te definiëren en duidelijk te maken op basis van welke grootheden de uitkomstmaat wordt vastgesteld.

Wat nu precies wordt bedoeld met de beleidsinterventie waarvan het effect in een effectiviteitsanalyse wordt gemeten, dient altijd zeer helder te worden gemaakt in de evaluatie. Bestaat de interventie bijvoorbeeld uit de inzet van een re-integratietraject of uit de inzet van een re-integratie-instrument waarvan er meerdere kunnen worden ingezet binnen een re-integratietraject? Ook het moment waarop een interventie begint en eindigt dient duidelijk te worden bepaald. Daar hangt immers de werkingsduur van de interventie op de uitkomstmaat van af. De empirische specificatie van beleidsinterventies in effectiviteitsanalyses is altijd een afweging tussen de gewenste definitie en de beschikbaarheid van gegevens. Voor de interpretatie van de analyseresultaten is het daarom van groot belang om keuzes in de uiteindelijke empirische specificatie te verantwoorden.

Wanneer men geïnteresseerd is in het gemiddelde effect van een interventie zoals die is ingezet in een populatie, dan kan men volstaan met het berekenen van een *overall* effect in die populatie. De informatieve waarde van een gemiddeld effect van ingezette instrumenten in de gehele populatie is echter beperkt, omdat die afhangt van de specifieke inzet van bepaalde instrumenten bij bepaalde (groepen van) personen. Wanneer men geïnteresseerd is in welke beleidsinterventie er nu precies voor wie werkt, dan is het van belang om het effect van de beleidsinterventie uit te splitsen naar doelgroepen en ingezette instrumenten.

Effecten van de beleidsinterventie kunnen apart worden vastgesteld voor verschillende omstandigheden waaronder de beleidsinterventie plaatsvindt. Zo kan het effect van bemiddeling afhangen van het moment waarop het wordt ingezet (vroeg of laat gedurende de werkloosheidsperiode). Een ander voorbeeld is dat het effect van scholing anders is tijdens hoogconjunctuur dan tijdens laagconjunctuur. In een effectiviteitsanalyse kunnen daarvoor verschillende effecten worden onderscheiden, afhankelijk van het moment van inzet en de stand van de conjunctuur.

Presentatie van effecten

Vastgestelde effecten kunnen op verschillende manieren worden gepresenteerd. Het ligt voor de hand om effecten uit te drukken in een verandering van de uitkomstmaat. Maar zowel de uitkomstmaat als effecten kunnen sterk variëren afhankelijk van de situatie waarin ze zich voordoen. Het is daarom van belang heel precies te zijn in de presentatie van effecten. Voor welke samenstelling van de populatie worden effecten gepresenteerd? Voor welke periode of voor

welk moment? Voor welke combinatie van beleidsinstrumenten? Bij inzet op welk tijdstip? Moeten de resultaten worden uitgedrukt in absolute effecten of relatieve effecten?

Belangrijk is om te benadrukken dat vastgestelde effecten van beleidsinterventies uitsluitend gelden voor de specifieke onderzoekspopulatie en onderzoeksperiode waarop de effectmeting heeft plaatsgevonden. Effecten kunnen ook worden berekend voor personen die buiten de onderzoekspopulatie vallen, bijvoorbeeld op grond van een extrapolatie van leeftijdseffecten, of buiten de onderzoeksperiode, bijvoorbeeld als ze een voorspelling voor de toekomst inhouden. Zo'n extrapolatie van effecten wordt uitgevoerd onder de veronderstelling dat de effectiviteit van de beleidsinterventie niet wezenlijk verschilt binnen en buiten de onderzochte onderzoekspopulatie en -periode. Die veronderstelling kan echter niet worden getoetst. Als er al effectberekeningen plaatsvinden buiten de onderzoekspopulatie en -periode, dienen deze daarom goed te worden verantwoord.

Omdat effecten vaak afhangen van de individuele situatie van personen, kan het beste een gemiddeld effect worden gepresenteerd. Belangrijk daarbij is de samenstelling van de populatie waarover dat gemiddelde wordt berekend. Aangezien men over het algemeen geïnteresseerd is in een gemiddeld effect voor de gehele onderzoekspopulatie, kan het beste het gemiddelde effect worden vastgesteld over de gehele onderzoekspopulatie. Dit wordt bereikt door het berekenen van de uitkomstmaat mét en zonder beleidsinterventie voor alle personen in de onderzoekspopulatie en daarover het gemiddelde verschil te nemen. Het zo berekende effect is specifiek voor de samenstelling van de onderzoekspopulatie en kan dus niet goed worden vergeleken met de inzet van dezelfde beleidsinterventie in een andere onderzoekspopulatie. Om de inzet van bijvoorbeeld bemiddeling te vergelijken tussen bijvoorbeeld de WW-populatie en de bijstandspopulatie, kan ervoor worden gekozen om uit te gaan van een referentie-persoon met typische karakteristieken die in beide populaties voorkomt (bijvoorbeeld oudere mannen met een grote afstand tot de arbeidsmarkt). Op die manier kan worden achterhaald wat het verschil in effectiviteit van bemiddeling is voor vergelijkbare personen in de WW- en bijstandspopulatie. Daarbij moet wel worden bedacht dat de referentiepersoon nooit volledig vergelijkbaar is tussen beide populaties, omdat onvergelijkbare kenmerken (zoals het arbeidsverleden) ervoor zorgen dat een persoon zich in de ene of de andere populatie bevindt.

Behalve tussen personen in de onderzoekspopulatie kunnen uitkomstmaten en effecten van beleidsinterventies ook sterk variëren in de tijd. Een effect kan een maand na de beleidsinterventie anders zijn dan een jaar na de beleidsinterventie, afhankelijk van de variatie in de uitkomstmaat en de toe- of afname van de effectiviteit in de loop van de tijd. Het kan daarom van belang zijn om een eenduidig tijdstip te kiezen waarop het effect van een beleidsinterventie wordt berekend. Aanbevolen wordt om dat tijdstip kort te leggen na de volledige implementatie van de interventie. Bij re-integratietrajecten moet bijvoorbeeld rekening worden gehouden met de duur tot aan de start van een traject en de doorlooptijd van het traject zelf. In de praktijk betekent dat een evaluatie van de werkloosheidsduur of de kans op werkherhervatting het beste ongeveer anderhalf jaar na de start van een traject kan plaatsvinden, of rekening houdend met het feit dat de meeste trajecten binnen een jaar na aanvang van werkloosheid worden ingezet, tussen twee en twee en een half jaar na aanvang van de werkloosheid.

Ook bij een tijdsvariabele inzet van beleidsinterventies treden er in de praktijk verschillen op in uitkomstmaten en effecten tussen personen in de populatie. Voor de berekening van een gemiddeld effect dient een 'gemiddeld' moment van inzet te worden gekozen om te abstraheren van de tijdsvariabele inzet van beleidsinterventies en zo individuele personen vergelijkbaar te maken. Dit 'gemiddelde' moment van inzet dient voor elke persoon in de populatie gelijk te zijn, omdat anders geen zuivere vergelijking plaatsvindt tussen personen. Dit geldt in het bijzonder wanneer het werkelijke moment van inzet bij personen selectief is, bijvoorbeeld op grond van leeftijd of afstand tot arbeid. In de berekening en presentatie van effecten dient voor die selectiviteit te worden gecorrigeerd. Het gekozen 'gemiddelde' moment van inzet dient bij voorkeur dicht bij het gemiddelde moment te liggen waarop interventies in werkelijkheid worden

ingezet. Bij de inzet van re-integratietrajecten voor werklozen dus ergens tussen een half jaar en een jaar na aanvang van de werkloosheid.

Tot slot wordt nog gewezen op de relatie tussen de uitkomstmaat en de gepresenteerde effecten. Zo kan een effect worden uitgedrukt in het absolute verschil in uitkomstmaat, bijvoorbeeld een absolute afname van de werkloosheidsduur, uitgedrukt in dagen of maanden. Maar een effect kan ook worden gerelateerd aan de uitkomstmaat en worden uitgedrukt als relatieve verschil in uitkomstmaat, bijvoorbeeld een procentuele stijging van het inkomen. Aanbevolen wordt om effecten bij een zogenaamde continue uitkomstmaat (zoals uitkeringsduur of inkomen) uit te drukken in een relatieve verandering van de uitkomstmaat (procenten) en effecten bij een zogenaamde discrete uitkomstmaat (zoals werkhervattingkans of aandeel moeilijk vervulbare vacatures) uit te drukken in een absolute verandering (procentpunten).

5 Evaluatiemethoden

In dit hoofdstuk worden de belangrijkste statistische evaluatiemethoden beschreven die beschikbaar zijn voor het vaststellen van het effect van (beleids)interventies (bijvoorbeeld een re-integratietraject) op van te voren vastgestelde individuele uitkomstmaten (bijvoorbeeld de kans op werkhervatting). Doel van het hoofdstuk is om een idee te geven van de mogelijkheden die statistische evaluatiemethoden bieden om onder verschillende omstandigheden, afhankelijk van het soort uitkomstmaat, het soort interventie en beschikbare gegevens, het effect van beleidsinterventies te meten. Om één en ander te verduidelijken wordt in dit hoofdstuk veelvuldig gebruik gemaakt van gestileerde voorbeelden uit de beleidspraktijk.

Counterfactual

Als men het daadwerkelijke netto-effect van maatregelen van activerend arbeidsmarktbeleid op individueel niveau zou willen meten, dan zou dezelfde persoon tegelijkertijd moeten worden vergeleken in een situatie waarin hij of zij wél met deze maatregelen heeft te maken en een situatie waarin hij of zij daar niet mee te maken heeft. Zo zijn er voor één en dezelfde werkzoekende in feite twee mogelijke uitkomsten voor de kans op werkhervatting, namelijk de kans op werkhervatting bij deelname aan een re-integratietraject en de kans op werkhervatting zonder deelname aan een re-integratietraject. In werkelijkheid kan men echter slechts één van deze twee situaties waarnemen. De werkzoekende neemt wel of niet deel aan een re-integratietraject, maar nooit allebei tegelijk. De niet waargenomen situatie blijft per definitie onbekend: dit staat in de literatuur bekend als de *counterfactual*.¹⁴ Het effect van het re-integratietraject op de kans op werkhervatting kan voor individuele werkzoekenden daarom nooit direct worden waargenomen. Maar dat betekent niet dat het effect van een beleidsinterventie niet zou kunnen worden vastgesteld.

Vergelijking tussen proef- en controlegroep

Om effecten van beleidsinterventies vast te kunnen stellen, dient een persoon die te maken heeft met een beleidsinterventie te worden vergeleken met een persoon die daar niet mee te maken heeft. Zo kan de totale groep werkzoekenden worden verdeeld in twee groepen, namelijk een groep personen die deelneemt aan een re-integratietraject (de proefgroep) en een groep personen die niet deelneemt aan een re-integratietraject (de controlegroep). Wanneer beide groepen goed vergelijkbaar zijn, dan kunnen verschillen in de uitkomstmaat (zoals de kans op werkhervatting) tussen beide groepen volledig worden toegeschreven aan de werking van de beleidsinterventie. Dat verschil is de netto-effectiviteit of toegevoegde waarde van de interventie.

Onderzoeksmethoden gericht op de evaluatie van de netto-effectiviteit van (beleids)interventies zijn dan ook primair gericht op het zo goed mogelijk vergelijken van de proefgroep (de groep die wel te maken heeft met een bepaalde interventie) met de controlegroep (de groep die niet te maken heeft met die interventie). Merk op dat de proef- en controlegroep worden gedefinieerd aan de hand van de inzet van een *bepaalde* interventie. In de controlegroep kan zowel sprake zijn van het ontbreken van een beleidsinterventie (geen re-integratietraject in het voorbeeld) als van een alternatieve beleidsinterventie (een ander re-integratietraject). In het eerste geval is bij de evaluatie van beleidsmaatregelen sprake van een *absolute* netto-effectiviteit (toegevoegde waarde van de interventie), in het tweede geval van een *relatieve* netto-effectiviteit (verschil in toegevoegde waarde tussen interventies). Het is daarom altijd van belang goed vast te stellen ten opzichte van welke situatie (of 'nul-alternatief') een effect wordt gemeten.

¹⁴ Degene die dit probleem identificeerde en de eerste oplossingsmethoden uitwerkten zijn Roy (1951), Rubin (1974), Holland (1986) and (Splawa)-Neyman (1990).

Samenstelling- en selectie-effecten

In de praktijk zijn de proefgroep en controlegroep niet altijd goed vergelijkbaar. Daardoor is het verschil in de uitkomstmaat (zoals de kans op werkhervatting) tussen de proef- en controlegroep niet altijd volledig toe te schrijven aan de werking van de beleidsinterventie (het re-integratietraject). Ten eerste kunnen personen in beide groepen verschillen in kenmerken die de uitkomstmaat direct beïnvloeden, zoals leeftijd en opleidingsniveau. Verschillen in de uitkomstmaat die direct worden veroorzaakt door verschillen in de samenstelling van beide groepen, worden in de literatuur *samenstellingseffecten* genoemd. Ten tweede kan deelname aan een re-integratietraject (en daarmee aan de proefgroep) afhangen van bepaalde kenmerken die soms wel worden geobserveerd, bijvoorbeeld de werkloosheidsduur zelf, en soms niet, bijvoorbeeld motivatie. Verschillen in de uitkomstmaat die worden veroorzaakt door een niet-willekeurige inzet van interventies worden *selectie-effecten* genoemd. Dit kan worden geïllustreerd aan de hand van een verschil in motivatie tussen de groep met een re-integratietraject en de groep zonder een re-integratietraject. Naarmate personen gemotiveerder zijn, zullen ze waarschijnlijk eerder deelnemen aan re-integratie activiteiten, maar mogelijk ook eerder een nieuwe baan vinden. Die grotere kans op werkhervatting in de groep met een interventie is in zo'n geval niet uitsluitend toe te schrijven aan deelname aan het re-integratietraject. Zowel de grotere kans op werkhervatting als deelname aan een re-integratietraject hangen positief samen met de motivatie van deze personen. Deelnemers aan het re-integratietraject (de proefgroep) vormen in dat geval een *selectieve groep* die niet zonder meer is te vergelijken met de groep die niet deelneemt aan een re-integratietraject (de controlegroep).

Naast samenstelling- en selectie-effecten bestaan er nog een aantal effecten die de vergelijkbaarheid tussen proef- en controlegroep kunnen beïnvloeden. Zo is de zogenaamde Ashenfelter dip een bijzondere vorm van selectie in de proefgroep (zie Ashenfelter, 1978). Het idee daarbij is dat de omstandigheden voorafgaand aan de interventie ertoe hebben geleid dat er een interventie heeft plaatsgevonden. Het zonder meer vergelijken van proef- en controlegroep voorafgaand aan de interventie is dan niet zuiver. Een illustratie van zo'n situatie is het ontslag van een voetbalcoach (de interventie). Vaak wordt er na het ontslag waargenomen dat het team beter gaat presteren (uitkomstmaat), waaruit vaak de conclusie wordt getrokken dat clubs de coach moeten ontslaan als de prestaties van een team ondermaats zijn (positieve effectiviteit). Ashenfelter's punt is dat het ontslaan van de coach gebeurt als het team in een dip zit en zo lijkt het alsof de coachwissel betere prestaties met zich meebrengt. In werkelijkheid is de waargenomen verbetering een natuurlijk proces, waarbij een terugkeer van het prestatieniveau naar het gemiddelde wordt waargenomen. Bij een goede vergelijking tussen proef- en controlegroep zal daarom ook voor het verschil in omstandigheden voorafgaand aan de interventie moeten worden gecorrigeerd.

Anticipatie effecten

Ten slotte kan de vergelijkbaarheid tussen de proef- en controlegroep ook nog worden beïnvloed door anticipatie-effecten. Als werkzoekenden weten dat ze op korte termijn in aanmerking komen voor een re-integratietraject, dan kunnen zij hun inspanning om een baan te vinden alvast aanpassen. Dat beïnvloedt de vergelijkbaarheid met werkzoekenden die niet in aanmerking komen voor een re-integratietraject of wel in aanmerking komen maar dat niet weten. Hoe het zoekgedrag wordt aangepast in de proefgroep ten opzichte van de controlegroep kan niet van tevoren worden vastgesteld. Pas wanneer beide groepen goed vergelijkbaar zijn, rekening houdend met alle relevante kenmerken die naast de interventie bepalend zijn voor de uitkomstmaat, dan kunnen verschillen in de uitkomstmaat tussen beide groepen volledig worden toegeschreven aan de werking van de beleidsinterventie.

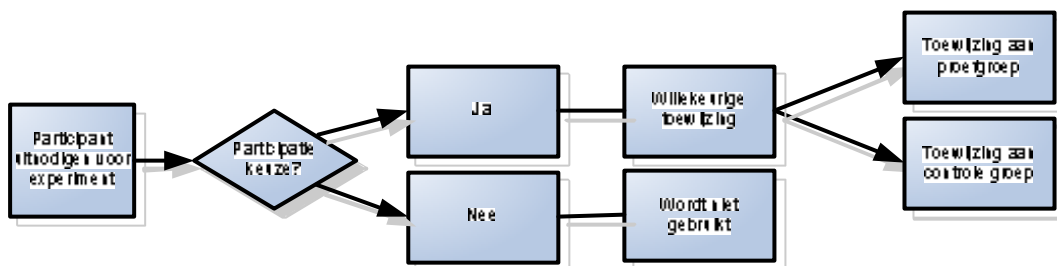
Afhankelijk van de uitkomstmaat, het type beleidsinterventie en beschikbare gegevens voor analyse, kan door middel van specifieke evaluatiemethoden worden gecorrigeerd voor samenstellingseffecten, selectieeffecten en anticipatie-effecten. De onderzoeksmethoden die

hieronder worden besproken zijn primair gericht op het zo goed mogelijk vergelijken, of het zo goed mogelijk vergelijkbaar maken, van de proef- en controlegroep.

5.1 Sociaal experiment

Een eerste methode voor het bepalen van de effectiviteit van een interventie is een sociaal experiment. Figuur 5.1 laat zien hoe een sociaal experiment wordt uitgevoerd. De figuur kan worden uitgelegd aan de hand van het voorbeeld waarin het effect van een re-integratietraject op werkhervatting van personen wordt gemeten. Uit de totale groep werklozen worden willekeurig potentiële deelnemers uitgenodigd om deel te nemen aan het experiment. Dat kan de volledige groep werklozen zijn. Deze potentiële deelnemers maken vervolgens een keuze of zij dit wel of niet willen. Vervolgens worden de daadwerkelijke deelnemers die aangeven te willen participeren in het experiment willekeurig ingedeeld in twee groepen. Eén van deze twee groepen krijgt een re-integratietraject (de proefgroep), de andere groep krijgt geen re-integratietraject (de controlegroep).

Figuur 5.1: Maatschappelijk experiment met willekeurige toewijzing, Cameron & Trivedi (2005)



De belangrijkste veronderstelling bij een experiment is dat de toekenning van de interventie onafhankelijk is van de uitkomstmaat (geen selectie-effecten). Dit wordt bereikt door de willekeurige toewijzing aan de proef- en controlegroep. Door deze manier van toewijzing zijn ook de kenmerken van deelnemers willekeurig verdeeld over de twee groepen en die willekeur wordt groter naarmate de groepsgrootte toeneemt. Door het ontbreken van systematische verschillen in samenstelling tussen de proef- en controlegroep (geen samenstellingseffecten) kunnen op basis van deze experimentele setting verschillen in de gemiddelde werkloosheidsduur tussen de proefgroep en controlegroep toegewezen worden aan de werking van het re-integratietraject.¹⁵ Bolhaar, de Jong, van der Klaauw en Lindeboom (2004) voerden een experiment uit om het effect van intensieve toetsing van het reïntegratieverslag op onder andere de WAO-instroom te meten, waarbij in een beperkt aantal regio's deze intensieve toetsing werd toegepast en in andere niet. Uit die studie bleek dat de intensieve toetsing er toe leidde dat minder mensen die langdurig ziek waren een aanvraag voor een WAO-uitkering deden.

Een experiment is een gecontroleerde onderzoekssituatie, waarmee het de meest ideale manier is om effecten van beleidsinterventies te isoleren van andere factoren die invloed hebben op de uitkomstmaat. Toch is het in de praktijk niet altijd even gemakkelijk om een sociaal experiment op te zetten. Door de gecontroleerde onderzoekssituatie zijn de kosten van het uitvoeren van een experiment relatief hoog, terwijl de verkregen resultaten onderwerpsspecifiek zijn. Het generaliseren van de resultaten naar andere beleidsinterventies is niet per definitie mogelijk. Ook neemt de organisatie en uitwerking van experimenten relatief veel tijd in beslag, tijd die vaak niet beschikbaar is binnen het beleidsproces. Sociale experimenten waarbij mogelijk sprake is van rechtsongelijkheid (wel versus geen beleidsinterventie) zijn juridisch niet haalbaar. Wel kunnen twee alternatieve beleidsinterventies tegenover elkaar worden gezet, waarbij niet van te voren

¹⁵ Zie voor toepassingen van experimenten voor het meten van de effectiviteit van beleidsinterventies onder meer studies door Sacerdote (2001), Angrist en Lavy (20002) en Leuven, Oosterbeek en van der Klaauw (2005).

duidelijk is welke gunstiger is voor de deelnemer. Daarmee zijn relatieve netto-effecten vast te stellen (verschil in toegevoegde waarde van interventies), maar geen absolute netto-effecten (toegevoegde waarde van de interventie). Bovendien kunnen personen omwille van ethische redenen niet gedwongen worden om deel te nemen aan een experiment en moeten zij van tevoren worden ingelicht over deelname aan het experiment. Dit kan leiden tot selectieve deelname aan het experiment, waardoor de resultaten van de evaluatie niet kunnen worden veralgemeniseerd naar de totale doelpopulatie.

5.2 Natuurlijke of quasi experimenten

Wanneer het niet mogelijk is om een sociaal experiment uit te voeren, dan kan men op zoek gaan naar een natuurlijk of quasi experiment. Daarbij is sprake van een situatie in de werkelijkheid die een gecontroleerde onderzoekssituatie benadert. Voorwaarde is dat een min of meer willekeurig deel van de populatie wordt geconfronteerd met een verandering (de interventie) waarop zij niet heeft kunnen anticiperen, terwijl er voor het overige deel van de populatie niks verandert. Stel bijvoorbeeld dat een volstrekt willekeurig deel van de populatie werklozen tijdelijk wordt uitgesloten van de inzet van re-integratietrajecten als gevolg van een administratieve fout, dan zou die situatie kunnen worden gezien als een natuurlijk experiment. Een dergelijke min of meer willekeurige interventie wordt een exogene verandering of een exogene schok genoemd. Andere voorbeelden van exogene veranderingen die groepen binnen een populatie treffen zijn plotselinge veranderingen of toevalligheden in beleid of wetgeving¹⁶ of een natuurramp die slechts een willekeurig deel van de bevolking treft. Exogene schokken op individueel niveau zijn bijvoorbeeld de volgorde waarin mensen worden geboren of de verschillende levenservaringen van tweelingen.¹⁷ Of het effect van een interventie geëvalueerd kan worden door middel van een vergelijking van de op natuurlijke wijze ontstane proef- en controlegroep, hangt af van de volgende criteria:

- De interventie moet daadwerkelijk exogeen zijn, personen moeten niet hebben kunnen anticiperen op de interventie, waardoor de inzet niet meer volstrekt willekeurig is¹⁸. Als werklozen in het gegeven voorbeeld tijdig op de hoogte zijn van de administratieve fout, zullen ze actie ondernemen om andersoortige hulp bij hun re-integratie te ontvangen;
- De exogene verandering moet hebben geleid tot een willekeurige toewijzing aan een proef- en controlegroep die beide van voldoende omvang zijn voor nadere evaluatie. De administratieve fout moet dus een voldoende grote groep werklozen treffen;
- De overige omstandigheden waaronder het natuurlijke of quasi experiment heeft plaatsgevonden moeten nauwkeurig kunnen worden vastgesteld, omdat (veranderingen in) deze overige omstandigheden verantwoordelijk kunnen zijn voor verschillen tussen de proef- en controlegroep.

Natuurlijke experimenten zijn wetenschappelijk gezien zeer interessant, omdat ze een gecontroleerde onderzoekssituatie nabootsen in de werkelijkheid. Daardoor kunnen effecten relatief zuiver worden gemeten zonder de hoge kosten van een vooropgezet sociaal experiment. Tegelijkertijd zijn natuurlijke experimenten zeldzaam. Om die reden wordt bij effectiviteitsanalyse vaak gebruik gemaakt van een vergelijking tussen de proefgroep en een controlegroep die op basis van beschikbare gegevens wordt geconstrueerd. In dat geval dient altijd gecontroleerd te worden voor samenstellingseffecten en selectie effecten. Deze effecten kunnen worden geïdentificeerd door gebruik te maken van variatie in relevante factoren binnen de

¹⁶ Een voorbeeld is het leeftijdsverschil binnen hetzelfde schooljaar (vroeg versus late leerlingen).

¹⁷ Studies die gebruik maken van exogene variatie zijn bijvoorbeeld Angrist en Krueger (1991), Ashenfelter en Krueger (1994), Angrist en Evans (1996, 1998) en Plug (2004).

¹⁸ Technisch betekent dit dat de interventie exogeen moet zijn met betrekking tot alle factoren in de effectmeting. De interventie kan namelijk weliswaar exogeen zijn met betrekking tot een bepaalde uitkomstmaat, maar endogeen met betrekking tot factoren die deze uitkomstmaat mede beïnvloeden.

onderzoekspopulatie op één bepaald moment (cross sectie variatie). Relevante factoren zijn de uitkomstmaat (bijvoorbeeld de kans op werkhervatting), factoren die van invloed zijn op de uitkomstmaat (bijvoorbeeld opleidingsniveau) en de interventie zelf (bijvoorbeeld een re-integratietraject). Maar identificatie kan ook plaatsvinden op basis van variatie in relevante factoren over tijd. Voor elk van deze gevallen zijn er verschillende evaluatiemethoden beschikbaar, die hieronder apart worden besproken.

5.3 Het benutten van cross sectie variatie

Cross sectie variatie is de variatie in factoren die geldt op één bepaald moment tussen individuen of groepen van individuen binnen de onderzoekspopulatie. Zo is bijvoorbeeld wel bekend of er een interventie heeft plaatsgevonden, maar niet precies wanneer. Evaluatiemethoden die zich voor de identificatie van effecten richten op het benutten van cross sectie variatie, zijn de discontinuïteiten benadering en de matching methode.

5.3.1 Discontinuïteiten benadering

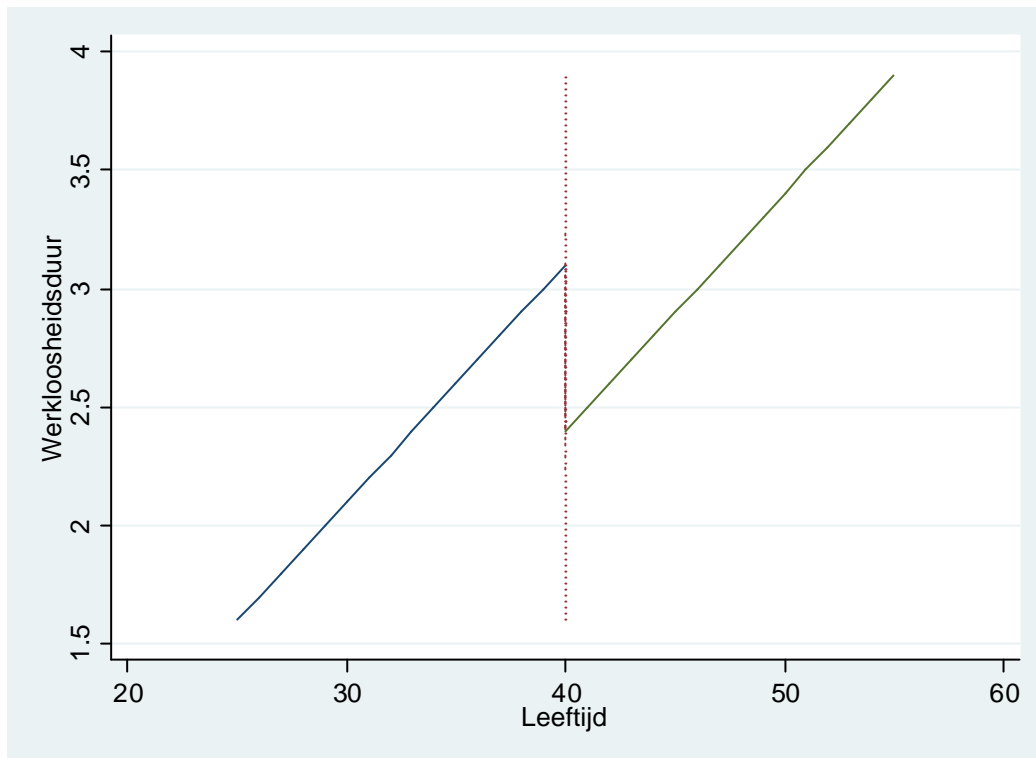
Een discontinuïteiten benadering is een bijzondere toepassing van een natuurlijk experiment waarbij de kans op een beleidsinterventie een discontinue functie is van één of meerdere onderliggende factoren. Dit kan worden verduidelijkt aan de hand van een voorbeeld. Stel dat werklozen boven de 40 eerder in aanmerking komen voor een re-integratietraject dan personen beneden de 40 jaar. Figuur 5.2 toont gestileerd de waargenomen gemiddelde werkloosheidsduur in maanden voor personen van verschillende leeftijden.¹⁹ De discontinuïteiten benadering is gebaseerd op de aanname dat personen om en nabij de 40 jaar (het breekpunt) vergelijkbaar zijn. Zonder de inzet van een re-integratietraject is de werkloosheidsduur voor personen net boven de 40 jaar (de proefgroep) naar verwachting gelijk aan die voor personen net onder de 40 jaar (de controlegroep). Het mechanisme dat bepaald of een persoon in de proef- dan wel controlegroep zit, is de bepaling dat personen vanaf 40 jaar eerder in aanmerking komen voor een re-integratietraject. Omdat personen om en nabij de 40 jaar verder vergelijkbaar zijn in hun kenmerken, verwachten we dat de duur van werkloosheid niet veel zal verschillen tussen personen die 39 jaar oud zijn en personen die 41 jaar oud zijn. Echter, uit Figuur 5.2 blijkt dat de werkloosheidsduur voor personen die iets ouder zijn dan 40 korter is in vergelijking met die van personen die iets jonger zijn dan 40. Dit duidt erop dat een grotere deelname aan een re-integratietraject samenhangt met een kortere werkloosheidsduur.²⁰

Omdat een discontinuïteiten benadering een toepassing is van een natuurlijk experiment, kan deze methode alleen worden toegepast wanneer de interventie gezien kan worden als exogene schok. Er moet dus sprake zijn van een min of meer arbitraire beleidsinterventie. Wanneer werklozen weinig invloed hebben op het moment van werkloosheid en de inzet van een re-integratietraject afhankelijk is van leeftijd, dan is aan die voorwaarde voldaan. Maar wanneer (een deel van de groep) werklozen in staat zijn het moment van werkloosheid uit te stellen tot net na de leeftijd van 40 jaar, dan zijn de proefgroep (boven 40 jaar) en controlegroep (beneden 40 jaar) niet meer goed vergelijkbaar. Er is dan sprake van mogelijke anticipatie op de beleidsinterventie.

¹⁹ Figuur 5.2 is niet gebaseerd op werkelijke cijfers en dus op geen enkele wijze een weergave van de werkelijkheid.

²⁰ Zie ook Berkhout e.a. (2008) waar een discontinuïteiten benadering wordt gebruikt om het effect te meten van de wijze waarop het budget van het inkomensdeel van de Wet Werk en Bijstand wordt bepaald op het bijstandsvolume in gemeenten. Deze wijze hangt af van de omvang van gemeenten. Of mensen in soortgelijke gemeenten, maar van verschillende grootte, in de bijstand (blijven) zitten mag vergelijkbaar worden verondersteld, met uitzondering van de wijze waarop het WWB budget wordt vastgesteld. Deze discontinuïteit (of breekpunt) wordt gebruikt om het effect op het bijstandsvolume te identificeren. Leuven en Oosterbeek (2004) gebruiken een soortgelijke evaluatiemethode om te onderzoeken wat het effect is van een maatregel die scholing aftrekbaar maakt voor personen die ouder zijn dan 40 jaar.

Figuur 5.2: Gestileerde waargenomen relatie tussen leeftijd en werkloosheidsduur



5.3.2 Matching

Het belangrijkste probleem dat zich bij een discontinuïteiten benadering voordoet, is dat onduidelijk is of de controlegroep (zonder interventie) wel goed vergelijkbaar is met de proefgroep (met interventie). De discontinuïteiten benadering richt zich op een vergelijking van de uitkomstmaat tussen de proef- en controlegroep, waarbij zoveel mogelijk wordt gecorrigeerd voor de invloed van samenstellingseffecten door middel van een econometrisch model. Hoe meer informatie beschikbaar is over personen en de omstandigheden waarin ze verkeren, hoe beter het verschil in de uitkomstmaat tussen de proef- en controlegroep kan worden verklaard. Des te zuiverder het overgebleven verschil in de uitkomstmaat kan worden toegeschreven aan de beleidsinterventie.

Een econometrisch model en bijbehorende modelspecificatie blijven echter altijd een benadering van de werkelijke invloed van kenmerken van personen en van de omgeving waarin ze verkeren op de uitkomstmaat. De vaststelling van het effect van een beleidsinterventie is daardoor een schatting van het werkelijke effect. Het vaststellen van een verschil in het effect van kenmerken tussen twee groepen is altijd minder nauwkeurig dan het vaststellen van het verschil in die kenmerken zelf. Wanneer de proef- en controlegroep volledig vergelijkbaar zijn, met uitzondering van de beleidsinterventie, kunnen overgebleven verschillen in de uitkomstmaat volledig worden toegeschreven aan die beleidsinterventie. Alleen bij een sociaal experiment wordt in een gecontroleerde onderzoekssituatie een min of meer vergelijkbare controlegroep gecreëerd.

Het doel van de matching methode is het construeren van een zo vergelijkbaar mogelijke proef- en controlegroep. Dit gebeurt door personen voor wie een interventie wordt waargenomen (de proefgroep) te koppelen aan personen voor wie geen interventie wordt waargenomen (de zogenaamde referentiegroep) op basis van zoveel mogelijk dezelfde kenmerken die van invloed zijn op de uitkomstmaat. Zo wordt bijvoorbeeld een hoogopgeleide oudere vrouw uit Groningen die in een re-integratietraject zit gekoppeld aan een andere hoogopgeleide oudere vrouw uit

Groningen die (toevallig) niet in een re-integratietraject zit. Door koppeling van personen uit een relevante referentiegroep ontstaat een met de proefgroep vergelijkbare controlegroep. Een verschil in de uitkomstmaat tussen beide groepen kan dan worden toegeschreven aan de werking van de beleidsinterventie. De matching methode is vooral relevant als er veel informatie beschikbaar is over personen en de omstandigheden waarin zij verkeren. Alleen dan kunnen proef- en controlegroep daadwerkelijk vergelijkbaar worden gemaakt, waarna het verschil in uitkomstmaat niet verder hoeft te worden verklaard.

De koppeling van personen tussen proefgroep en referentiegroep kan plaatsvinden op twee verschillende manieren, namelijk via *exact matching* en via *matching gebruik makend van een afstandsmaat*. Bij exact matching worden personen uit de proefgroep gekoppeld aan personen in de referentiegroep onder de strikte aanname dat de relevante kenmerken van beide personen exact overeen komen.²¹ In ons voorbeeld van de hoogopgeleide oudere vrouw uit Groningen komt uitsluitend een andere hoogopgeleide oudere vrouw uit Groningen die geen re-integratietraject heeft ontvangen in aanmerking voor koppeling. Bij deze methode wordt de kans op een match kleiner naarmate kenmerken van een persoon minder gangbaar zijn. Wanneer hoogopgeleide oudere vrouwen verder niet voorkomen in Groningen, wordt dit type persoon uitgesloten van verdere evaluatie. Er is dan sprake van selectieve matching, waardoor vooral effecten kunnen worden gemeten van personen met gemiddelde kenmerken.²²

Vanwege deze selectiviteit wordt in de praktijk vaak gebruik gemaakt van een matching methode die er voor zorgt dat de mate waarin personen in de proef- en controlegroep van elkaar verschillen wordt geminimaliseerd. Deze methode maakt gebruik van een afstandsmaat die het verschil in meerdere kenmerken samenvat. Zo zou de hoogopgeleide oudere vrouw uit Groningen wellicht kunnen worden gekoppeld aan een wel bestaande hoogopgeleide oudere man uit Limburg, wanneer de combinatie van geslacht en regio een beperkte invloed heeft op de uitkomstmaat. Personen met incidenteel voorkomende kenmerken kunnen op die manier wel in de controlegroep worden meegenomen, wanneer de overige kenmerken van deze personen meer gemiddeld zijn.

In Hoofdstuk 6 wordt nader ingegaan op de manier waarop verschillen in kenmerken worden uitgedrukt in een afstandsmaat, die via de matching methode wordt geminimaliseerd. Hier beperken we ons tot de vraag wanneer een match tussen personen in de proef- en controlegroep kan worden opgevat als een goede match. De voorwaarden waaraan een goede match moet voldoen zijn de volgende:

- *Volledigheid van kenmerken voor matching*: Er dient gematcht te worden op alle factoren die invloed hebben op de uitkomstmaat. Als belangrijke invloedsfactoren niet zijn opgenomen in de kenmerken waarop wordt gematcht, dan zijn personen in de proef- en controlegroep uiteindelijk niet goed vergelijkbaar. Daardoor is een effectmeting op basis van deze matching ook niet zuiver.²³
- *Onafhankelijkheid van beleidsinterventie*: Na constructie van de proef- en controlegroep door de koppeling van personen op basis van kenmerken, dient het verschil in samenstelling van de twee groepen onafhankelijk te zijn van de beleidsinterventie. Anders gezegd, de gemiddelde kans op een beleidsinterventie mag niet verschillen tussen de proef- en controlegroep als rekening wordt gehouden met de waargenomen kenmerken waarop is gematcht. Dit kan alleen als de inzet van de interventie uitsluitend afhangt van

²¹ Zie bijvoorbeeld van Klaveren en Maassen van den Brink (2007).

²² Voor personen die in de staart van de verdeling van bepaalde kenmerken zitten, geldt dat er met een kleinere kans een match wordt gevonden in vergelijking met personen met meer gemiddelde kenmerken. Dit wordt ook wel 'regression towards the mean' genoemd.

²³ Deze voorwaarde staat in de literatuur bekend als de 'unconfoundedness' voorwaarde (Imbens, 2005) of de 'ignorability' voorwaarde (Rubin, 1978; Wooldridge, 2001) en kan worden vergeleken met de beter bekende 'omitted variable bias' die zich voordoet als niet alle relevante variabelen in een regressie worden opgenomen.

deze kenmerken en niet wordt beïnvloed door kenmerken die niet in de matching zijn betrokken. Als de toewijzing naar de proef- en controlegroep willekeurig heeft plaatsgevonden, dan wordt per definitie aan deze voorwaarde voldaan.²⁴

- *Gelijke basis van kenmerken*: Alle kenmerken die worden waargenomen in de proefgroep, dienen ook waargenomen te worden in de controlegroep en andersom, hoewel niet noodzakelijk in gelijke mate. Dus als er in het voorbeeld van de hoger opgeleide oudere vrouw uit Groningen die een re-integratietraject ontvangt niemand uit Groningen in de controlegroep zit, dan is de vergelijking tussen proef- en controlegroep niet zuiver. De intuïtieve interpretatie van deze voorwaarde is dat de controlegroep op alle dimensies een goede vergelijkingsgroep moet zijn voor de proefgroep.²⁵

Wanneer aan deze voorwaarden wordt voldaan, leidt de matching methode in een situatie met uitsluitend cross sectie gegevens tot een zuiverdere schatting van het effect van een beleidsinterventie dan andere evaluatiemethoden, zoals de discontinuïteiten benadering. Verschillen tussen de proef- en controlegroep zijn immers zo klein mogelijk gemaakt, zodat het effect ervan niet meer benaderd hoeft te worden door een noodzakelijk beperkte modelspecificatie in een econometrische analyse.

In 1983 hebben Rosenbaum en Rubin aangetoond dat de kans om een combinatie van relevante kenmerken waar te nemen in de proefgroep gelijk is aan de kans op een beleidsinterventie (kans om in de proefgroep terecht te komen), geconditioneerd op deze relevante kenmerken. Dus de kans dat een hoogopgeleide oudere vrouw uit Groningen in de proefgroep zit is even groot als de kans op een interventie voor een hoogopgeleide oudere vrouw uit Groningen. Dat betekent dat personen gekoppeld kunnen worden op basis van de kans om in de proefgroep of referentiegroep terecht te komen, waarbij die kans wordt verklaard op grond van deze relevante kenmerken. Deze bevinding heeft als grote voordeel dat matching kan worden gebaseerd op de individuele kans op een interventie die afhankelijk is van alle relevante kenmerken voor matching. Om die kans te berekenen kan worden volstaan met een simpele kansanalyse (Probit of Logit). Zo'n kansanalyse maakt het voldoen aan de voorwaarde van een gelijke basis van kenmerken een stuk eenvoudiger. Een gelijke basis van kenmerken kan worden bereikt door alleen personen in de referentiegroep te beschouwen waarvoor geldt dat hun individuele kans op een beleidsinterventie niet hoger of lager is dan van alle personen in de controlegroep (en andersom).

Het matchen van personen op basis van de kans op een interventie wordt 'Propensity Score Matching' genoemd. De voorspelde kans waarop wordt gematcht, wordt de Propensity Score genoemd. Een groot voordeel van Propensity Score Matching is dat door de monotoon stijgende en continue Propensity Score geen dimensieproblemen ontstaan bij koppeling van personen in de proef- en referentiegroep, zoals die zich kunnen voordoen bij matching op meerdere individuele relevante kenmerken samen (zoals bij exact matching). De Propensity Score kan worden gezien als een afstandsmaat die zich leent voor elke willekeurige toepassing van de matching methode. Die toepassing betreft de manier waarop een koppeling met een gegeven afstand plaatsvindt. Hierop wordt in Hoofdstuk 6 verder ingegaan.

De matching methode richt zich volledig op het zo vergelijkbaar mogelijk krijgen van de proef- en controlegroep. De invloed van samenstelling- en selectie-effecten op de effectmeting, voorzover gebaseerd op waargenomen kenmerken, is dan ook kleiner bij toepassing van matching dan bij elke andere evaluatiemethode, met uitzondering van de experimentele setting. Studies hebben laten zien dat via matching de effectiviteit van beleidsinterventies zuiverder wordt vastgesteld dan via andere ex-post evaluatiemethoden, zolang alle relevante invloedsfactoren

²⁴ Deze voorwaarde staat in de literatuur bekend als de 'conditional mean independence assumption'.

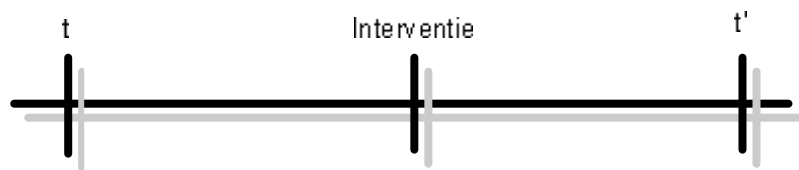
²⁵ In de literatuur staat deze voorwaarde bekend als de 'common support' aanname. Het idee erachter is dat wanneer zowel uit de proefgroep als de controlegroep een persoon wordt geselecteerd op basis van de kenmerken voor matching, beide personen geselecteerd moeten zijn uit dezelfde kansverdeling.

maar in de analyse betrokken worden. Soms variëren relevante invloedsfactoren echter in de tijd, hetgeen niet op basis van cross sectie gegevens kan worden waargenomen. Soms worden relevante invloedsfactoren in z'n geheel niet waargenomen. In die gevallen kan het gebruik van longitudinale gegevens met variatie in de tijd een oplossing bieden.

5.4 Benutten van variatie in de tijd

Wanneer er informatie beschikbaar is over personen op meerdere momenten in de tijd, waaronder informatie over de uitkomstmaat, dan kan ook variatie in de tijd gebruikt worden om verschillen in een uitkomstmaat toe te schrijven aan verschillen in kenmerken en omstandigheden tussen personen. Om het effect te meten van een interventie kan voor- en na de interventie de uitkomstmaat worden vergeleken voor dezelfde persoon.²⁶ Personen fungeren in dat geval zelf als controlegroep. Figuur 5.3 illustreert die situatie. Op tijdstip t is er een voormeting, vervolgens vindt de interventie plaats en daarna is er een nameting op tijdstip t' . Op tijdstip t' wordt de uitkomst waargenomen inclusief het effect van de interventie. De benodigde gegevens hebben een longitudinaal karakter, omdat er informatie aanwezig moet zijn voor personen over tijd.²⁷ Effecten kunnen in deze opzet alleen zuiver worden gemeten onder de veronderstelling dat het verschil in de uitkomstmaat tussen t en t' niet wordt beïnvloed door persoonlijke, huishoudelijke en/of economische veranderingen in de tijd. In de praktijk is aan deze voorwaarde alleen in uitzonderlijke gevallen voldaan. Daarom wordt in werkelijkheid vaak gebruik gemaakt van de zogenaamde difference-in-difference methode, die hieronder wordt toegelicht.

Figuur 5.3: Vergelijking van uitkomsten voor en na een interventie voor personen waarbij een interventie heeft plaatsgevonden



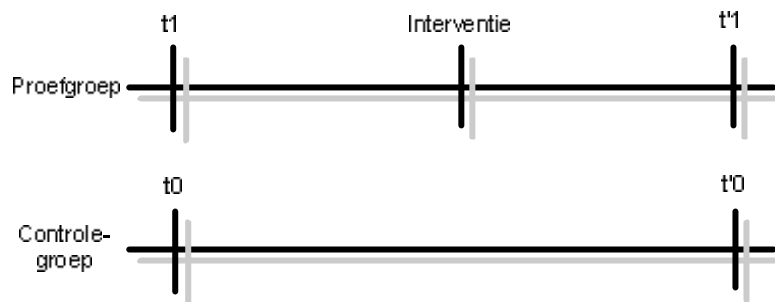
5.4.1 Difference-in-difference methode

De zogenaamde difference-in-difference methode, ook wel kortweg aangeduid met ‘diff-in-diff’, houdt rekening met veranderingen in de tijd door te kijken naar de invloed van deze veranderingen op de uitkomstmaat in een controlegroep. Figuur 5.4 geeft een indruk van de werking van de diff-in-diff methode. Apart voor de proef- en controlegroep is een tijdas getekend. De voormeting vindt plaats op tijdstip t en de nameting op tijdstip t' . Een 1 (0) achter het tijdstip geeft aan of het een meting betreft voor een persoon in de proefgroep (controlegroep). Volgens de diff-in-diff methode kan het effect van een interventie op een uitkomstmaat bepaald worden door eerst afzonderlijk voor de proefgroep en de controlegroep de uitkomstverschillen in de tijd te meten en vervolgens het verschil van deze verschillen te meten tussen de proef- en controlegroep. Een groot voordeel van de diff-in-diff methode is dat de proef- en controlegroep niet volledig vergelijkbaar hoeven te zijn, zolang verschillen maar constant zijn in de tijd.

²⁶ In de literatuur spreekt men over de vergelijking voor en na de interventie, of zoals in Heckman (1999) ‘The Before-After Estimator’. Een correctere benaming is de vergelijking voor en na de interventie voor personen waarbij een interventie heeft plaatsgevonden.

²⁷ Heckman & Rob (1985a,b) laten zien dat ook een herhaalde cross-sectie steekproef voldoende is om de voor en na schatter toe te passen. Het idee is dat een representatieve steekproef voor en na de interventie allebei een onnauwkeurigheid met zich meebrengt. Bij de bepaling van het gemiddelde verschil in uitkomst zal deze onnauwkeurigheid gemiddeld 0 zijn.

Figuur 5.4: Difference-in-difference methode: vergelijking tussen personen in proef- en controlegroep van de uitkomst voor en na een interventie



Toepassing van de difference-in-difference methode kan worden verduidelijkt aan de hand van een voorbeeld. Stel dat werklozen jonger dan 25 jaar vanaf tijdstip t^* plotseling geen werkloosheidsuitkering meer ontvangen wanneer ze hun baan verliezen. Wat voor effect heeft dat op de kans om binnen een jaar na aanvang van de werkloosheid een nieuwe baan te hebben? Het effect van deze beleidsinterventie kan worden gemeten door de populatie werklozen op te delen in een proefgroep jonger dan 25 jaar en een controlegroep ouder dan 25 jaar. De eerste meting dient ruim een jaar voorafgaand aan t^* plaats te vinden (voormeting) en de tweede meting ruim een jaar na t^* (nameting). In totaal zijn er dus vier metingen nodig: een voormeting en een nameting in zowel de proefgroep als de controlegroep. De difference-in-difference methode vergelijkt nu het verschil in de uitkomstmaat (aandeel werkhervattingen binnen een jaar na aanvang van de werkloosheid) tussen de twee metingen voor zowel de proef- als de controlegroep. Het verschil in de uitkomstmaat tussen de voor- en nameting binnen de controlegroep wordt alleen beïnvloed door veranderingen in de tijd. Het verschil in de uitkomstmaat tussen de voor- en nameting in de proefgroep wordt naast veranderingen in de tijd ook beïnvloed door de beleidsinterventie. Onder de veronderstelling dat veranderingen in de tijd op dezelfde manier doorwerken op de proef- en controlegroep, kan het effect van het plotseling afschaffen van de werkloosheidsuitkering voor jongeren worden geïdentificeerd door een vergelijking van de verschillen tussen de voor- en nameting tussen de proef- en controlegroep. Natuurlijk wordt de vergelijking beter en de aanname van een gelijke doorwerking van veranderingen in de tijd op de proef- en controlegroep meer valide wanneer beide groepen beter op elkaar lijken. Dit kan worden bewerkstelligd door voorafgaand aan de diff-in-diff evaluatie beide groepen vergelijkbaar te maken door middel van matching.

De Jong, Lindeboom en van der Klaauw (2005) hebben de difference-in-difference methode toegepast bij een experiment waarin het effect van intensieve toetsing van het re-integratieverslag op onder meer de WAO-instroom werd gemeten. Daarbij werd in een aantal regio's vanaf een bepaald moment de intensieve toetsing toegepast. Door het meten van het verschil in WAO-instroom voor en na de start van de intensieve toetsing in zowel de regio's waar deze werd toegepast (de proefgroep) als de regio's waar deze niet werden toegepast (de controlegroep), kon het effect van deze beleidsinterventie worden geïsoleerd van autonome veranderingen in de WAO-instroom tussen beide tijdstippen.

De difference-in-difference methode combineert de eenvoud van een voor- en nameting met een correctie voor tijdsvariërende invloedsfactoren. De diff-in-diff methode wordt dan ook vaak gebruikt in combinatie met een natuurlijk experiment, waarbij sprake is van een exogene verandering op een bepaald moment (de interventie) voor een deel van de (onderzoeks)populatie. De diff-in-diff methode leidt tot een zuiverdere vaststelling van effecten naarmate andere (autonome) veranderingen in de tijd stabiel zijn. Een belangrijke aanname is immers dat de gemiddelde persoon in de proefgroep met dezelfde veranderingen in de tijd te maken heeft gehad als de gemiddelde persoon in de controlegroep, met uitzondering van de interventie. Deze

aanname kan worden getest voor zover het waargenomen invloedsfactoren betreft. Als niet aan de aanname wordt voldaan, kan de diff-in-diff methode beter niet worden toegepast.

Een belangrijke situatie waarin niet aan de aanname van stabiele veranderingen in de tijd wordt voldaan, is die waarin het tijdstip waarop een beleidsinterventie plaatsvindt varieert in de tijd. Wanneer in Figuur 5.4 het moment van interventie varieert tussen t_1 en t'_1 , dan zijn personen in de proefgroep niet allemaal op dezelfde manier vergelijkbaar met personen in de controlegroep. Wanneer bovendien de kans op een interventie afhangt van de tijd tussen t_1 en t'_1 , dan is de difference-in-difference methode alleen niet langer toepasbaar. Dit doet zich bij uitstek voor in een situatie waarin het effect van re-integratietrajecten op de werkloosheidsduur wordt gemeten. Door allerlei min of meer toevallige oorzaken starten re-integratietrajecten lang niet allemaal op hetzelfde moment. Daarnaast hangt de inzet van een re-integratietraject ook af van de werkloosheidsduur. In zo'n geval is het niet mogelijk om een vergelijkbaar moment van voor- en nameting te construeren en moet rekening worden gehouden met de tijdsvariërende inzet van een beleidsinterventie. Dit kan bij uitstek worden gedaan met behulp van zogenaamde duurmodellen.

5.4.2 Duurmodellen (Timing-of-events methode)

Tot dusver is aangenomen dat er één situatie is voor de interventie en één na de interventie, zoals is weergegeven in Figuur 5.3 en Figuur 5.4. In de praktijk hangt de gemeten uitkomst vaak af van het tijdstip waarop men meet. Dit is met name het geval als de uitkomstmaat zelf een functie is van de tijd, bijvoorbeeld de werkloosheidsduur of de baanduur. Ook kan een interventie variabel in de tijd plaatsvinden en bovendien afhankelijk zijn van de uitkomstmaat. Al die variatie in de tijd beïnvloedt de uitkomstmaat, zodat daar in de effectiviteitsanalyse rekening mee moet worden gehouden. Tegelijkertijd kan juist die variatie in de tijd worden gebruikt om effecten van een interventie te identificeren. Beide worden bereikt met de zogenaamde timing-of-events methode.

Als we teruggaan naar het voorbeeld waarbij het effect van een re-integratietraject op de werkloosheidsduur wordt gemeten, geldt dat personen die deelnemen aan een re-integratietraject vaak niet allemaal tegelijkertijd starten met een dergelijk traject. Een deel van die tijdsvariabele inzet is voor werkloze personen zelf volstrekt willekeurig, want afhankelijk van voor hen toevallige omstandigheden, zoals de aanwezige capaciteit bij case-managers of de startdatum van een cursus. Personen die langer op een re-integratietraject wachten hebben langer de tijd om nog voor het begin van het traject een baan te vinden. Aan de andere kant daalt over het algemeen de kans om alsnog een baan te vinden naarmate men langer zonder werk zit. Deze duurzaamheidsafhankelijkheid kan worden verklaard door het feit dat kennis en vaardigheden van personen veroudert en werklozen ook steeds meer het contact met de arbeidsmarkt, inclusief alle daarmee samenhangende sociale netwerken, verliezen. Bovendien kan een langere werkloosheidsduur demotiverend en zelfs stigmatiserend werken, zodat werkzoekenden minder hard naar werk zoeken en werkgevers minder geneigd zijn om werkzoekenden aan te nemen. Daardoor hebben twee identieke personen die op verschillende momenten een re-integratietraject ontvangen bij aanvang van het re-integratietraject niet dezelfde verwachte werkloosheidsduur, los van het effect van het re-integratietraject zelf.

Daarnaast kan sprake zijn van een gerichte en dus niet willekeurige tijdsafhankelijke inzet van re-integratietrajecten. Personen waarbij de verwachting is dat zij niet lang werkloos zijn komen niet onmiddellijk in aanmerking voor een traject. Daarmee is de inzet van re-integratietrajecten selectief in de tijd. Als een latere inzet van een re-integratietraject (tijdstip van interventie) samenhangt met een grotere kans op werkhervatting (de uitkomstmaat), dan zal daar in de vergelijking tussen proef- en controlegroep rekening mee moeten worden gehouden.

In het geval dat uitkomsten en effecten afhangen van het tijdstip waarop beleid wordt ingezet en uitkomsten worden gemeten, is het raadzaam een effectiviteitsanalyse uit te voeren met een zogenaamd duurmodel. Bij een duurmodel wordt de duur van een bepaalde periode, zoals de werkloosheidsduur, baanduur of de duur tot aan de start van een re-integratietraject, expliciet

afhankelijk gemaakt van allerlei invloedsfactoren. Eén van die factoren is de reeds verstreken duur zelf. Op die manier wordt rekening gehouden met de duurzaamheid van de uitkomstmaat. Zo kan blijken dat hoe langer men werkloos is, hoe kleiner de kans om alsnog een baan te vinden (negatieve duurzaamheid op de uitstroomkans, positieve duurzaamheid op de duur tot uitstroom). Een andere invloedsfactor op de werkloosheidsduur is de inzet van een re-integratietraject vanaf een bepaald moment in de tijd. Door dit kenmerk als tijdsvariabele factor mee te nemen in een duurmodel, wordt voor verschillende personen de invloed van de werking van het re-integratietraject gedurende verschillende perioden gemeten, afhankelijk van het tijdstip waarop het re-integratietraject wordt ingezet. Verschillen in de werkloosheidsduur kunnen daarmee dus zowel worden gerelateerd aan verschillen in de reeds verstreken werkloosheidsduur (duurafhankelijkheid) als aan verschillen in de werkingsduur van re-integratietrajecten. Daarmee kan het effect van een tijdsafhankelijke beleidsinterventie worden geïdentificeerd.

Er zijn allerlei omstandigheden die ervoor kunnen zorgen dat uitkomstmaten kunnen variëren in de tijd, zoals leer- en aanpassingsprocessen en allerlei institutionele factoren. Wanneer bijvoorbeeld de uitkeringshoogte na een jaar werkloosheid daalt, dan wordt het na een jaar werkloosheid aantrekkelijker om te gaan werken. Al die structurele en incidentele factoren die de uitkomstmaat op verschillende manieren in de tijd beïnvloeden kunnen worden opgenomen in een duurmodel als tijdsvariërende variabele of via de specificatie van duurzaamheid. Een tijdsvariërende inzet van de beleidsinterventie is slechts één van de factoren die invloed kunnen uitoefenen op de uitkomstmaat over de tijd.

Daarnaast heeft een duurmodel als voordeel dat het gebruik maakt van longitudinale gegevens (meerdere meetmomenten in de tijd), waardoor het mogelijk is om naast waargenomen kenmerken expliciet rekening te houden met de invloed van niet-waargenomen kenmerken die constant is in de tijd (zie Hoofdstuk 6 voor details). Daarmee zijn duurmodellen beter in staat dan meer statische evaluatiemethoden (zoals matching) om rekening te houden met factoren die wel invloed uitoefenen op de uitkomstmaat, maar niet direct (kunnen) worden geobserveerd.

In Nederland zijn de meeste netto-effectiviteitsanalyses van activerend arbeidsmarktbeleid uitgevoerd door middel van duurmodellen. Zo is het netto-effect van sancties op de kans op werkhervatting door middel van een duuranalyse onder andere gemeten door Van den Berg, Van der Klaauw en Van Ours (2004) voor bijstandsgerechtigden en door Abbring, Van den Berg en Van Ours (2005) voor WW-gerechtigden. Ook de netto-effectiviteit van re-integratietrajecten, inclusief bemiddeling en (intensieve) begeleiding, is vaak met behulp van duurmodellen geanalyseerd. Voor WW-gerechtigden is dat gedaan door onder meer Gorter en Kalb (1996), Centraal Planbureau (2000), Heyma, Zwinkels en Van Seters (2003), Heul, Berendsen, van der Eijken en Vlek (2003), Heyma (2005), Graaf-Zijl, Groot, Heyma, Hop en Janssen (2005), Graaf-Zijl, Groot en Hop (2006) en Graaf-Zijl en Hollanders (2007). Voor de meting van de netto-effectiviteit van re-integratietrajecten bij bijstandsgerechtigden zijn onder meer duurmodellen toegepast door Graaf-Zijl, Heyma en de Hoop (2006) voor de gemeente Amsterdam, Graaf-Zijl, Groot en Hop (2006) op basis van de MOSA en Zwinkels, Hekelaar en Braat (2006) voor de gemeente Rotterdam. Duurmodellen voor de meting van de netto-effectiviteit van re-integratietrajecten voor arbeidsgehandicapten zijn gebruikt in Heyma (2002), Heyma, Zwinkels en van Seters (2003), Heyma en Zijl (2005), waar het ging om herbeoordeelde WAO-gerechtigden, Graaf-Zijl, Groot en Hop (2006) voor alleen volledig arbeidsongeschikten en UWV (2008) waarin ook voor gedeeltelijk arbeidsongeschikten waren betrokken. Ten slotte is er een duuranalyse uitgevoerd door Groot, de Graaf-Zijl en Hop (2007) naar de netto-effectiviteit van re-integratietrajecten bij niet-uitkeringsgerechtigden.

5.5 Combinatie van evaluatiemethoden

Uit het overzicht van evaluatiemethoden tot dusver is gebleken dat matching zich onderscheidt van andere methoden door de primaire gerichtheid op het vergelijkbaar krijgen van de proef- en controlegroep. Na matching wordt de uitkomstmaat vergeleken tussen proef- en controlegroep en het verschil daartussen toegeschreven aan de werking van de beleidsinterventie. Indien er factoren zijn die niet in de matching kunnen worden betrokken maar toch invloed uitoefenen op de uitkomstmaat, zoals variatie in de tijd of niet-waargenomen invloedsfactoren, dan heeft het zin om na matching nog een van de andere evaluatiemethoden toe te passen.

Het beste voorbeeld is de combinatie van matching met een duurmodel. Matching zorgt eerst voor een goede vergelijkbaarheid van de proef- en controlegroep op basis van waargenomen invloedsfactoren die worden gemeten op één bepaald moment in de tijd. Correctie voor verschillen in de uitkomstmaat tussen proef- en controlegroep als gevolg van de (verschillende) invloed van die factoren hangt na matching dus minder af van de functionele vorm in het duurmodel. Vervolgens wordt de tijdsafhankelijkheid van de uitkomstmaat en de inzet van de beleidsinterventie vormgegeven in het duurmodel. Daarnaast heeft een duurmodel het voordeel boven matching dat de (tijdsinvariante) invloed van niet-waargenomen factoren op de uitkomstmaat kan worden geïdentificeerd aan de hand van een vergelijking van de uitkomstmaat op meerdere momenten in de tijd. Daardoor wordt de vergelijkbaarheid van proef- en controlegroep verder verbeterd (zie ook Paragraaf 6.3.4). Bij matching voorafgaand aan een duuranalyse kunnen niet alle koppelingenregels voor matching worden toegepast, omdat in een duurmodel personen worden vergeleken en niet alleen de uitkomstmaat van personen (zie ook Paragraaf 6.2.1).

Een ander voorbeeld van de combinatie van evaluatiemethoden is het gebruik van de difference-in-difference methode in combinatie met een duurmodel. Wanneer de werkloosheidsduur als uitkomstmaat wordt gekozen bij de evaluatie van het effect van het afschaffen van het recht op een werkloosheidsuitkering voor jongeren tot 25 jaar (zie het voorbeeld in Paragraaf 5.4.1), dan kan een duurmodel worden gebruikt voor de voor- en nameting van de werkloosheidsduur in de proef- en controlegroep. Door een vergelijking van de werkloosheidsduur in de proef- en controlegroep voor en na afschaffing van het recht op een werkloosheidsuitkering voor jongeren tot 25 jaar (difference-in-difference), kan het effect van deze beleidsinterventie worden geïdentificeerd.

5.6 Welke evaluatiemethode is het meest geschikt?

Welke evaluatiemethode in welk geval het meest geschikt is hangt af van veel verschillende factoren. De keuze van een evaluatiemethode en vooral de toegepaste invulling ervan is een min of meer subjectieve keuze van degene die de effectiviteitsanalyse uitvoert. Op voorhand is niet voor elke situatie te zeggen wat de beste evaluatiemethode is. Een empirische specificatie dient uiteindelijk empirisch te worden bepaald én getoetst. Toch kunnen op grond van de bestaande literatuur in algemene termen wel handreikingen worden gedaan over welke evaluatiemethoden in welke situaties goed presteren. Dat is ten slotte ook het doel van dit handboek.

Grofweg hangt de keuze van de optimale evaluatiemethode af van de volgende factoren:

- het moment van evalueren: gedurende het beleidsproces of via een ex-post analyse;
- de beschikbare gegevens: op meerdere momenten of niet;
- het type beleidsinterventie: inzet varieert in de tijd of alleen tussen groepen;
- vergelijkbaarheid van en mate van overlap tussen proef- en controlegroep, en
- de uitkomstmaat: duurafhankelijk, kansmaat of anders.

Belangrijk is dat de inzet van een beleidsinterventie varieert ten opzichte van de te meten uitkomstmaat. Die variatie kan zich voordoen tussen groepen binnen de doelpopulatie, maar ook

in de tijd. Variatie zorgt voor verschillen en verschillen zijn nodig om effecten te kunnen identificeren. Als iedereen binnen de doelpopulatie op hetzelfde moment te maken krijgt met een beleidsinterventie, dan is het effect ervan niet te meten.

Om enige richting te geven in het antwoord op de vraag welke evaluatiemethode in welke situatie het meest geschikt is, kan een aantal aanvullende vragen worden gesteld. Via het antwoord op die vragen kan een indicatie worden gegeven welke evaluatiemethode voor de specifieke beleidspraktijk geschikt zou zijn voor het meten van de netto-effectiviteit. Die aanvullende vragen en antwoorden worden samengevat in het schema van Figuur 5.1.

Ten eerste is het de vraag of de effectiviteitsanalyse gedurende de beleidsinterventie kan plaatsvinden of pas achteraf. Als er ruimte is om vooraf de evaluatie te organiseren kan het beste gebruik worden gemaakt van een *experiment*. In een experiment worden de proef- en controlegroep bij voorbaat vergelijkbaar gemaakt door de beleidsinterventie willekeurig in de populatie in te zetten. Bijvoorbeeld door slechts een willekeurig deel van de populatie werklozen actief te bemiddelen naar werk. De andere groep ontvangt dan geen bemiddeling (voor de berekening van absolute netto-effecten) of een alternatieve beleidsinterventie (voor de berekening van relatieve netto-effecten). In de praktijk stuit het willekeurig inzetten van beleidsinterventies vaak op financiële, juridische en ethische bezwaren. De meeste evaluatiemethoden bestaan daarom uit ex-post evaluaties, waarbij de mogelijk selectieve inzet van de beleidsinterventie al heeft plaatsgevonden.

In zeer uitzonderlijke gevallen is achteraf sprake van een spontane experimentele situatie die toevallig is ontstaan en niet van te voren is opgezet. Zo'n situatie wordt aangeduid met een *natuurlijk of quasi experiment*. De doelpopulatie is in dat geval bij toeval willekeurig toegedeeld aan de proef- en controlegroep, denk bijvoorbeeld aan willekeurige administratieve fouten, waardoor een deel van de doelpopulatie niet of pas later te maken krijgt met de beleidsinterventie. Ook kan het voorkomen dat sterk gelijkende groepen wel en niet te maken krijgen met een beleidsinterventie, bijvoorbeeld als gevolg van een leeftijds criterium in regelgeving. In dat geval kan gebruik worden gemaakt van een *discontinuïteiten benadering*, waarbij een proef- en controlegroep worden samengesteld uit personen die net wel, respectievelijk net niet aan het leeftijds criterium voldoen. Door vergelijking van deze sterk gelijkende (deel)groepen in de onderzoekspopulatie kan het netto-effect van de beleidsinterventie worden geïdentificeerd. Bestaan dit soort situaties niet, dan zal een ex-post evaluatie op één of andere manier rekening moeten houden met samenstelling- en selectie-effecten voor een goede vergelijking van de uitkomstmaat tussen proef- en controlegroep.

De beste ex-post methode om de proef- en controlegroep vergelijkbaar te maken, is de zogenaamde *matching* methode. Met de matching methode wordt een controlegroep samengesteld die qua kenmerken zoveel mogelijk lijkt op de proefgroep. Een vergelijking tussen de uitkomstmaat tussen beide groepen geeft vervolgens het netto-effect van de inzet van de beleidsinterventie. Omdat de samenstelling van de controlegroep wordt gebaseerd op waargenomen kenmerken van personen en de situatie waarin zij verkeren, leent deze methode zich bij uitstek voor een netto-effectmeting als er gegevens over veel verschillende kenmerken beschikbaar zijn, die bovendien zowel binnen als buiten de proefgroep voorkomen. In dat geval is de kans groot dat een soortgelijke groep als de proefgroep kan worden samengesteld, waardoor verschillen in samenstelling en selectie worden geëlimineerd.

Er is echter een aantal situaties waarin de matching methode (alleen) geen goede oplossing biedt voor het verschil tussen de proef- en controlegroep. Het eerste voorbeeld daarvan is de situatie waarin de beleidsinterventie variabel in de tijd wordt ingezet, zoals het geval is bij de meeste re-integratietrajecten. Niet iedereen krijgt in de praktijk op hetzelfde moment een re-integratietraject aangeboden, waardoor de proef- en controlegroep niet alleen qua kenmerken, maar ook in de tijd vergelijkbaar moet worden gemaakt. Dit is met name belangrijk wanneer de uitkomstmaat zelf afhankelijk is van de tijd, zoals de werkloosheidsduur, baanduur en zoekduur. De matching

methode is daar minder geschikt voor dan bijvoorbeeld een *duurmodel*. Een duurmodel berekent de kans op beëindiging van bijvoorbeeld de werkloosheidsduur, afhankelijk van allerlei invloedsfactoren, waaronder het tijdstip van inzet van de beleidsinterventie. De uitkomstmaat wordt dus afhankelijk van de verstreken duur mét en zonder beleidsinterventie gemodelleerd. Overigens zijn veel kansmaten ook afhankelijk van de (verstreken) tijd, zoals de kans op werkhervatting, de kans op een stapje op de re-integratieladder en de kans op een hernieuwde instroom in de SUWI-keten. Wanneer de inzet van de beleidsinterventie variabel in de tijd plaatsvindt en de uitkomstmaat duurafhankelijk is, dan kan het beste gebruik worden gemaakt van een duurmodel. Voorwaarde is wel dat de beschikbare gegevens het mogelijk maken om variatie in de tijd op individueel niveau te meten (longitudinale gegevens).

Overigens sluiten de matching methode en een duurmodel elkaar niet uit. Zo kan matching worden gebruikt voor het zo goed mogelijk vergelijkbaar maken van de proef- en controlegroep op basis van kenmerken op één moment in de tijd. Vervolgens kan een duurmodel worden gebruikt om de tijdsafhankelijke ontwikkeling tussen de proef- en controlegroep te vergelijken. Deze combinatie van evaluatiemethoden wordt sterk aanbevolen.

Een andere situatie waarin de matching methode geen goede oplossing biedt voor het verschil tussen proef- en controlegroep, is die waarin er relatief weinig overlap is tussen beide groepen. Dat is onder meer het geval wanneer de inzet van een te evalueren beleidsinterventie sterk selectief is. Als voorbeeld kan de inzet van uitstroompremies worden genoemd wanneer deze uitsluitend én talrijk worden ingezet bij ouderen. De samenstelling van de proef- en controlegroep zijn op basis van matching dan niet goed in overeenstemming te krijgen, omdat er een te groot verschil in leeftijd bestaat tussen proef- en controlegroep. Leeftijd is nu eenmaal een belangrijke verklarende factor voor de meeste uitkomstmaten. Matching gaat in dat geval ten koste van een voldoende aantal waarnemingen voor de effectiviteitsanalyse, omdat de proefgroep alleen gematched kan worden met de zeer kleine groep ouderen die géén uitstroompremie hebben ontvangen. In zo'n geval kan de *difference-in-difference* methode uitkomst bieden. Bij de diff-in-diff methode wordt het verschil in uitkomstmaat voor en na de interventie gemeten voor zowel de proefgroep als de controlegroep. Het 'verschil van het verschil' kan vervolgens – na correctie voor samenstelling- en selectie-effecten – worden toegeschreven aan de inzet van uitstroompremies. Als matching het verschil in samenstelling niet kan elimineren, dan biedt de difference-in-difference methode een manier om met dat verschil om te gaan.

Bij de difference-in-difference methode dient nog wel gecorrigeerd te worden voor samenstelling- en selectie-effecten. Dat kan worden gerealiseerd met behulp van een econometrisch model (zie ook Paragraaf 6.1). Ook hier geldt dat in het geval van een duurafhankelijke uitkomstmaat, zoals een duurvariabele of een tijdsafhankelijke kansmaat, gebruik kan worden gemaakt van een duurmodel.

Alle ex-post evaluatiemethoden tot dusver gaan uit van de beschikbaarheid van gegevens op meerdere meetmomenten in de tijd, namelijk voorafgaand en volgend op de beleidsinterventie.²⁸ Er zijn situaties waarin achteraf niet meer kan worden vastgesteld wat de uitkomstmaat was voorafgaand aan de beleidsinterventie. Er is dan geen informatie beschikbaar over de variatie in de uitkomstmaat in de tijd, zodat evaluatiemethoden die daarvan gebruik maken, zoals een duuranalyse of een difference-in-difference methode, niet toepasbaar zijn. De matching methode blijft in dat geval wel toepasbaar, omdat het een controlegroep samenstelt die sterke gelijkenissen vertoont met de proefgroep. Die sterke gelijkenis maakt het waarschijnlijk dat ook de uitkomstmaat ten tijde van de ontbrekende voormeting vergelijkbaar is. Dit gaat echter alleen op

²⁸ Relevante gegevens over de situatie voorafgaand aan de inzet van een beleidsinterventie kunnen in veel gevallen ook achteraf worden vastgesteld, bijvoorbeeld op basis van administratieve gegevens. Pas wanneer deze gegevens niet meer zijn te achterhalen, spreken we van een niet beschikbare voormeting.

zolang er gegevens over voldoende invloedfactoren beschikbaar zijn om de controlegroep echt een vergelijkbare samenstelling te geven als de proefgroep. Wanneer het aantal waargenomen invloedfactoren beperkt is en de overlap in kenmerken gering, zal ook de vergelijking tussen de proefgroep en de geconstrueerde controlegroep niet zuiver zijn. In dat geval kan niet langer gebruik worden gemaakt van een econometrisch model voor de effectiviteitsanalyse, maar zal gebruik moeten worden gemaakt van andere evaluatiemethoden.

Figuur 5.5: Beslisboom voor keuze van de meest geschikte evaluatiemethode



6 Toegepaste specificaties van methoden

Aan het eind van Hoofdstuk 5 is een indruk gegeven van geschikte evaluatiemethoden voor het meten van effecten van maatregelen van activerend arbeidsmarktbeleid, afhankelijk van de beschikbare gegevens voor evaluatie en van de gekozen uitkomstmaat. Als eenmaal voor een evaluatiemethode is gekozen, dient de manier waarop die wordt toegepast nog verder te worden gespecificeerd. Daarbij dient zo veel mogelijk rekening te worden gehouden met de beschikbaarheid van gegevens, de gekozen uitkomstmaat en de specifieke situatie waarin de uitkomstmaat wordt gemeten. Zo wordt werkloosheid bij een WW-gerechtigde overwegend beëindigd door werkhervatting, maar bij een bijstandsgerechtigde ook vaak door terugtrekking van de arbeidsmarkt. Dat verschil heeft consequenties voor de manier waarop de kans op werkhervatting in bijvoorbeeld een duurmodel dient te worden gespecificeerd. Dit hoofdstuk vormt een leidraad om bij specifieke effectiviteitsanalyses tot een optimale specificatie van de evaluatiemethodiek te komen.

Er kan onderscheid worden gemaakt in drie belangrijke groepen van evaluatiemethoden waarvoor in dit hoofdstuk enkele specificaties nader worden uitgelegd. Dit zijn de min of meer standaard econometrische vergelijkingen die vooral van toepassing zijn op de discontinuïteiten benadering en de difference-in-difference methode (Paragraaf 6.1), de matching methode (Paragraaf 6.2) en duurmodellen (Paragraaf 6.3). Bij elke specificatie wordt een voorbeeld gegeven van een situatie waarin deze kan worden toegepast, wordt het idee achter de specificatie beschreven en wordt indien mogelijk een aanbeveling gedaan over de te kiezen specificatie.

6.1 Standaard econometrische vergelijkingen

In standaard econometrische vergelijkingen wordt een uitkomstmaat UM ‘verklaard’ uit een aantal invloedsfactoren X , zoals kenmerken van de persoon en van omstandigheden waarin hij of zij zich bevindt, en de beleidsinterventie BI .²⁹ Dit kan formeel worden beschreven als

$$UM = f(X, BI) \quad (6.1)$$

De relatie $f(\cdot)$ tussen de uitkomstmaat en de invloedsfactoren kan elke willekeurige lineaire of niet-lineaire vorm aannemen. Als men bijvoorbeeld wil meten of de inzet van scholing leidt tot een hoger loon in de daarop volgende baan, dan is UM het loon, BI een indicator van de inzet van scholing en X een vector van kenmerken die de hoogte van het loon bepalen. In de praktijk wordt bij loonvergelijkingen vaak een exponentiële functie voor $f(\cdot)$ gebruikt, zodat de standaard econometrische loonvergelijking gelijk is aan

$$UM = \exp[(X \times \mathbf{b}) + (BI \times \mathbf{g}) + \mathbf{e}] \quad (6.2)$$

De invloed van de factoren X op het loon UM wordt weergegeven door \mathbf{b} en de toegevoegde waarde van scholing door \mathbf{g} . De term \mathbf{e} houdt rekening met toevallige afwijkingen in het loon die niet kunnen worden gerelateerd aan de waargenomen invloedsfactoren of de werking van scholing. Veronderstellingen over de verdeling van \mathbf{e} bepalen uiteindelijk de functionele relatie tussen de uitkomstmaat en de ‘verklarende’ variabelen.

²⁹ Strikt genomen wordt de uitkomstmaat UM in een econometrische vergelijking niet verklaard, maar slechts gerelateerd aan de invloedsfactoren X en de beleidsinterventie BI . De richting van de correlatie tussen deze factoren, ofwel de causaliteit, wordt daarbij verondersteld te lopen in de richting van de uitkomstmaat. Dit dient in evaluatieonderzoek te worden verantwoord op basis van theoretische en empirische argumenten.

Wanneer de uitkomstmaat in plaats van het loon bijvoorbeeld werkhervatting is, dan kan de econometrische vergelijking worden uitgedrukt in een kansmodel, ofwel

$$KansopU = f(X, BI) \quad (6.3)$$

De kans dat de uitkomst U zich voordoet wordt uitgedrukt in de verklarende factoren X en de inzet van scholing. Een ander voorbeeld van een kansvergelijking is de kans dat een vacature binnen een bepaalde periode, zeg een half jaar, wordt vervuld (de uitkomst U). Die kans kan worden gerelateerd aan allerlei economische omstandigheden, zoals de algemene economische groei, het werkloosheidsniveau en de omvang van de uitstroom uit het onderwijs (factoren X), plus de beleidsinterventie dat een vacature wel of niet wordt gepubliceerd in de CWI-vacaturebank (BI). Daarmee kan het effect van het opnemen van een vacature in de CWI-vacaturebank worden geïdentificeerd en uitgedrukt in een toename van de kans op vervulling van de vacature. Door voor $f(\cdot)$ een normale verdeling te kiezen, wordt deze vergelijking een zogenaamd Probit model. Wordt er gekozen voor een exponentiële verdeling, dan volgt het zogenaamde Logit model.

In een discontinuïteiten benadering kunnen standaard econometrische vergelijkingen rechtstreeks worden toegepast. Stel dat men het effect van scholing op het aanbod van arbeid wil meten, waarbij scholing alleen wordt ingezet bij mensen tot 35 jaar. Dan kan men kijken naar het aanbod van arbeid van mensen van ongeveer 35 jaar (UM) en dat aanbod relateren aan allerlei kenmerken van deze mensen en de economische omstandigheden waarin zij zich bevinden (X). Voor de groep tot 35 jaar wordt de indicator voor scholing (BI) dan gelijkgesteld aan 1, voor mensen boven 35 jaar wordt deze gelijkgesteld aan 0. Op basis van een standaard econometrische vergelijking, zoals vergelijking (6.2) geeft g aan wat het effect is van scholing op het arbeidsaanbod.

Ook bij de difference-in-difference methode kunnen standaard econometrische vergelijkingen worden gebruikt. Als voorbeeld kan men denken aan de invoering van een vacaturebank in een bepaalde regio. Zorgt die er nu voor dat er meer vacatures worden vervuld? Met een difference-in-difference methode kan een vergelijking worden gemaakt tussen twee regio's voor en na de invoering van de vacaturebank, dus afhankelijk van tijdstip t . De vervulling van vacatures (UM) kan worden verklaard uit allerlei economische en regionale omstandigheden (X), plus het feit dat één van regio's vanaf een bepaald moment te maken krijgt met de vacaturebank (BI). De econometrische vergelijking kan dan worden geschreven als

$$UM(t, r) = a(t) + a(r) + (X(t, r) \times b) + (BI(t, r) \times g) + e(t, r) \quad (6.4)$$

Constante verschillen tussen de regio's worden verklaard door $a(r)$, constante verschillen in de tijd door $a(t)$, verschillen in omstandigheden die variëren tussen regio's en in de tijd door $X(t, r)$, zodat het effect van de vacaturebank g kan worden toegeschreven aan de inzet van de vacaturebank op een bepaald moment in een bepaalde regio, $BI(t, r)$.

6.2 Matching

Met de matching methode wordt geprobeerd om een zo goed mogelijke controlegroep te simuleren. Dit gebeurt door personen voor wie een interventie wordt waargenomen (de proefgroep) te koppelen aan personen uit de groep voor wie geen interventie wordt waargenomen (de referentiegroep). Het deel van de referentiegroep dat wordt gekoppeld aan de proefgroep vormt de controlegroep. Koppeling vindt plaats op basis van zoveel mogelijk kenmerken die invloed kunnen hebben op de uitkomstmaat. In deze paragraaf wordt ingegaan op de verschillende manieren waarop die matching kan plaatsvinden. Daarbij wordt onderscheid gemaakt tussen matching op statische kenmerken en matching op duurvariabelen.

6.2.1 Matching op cross-sectie gegevens

De koppeling van personen in de proefgroep met personen in de referentiegroep op basis van cross-sectie gegevens (gegevens die gelden op een bepaald moment in de tijd) vindt in twee stappen plaats. Eerst wordt een afstandsmaat gecreëerd voor het verschil in relevante kenmerken tussen de proef- en controlegroep. Deze afstandsmaat laat zien in welke mate personen in de proef- en referentiegroep nu eigenlijk op elkaar lijken. Vervolgens wordt een koppelingsregel toegepast die bepaalt welke personen uit de proef- en referentiegroep op basis van deze afstandsmaat met elkaar worden vergeleken.

Afstandsmaat voor koppeling

Stel men wil het effect meten van de inzet van scholing op het loonniveau van werklozen die vervolgens een baan vinden. Krijgen werklozen die scholing hebben gevolgd ook meer betaald dan werklozen die geen scholing hebben gevolgd? Bijvoorbeeld omdat scholing resulteert in een hogere productiviteit, die gedeeltelijk terechtkomt bij de werknemer in de vorm van een hogere beloning. Het loon dat werknemers ontvangen leent zich prima voor een directe vergelijking door middel van een econometrische analyse, waarin het loonniveau wordt verklaard uit kenmerken van personen, van hun baan en van de omstandigheden waarin ze verkeren, zoals de conjunctuur (zie de vorige paragraaf). Verschillen in het loon tussen de groep met scholing (de proefgroep) en de groep zonder scholing (de controlegroep) die niet kunnen worden verklaard door de in het model opgenomen kenmerken, kunnen worden toegeschreven aan de inzet van scholing.

De verklaring van het loonniveau aan de hand van een econometrische vergelijking is een benadering van het werkelijke proces waarmee lonen tot stand komen. De gebruikte functionele vorm kan onzuiverheden bevatten in de verklaring van het loonniveau aan de hand van invloedsfactoren X die kunnen doorwerken in de vaststelling van het effect van de beleidsinterventie, zoals het effect van scholing. Bij de matching methode worden onzuiverheden in de verklaring van het loonniveau vermeden door het loon tussen de proef- en controlegroep direct te vergelijken. Er hoeft na matching dus geen econometrische vergelijking plaats te vinden. Voorwaarde is wel dat beide groepen uit min of meer dezelfde personen bestaan. Of personen min of meer hetzelfde zijn, kan worden vastgesteld aan de hand van een afstandsmaat die het verschil in kenmerken tussen personen in beide groepen samenvat. De volgende afstandsmaten worden daarvoor in de praktijk gebruikt:

- Exact Matching: een nul-afstandsmaat
- Mahalanobis Distance: het numerieke verschil in kenmerken van personen tussen de proef- en referentiegroep, gewogen naar de statistische variantie van elk van de kenmerken
- Propensity Score: de kans op een interventie (kans om deel uit te maken van de proefgroep), verklaard door de waargenomen kenmerken van personen

Bij Exact Matching is de afstandsmaat in feite ook de koppelingsregel. Een persoon in de proefgroep wordt gekoppeld aan een persoon uit de referentiegroep als geldt dat de afstand tussen de kenmerken van beide personen gelijk is aan nul. Op alle relevante kenmerken zijn beide personen dan gelijk, behalve waar het gaat om de inzet van de beleidsinterventie. Bij Exact Matching kan het gebeuren dat er meerdere personen zijn in de referentiegroep waarvoor geldt dat zij exact dezelfde relevante kenmerken hebben als een persoon in de proefgroep. In dit geval wordt er meestal een persoon willekeurig uitgekozen voor deelname aan de controlegroep.

Als het aantal kenmerken toeneemt, wordt de kans op een exacte match kleiner. De kwaliteit van de koppeling hangt daarom af van de relevantie en beschikbaarheid van invloedsfactoren. Bovendien geldt dat de kans op een match kleiner wordt naarmate de kenmerken van een persoon minder gangbaar zijn. In beide gevallen is er sprake van selectieve matching, omdat vooral

personen met gemiddelde kenmerken worden vergeleken in de effectiviteitsanalyse. Als het echter mogelijk is om met Exact Matching voor iedere persoon in de proefgroep een persoon uit de referentiegroep te matchen, bijvoorbeeld omdat de referentiegroep veel groter in omvang is dan de proefgroep, dan is Exact Matching de optimale manier van koppeling.

In de praktijk is een exacte match veelal niet mogelijk en daarom wordt het verschil tussen personen in de proef- en referentiegroep vaak samengevat in een afstandsmaat. Zo drukt de Mahalanobis Distance het verschil tussen personen in de proef- en referentiegroep uit als het numerieke verschil in de relevante invloedsfactoren X . De mate waarin een invloedsfactor meetelt in de Mahalanobis Distance hangt af van de mate waarin die invloedsfactor varieert tussen de proef- en referentiegroep. Echter, hoe meer invloedsfactoren relevant zijn voor de te vergelijken uitkomstmaat, des te ingewikkelder het berekenen van de Mahalanobis Distance wordt. Er is dan sprake van de zogenaamde 'Curse of Dimensionality'. De Propensity Score vat het verschil in alle relevante invloedsfactoren X tussen personen in de proef- en referentiegroep samen in de kans op een beleidsinterventie. De Propensity Score kan worden berekend met behulp van een eenvoudige kansanalyse, bijvoorbeeld een Probit of Logit analyse, waarin de afhankelijke variabele een dummy is die aangeeft of iemand wel of niet in de proefgroep zit. Nadeel van de Propensity Score ten opzichte van de Mahalanobis Distance is dat deze 'afstandsmaat' afhankelijk is van een willekeurig gekozen functionele vorm (bijvoorbeeld Probit of Logit). Groot voordeel is echter dat de Propensity Score het verschil tussen personen in de proef- en controlegroep samenvat tot een ééndimensionale maat en tevens rekening houdt met het relatieve belang van invloedsfactoren op het verschil in samenstelling tussen proef- en referentiegroep. Ten opzichte van Exact Matching heeft het als voordeel dat niet alle kenmerken in een gelijke combinatie hoeven voor te komen in de proef- en referentiegroep.

De vraag is nu wanneer welke afstandsmaat het beste kan worden toegepast. In de bestaande literatuur worden geen standaard criteria genoemd. Toch spelen een drietal factoren een belangrijke rol:

- Het aantal waarnemingen in proef- en referentiegroep;
- Het aantal kenmerken die relevant zijn voor matching;
- Of de relevante kenmerken categoriale of min of meer continue variabelen zijn.

Als alle personen in de proefgroep gekoppeld kunnen worden aan identieke personen uit de referentiegroep dan is dat wenselijk. Naarmate het aantal waarnemingen in de referentiegroep groter wordt ten opzichte van het aantal waarnemingen in de proefgroep, neemt de kans op een exacte match toe. In de praktijk zal de onderzoekspopulatie niet altijd groot genoeg zijn om op basis van alle relevante kenmerken een exacte match te vinden. Ook als bepaalde kenmerken een min of meer continu karakter hebben, zoals het loon of het aantal arbeidsuren, dan wordt het vinden van een exacte match moeilijk zo niet onmogelijk. Koppeling op basis van een Propensity Score wordt aantrekkelijker naarmate het aantal voor matching relevante kenmerken toeneemt. Dit omdat ongeacht het aantal kenmerken waarmee rekening dient te worden gehouden de koppeling altijd plaatsvindt op een eendimensionale variabele. In de praktijk wordt dan ook het meest gebruik gemaakt van de Propensity Score als afstandsmaat.

Aanbeveling

Het gebruik van Exact Matching is alleen aan te raden wanneer een klein deel van een afgebakende populatie te maken krijgt met een interventie. Wanneer slechts een klein deel van bijvoorbeeld de populatie arbeidsgehandicapten met rugklachten afkomstig uit de bouwnijverheid een experimentele rugbehandeling zou krijgen, dan is het wellicht mogelijk om in het deel van dezelfde populatie dat die behandeling niet krijgt vrijwel exact dezelfde werknemers aan te treffen. Krijgt ongeveer de helft van die populatie zo'n rugbehandeling, dan is de kans klein dat er voldoende personen in de andere helft van de populatie zitten met exact dezelfde kenmerken. In

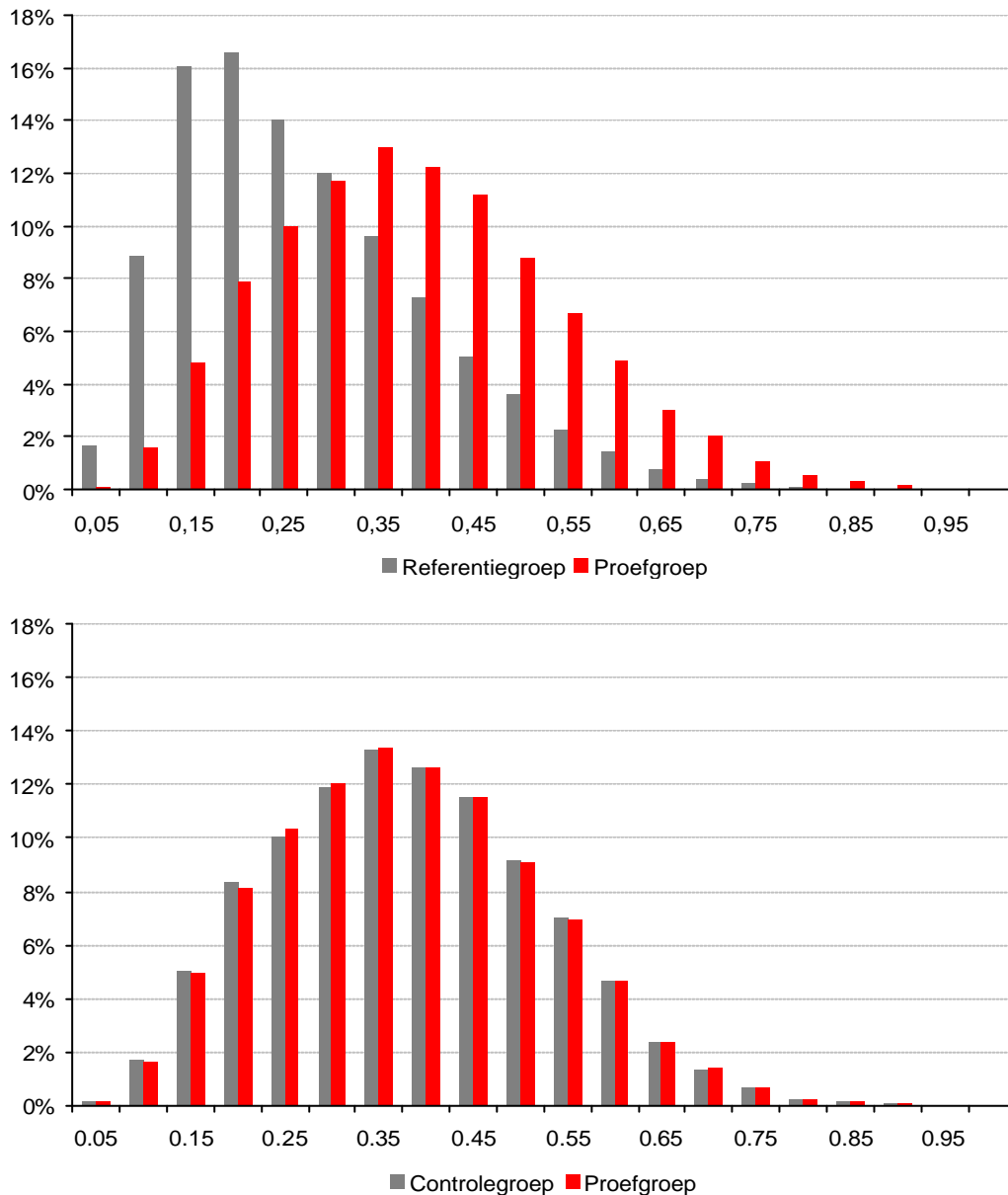
dat geval kan beter gebruik worden gemaakt van een niet-exacte afstandsmaat, waarbij wordt gezocht naar personen in de andere helft van de populatie die ongeveer dezelfde kenmerken hebben. In een afgebakende populatie als die van arbeidsongeschikten met rugklachten afkomstig uit de bouwnijverheid is het aantal kenmerken waarop personen van elkaar kunnen verschillen beperkt, waardoor de Mahalanobis Distance nog betekenisvol kan worden berekend. Als het aantal kenmerken waarop personen van elkaar kunnen verschillen groot is, dan kan het beste gebruik worden gemaakt van de Propensity Score. De Propensity Score berekend voor elke willekeurige persoon in de proef- en controlegroep de kans dat hij of zij te maken krijgt met de beleidsinterventie. Die kans kan worden vastgesteld op basis van alle relevante invloedsfactoren die worden waargenomen, zolang het aantal waarnemingen dit toelaat.

Methode voor koppeling

Met uitzondering van Exact Matching is tot dusver niet gedefinieerd hoe een koppeling tussen personen uit de proef- en referentiegroep tot stand kan komen, gegeven een afstandsmaat. Hieronder worden verschillende koppelingsregels besproken. Bij het concretiseren ervan gaan we uit van de Propensity Score als afstandsmaat. Deze wordt berekend als de kans op een beleidsinterventie in de proef- en controlegroep. Het spreekt voor zich dat deze kans in de proefgroep per definitie hoger ligt dan in de controlegroep, maar ook personen in de controlegroep hadden een positieve kans om de proefgroep terecht te komen.

- *Gestratificeerde Matching*: hierbij wordt de Propensity Score opgedeeld in een aantal intervallen. Dit wordt getoond in het bovenste gedeelte van Figuur 6.1. In het voorbeeld van Figuur 6.1 is de kans dat personen een interventie krijgen opgedeeld in 20 gelijke intervallen. Duidelijk is te zien dat de gemiddelde kans op een interventie hoger ligt in de proefgroep dan in de referentiegroep. Vervolgens worden personen uit de referentiegroep gekoppeld aan personen in de proefgroep die zich in hetzelfde interval bevinden. Het onderste gedeelte van Figuur 6.1 laat zien dat na matching de verdeling van de Propensity Score in de zo gevormde controlegroep hetzelfde is als in de proefgroep. Daardoor hebben we in beide groepen met min of meer dezelfde personen te maken, zodat het effect van de beleidsinterventie kan worden berekend als het verschil in de gemiddelde uitkomstmaat tussen personen in de proefgroep en controlegroep. Dat effect is gelijk aan het verschil in de uitkomstmaat per gekozen interval, gewogen met de verdeling van het aantal proefpersonen over de verschillende intervallen. Deze methode is toegepast in het onderzoek van UWV (2008) naar de netto-effectiviteit van re-integratie activiteiten bij arbeidsgehandicapten.
- *Nearest Neighbour Matching*: hiervoor worden personen in de proefgroep en referentiegroep gerangschikt naar de waarde van hun Propensity Score. Vervolgens wordt voor elke persoon uit de proefgroep een persoon in de referentiegroep gevonden die het minst afwijkt in termen van de Propensity Score. Een risico van deze koppelingsregel is dat er alleen wordt gekeken naar relatieve verschillen. Het is bijvoorbeeld mogelijk dat een 'buurman' uitgekozen wordt die in termen van de Propensity Score op een behoorlijke afstand ligt.

Figuur 6.1: Verdeling propensity score vóór (boven) en na matching (onder)



- *Alternative Neighbourhood Matching (Radius Matching)*: ook hier worden personen uit de proef- en referentiegroep gerangschikt naar de waarde van hun Propensity Score, maar nu wordt er geen persoon uit de referentiegroep gekoppeld aan een persoon in de proefgroep met de kleinste Propensity Score afwijking, maar wordt een gemiddelde van de uitkomstmaat van alle personen in de referentiegroep die het meest lijken op een persoon uit de proefgroep gekoppeld aan de waarde van de uitkomstmaat van deze persoon in de proefgroep. De betreffende personen uit de referentiegroep mogen daarbij niet meer dan een bepaalde hoeveelheid afwijken in hun Propensity Score. Deze koppelingsregel veronderstelt dat een gemiddelde waarde van de uitkomstmaat van meer dan één persoon uit de referentiegroep een betere vergelijking vormt dan de uitkomstmaat van één persoon uit de referentiegroep die het meest op de proefpersoon lijkt. Een mogelijk nadeel van deze regel is dat het op voorhand niet duidelijk is wat de afwijking van de Propensity Score zou mogen zijn.
- *Kernel Matching*: ook hier wordt de waarde van de uitkomstmaat van personen in de proefgroep gekoppeld aan een gemiddelde van de uitkomstmaat van meerdere personen

in de referentiegroep. Alleen betreft het hier een gemiddelde over *alle* personen in de referentiegroep, gewogen met een waarde die omgekeerd evenredig is met het verschil in Propensity Score. Dat betekent dat de uitkomstmaat van personen in de referentiegroep met een kleine afwijking in de Propensity Score sterker meetellen dan personen in de referentiegroep met een grote afwijking in de Propensity Score. Omdat de uitkomstmaten van alle personen in de referentiegroep in principe een kans hebben om ‘mee te tellen’ voor koppeling aan de uitkomstmaat van elke persoon in de proefgroep, resulteert Kernel Matching in een koppeling waarbij de uitkomstmaat een relatief grote variatie kent tussen personen. Dit heeft tot gevolg dat het effect van de beleidsinterventie relatief nauwkeurig (efficiënt) kan worden geschat. De mate waarin personen in de referentiegroep meetellen kan worden bepaald door een verdelingsfunctie met een bandbreedte waarover die kansverdeling wordt gebruikt voor koppeling. In de praktijk wordt gebruik gemaakt van de Gaussian, Biweight, Epanechnikov, Uniforme en Tricube verdelingsfuncties. Efficiëntie van de effectschatter wordt nauwelijks beïnvloed door het type kernelfunctie (Dehejia en Wahba, 1999)³⁰, wel door de bandbreedte die wordt gekozen. Uit de literatuur blijkt dat er geen standaard regels zijn over wat deze bandbreedte nu precies zou moeten zijn. In Software programma’s waarin matching als evaluatiemethode wordt aangeboden is de bandbreedte voorgeprogrammeerd, maar kan vaak door de gebruiker worden aangepast.

- *Mahalanobis Matching met Propensity Score*: deze manier van koppelen is gelijk aan de eerder besproken Nearest Neighbour Matching, maar in deze koppelingsmethode is de afstandsmaat de Mahalanobis Distance waarbij de Propensity Score één van de koppelingsvariabelen is.

Aanbeveling

Over de kwaliteit versus de eenvoud van koppeling geeft de literatuur geen eenduidig antwoord. Toch kunnen enkele aanbevelingen worden gedaan ten aanzien van de te hanteren koppelingsregel. Zo werd eerder al gesteld dat exact matching alleen in aanmerking komt wanneer iedere persoon in de proefgroep gekoppeld kan worden aan een persoon in de referentiegroep op basis van een beperkt aantal koppelingskenmerken, waarbij min of meer continue variabelen zo veel mogelijk zijn uitgesloten. Van de koppelingsregels op basis van een afstandsmaat zorgt Kernel Matching voor de meest efficiënte schatter van het effect van een interventie, omdat het de meeste variatie toelaat in de samenstelling van de controlegroep. Nearest Neighbour Matching en Gestratificeerde Matching zijn het meest eenvoudig maar ook het minst nauwkeurig. Bij de laatste methode wordt het aantal en de omvang van de intervallen altijd min of meer willekeurig gekozen. Wanneer matching wordt gebruikt voorafgaand aan een andere evaluatiemethode, zoals een duuranalyse, dan kan alleen gebruik worden gemaakt van koppelingsregels waarbij *personen* uit de referentiegroep worden gekoppeld aan *personen* uit de proefgroep. Koppelingsregels die *uitkomstmaten* uit de referentiegroep koppelen aan een *uitkomstmaat* in de proefgroep, zoals Alternative Neighbourhood Matching en Kernel Matching, kunnen in dat geval niet worden gebruikt.

6.2.2 Matching op duurvariabelen

Kenmerken waarop koppeling plaatsvindt zijn over het algemeen cross-sectie gegevens die gelden op een bepaald meetmoment. Voorbeelden zijn geslacht, leeftijd bij aanvang van werkloosheid en opleidingsniveau. Toch kan ook op basis van variabelen die variëren in de tijd worden gematched, zelfs als ze een relatie hebben met de uitkomstmaat. Hier kunnen we weer refereren aan het voorbeeld waarin we op zoek zijn naar het effect van re-integratietrajecten op de werkloosheidsduur van werklozen. De proefgroep met een re-integratietraject zal niet alleen verschillen van de referentiegroep zonder een re-integratietraject op basis van kenmerken op één

³⁰ Zie ook <http://www.quantlet.com/mdstat/scripts/spm/html/spmhtmlnode16.html>.

bepaald moment, maar ook naar de lengte van de periode gedurende welke personen worden waargenomen. Iemand die wat langer wordt waargenomen zal eerder een re-integratietraject hebben ontvangen, omdat het immers tijd vergt om met een traject te starten. Had deze persoon al eerder werk gevonden, dan kwam hij of zij met een korte werkloosheidsduur terecht in de referentiegroep. Door de snelle werkhervatting komt deze persoon niet in aanmerking voor een re-integratietraject en blijft dus buiten de proefgroep. Een willekeurig persoon in de proefgroep kan dus niet worden vergeleken met een willekeurig persoon in de referentiegroep, ook niet als alle overige kenmerken gelijk zijn. De lengte van de waargenomen zoekduur zegt al iets over de kans om een re-integratietraject te krijgen.

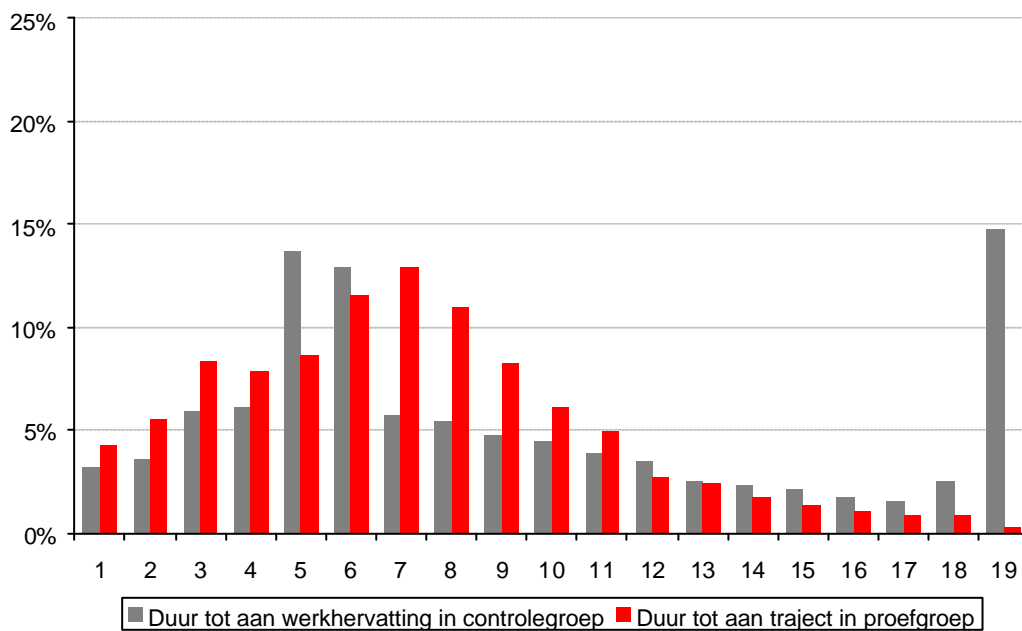
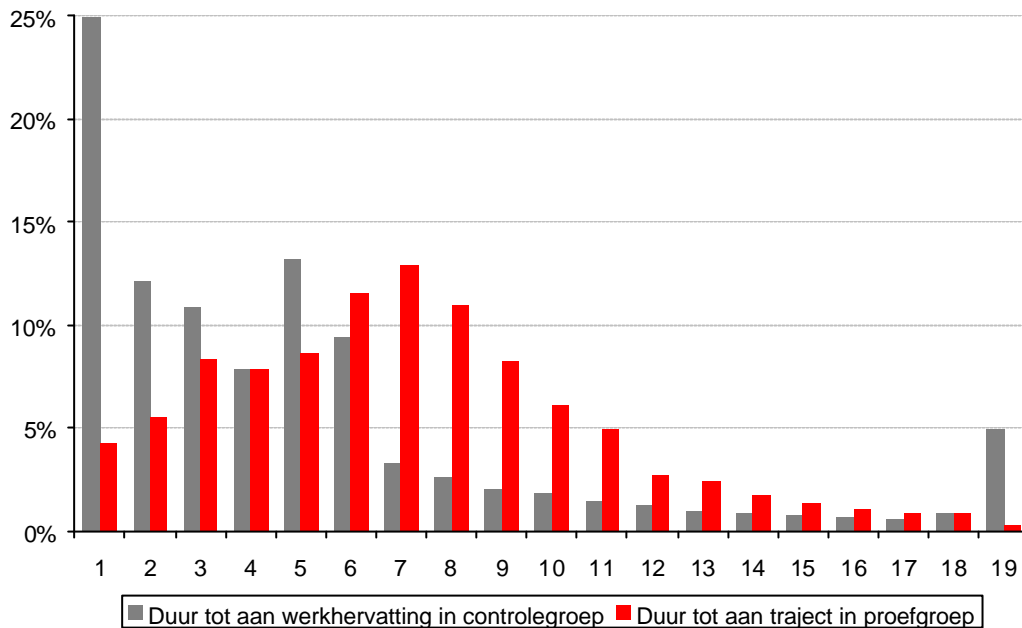
Dit wordt geïllustreerd aan de hand van Figuur 6.2. In het bovenste gedeelte van de figuur wordt de werkloosheidsduur tot aan werkhervatting in de referentiegroep vergeleken met de werkloosheidsduur tot aan de inzet van een re-integratietraject in de proefgroep. Wat opvalt is dat er veel personen in de referentiegroep relatief snel het werk hervatten, terwijl de meeste personen in de proefgroep dan nog een traject moet krijgen. Dit kan onder andere liggen aan selectieve inzet van re-integratietrajecten bij personen die het minst gemakkelijk werk vinden. Maar los daarvan wordt het verschil in het bovenste gedeelte van Figuur 6.2 veroorzaakt door de conditionele volgtijdelijkheid van de interventie en werkhervatting. Wanneer personen in de controlegroep bij toeval langer hadden gedaan over hun zoektocht naar werk, zouden ze een grotere kans hebben gehad om een re-integratietraject te krijgen en daarmee in de proefgroep terecht te komen. Als gevolg daarvan kan iemand die na een half jaar werkloosheid start met een re-integratietraject (proefgroep) niet worden vergeleken met iemand die na vijf maanden het werk heeft hervat. Wel met iemand die na 7 maanden of pas na 19 maanden het werk heeft hervat.

In ons voorbeeld is de uitkomstmaat de werkloosheidsduur. Omdat dit de afhankelijke variabele is die we juist willen verklaren, kunnen we hier niet op matchen. Waar we op zouden willen matchen is de duur tot de inzet van een re-integratietraject. Personen in de referentiegroep hebben echter geen startdatum voor een re-integratietraject waarop kan worden gematched. Wel hebben ze een waargenomen werkloosheidsduur. Voor een goede vergelijkbaarheid van proef- en controlegroep mag deze waargenomen werkloosheidsduur in ieder geval niet korter zijn dan de duur tot inzet van een re-integratietraject van de persoon in de proefgroep waarmee wordt gekoppeld. Iemand in de proefgroep die op tijdstip 10 een interventie ontvangt kan gekoppeld worden aan iedereen in de referentiegroep die niet eerder dan tijdstip 10 het werk heeft hervat. Of andersom gesteld, iemand in de referentiegroep die het werk hervat op tijdstip 10 kan worden gekoppeld aan iedereen in de proefgroep die vóór tijdstip 10 een interventie heeft ontvangen. Verder is er geen reden te veronderstellen dat het patroon van werkhervatting door de tijd heen anders is tussen proef- en referentiegroep.

Met dit uitgangspunt kan de volgende koppelingsprocedure worden opgesteld, zoals toegepast in UWV (2008), gebaseerd op gestratificeerde matching:

- De tijdsduur wordt opgedeeld in een aantal intervallen (19 intervallen in Figuur 6.2);
- Voor de proefgroep wordt het cumulatieve aantal waarnemingen per interval vastgesteld. Voor het n-de interval zijn dat alle personen die tot en met het n-de interval starten met een interventie;
- Vervolgens wordt voor de proefgroep het cumulatieve aantal waarnemingen uitgedrukt als aandeel van de som van het cumulatieve aantal waarnemingen over alle intervallen;
- Deze aandelen worden gedeeld door het maximale aandeel. Daarmee wordt het aandeel in het interval met het maximale aandeel (i.c. het laatste interval) gelijk gesteld aan 1 en worden de aandelen in alle andere intervallen kleiner dan 1;
- Dit geschaalde aandeel wordt gebruikt om een willekeurig aandeel waarnemingen uit de referentiegroep in hetzelfde interval te kiezen voor de controlegroep, waarbij de intervallen in de referentiegroep zijn gebaseerd op de duur tot werkhervatting.

Figuur 6.2: Verdeling werkloosheidsduur over duurperioden vóór en na matching



Toepassing van deze methode leidt tot een controlegroep waarin het aandeel van snelle werkherlevingen wordt beperkt ten opzichte van de referentiegroep, zoals kan worden vastgesteld in het onderste gedeelte van Figuur 6.2. Het idee achter deze aanpak is dat alle personen in de proefgroep gematched kunnen worden met alle personen in het laatste interval van de referentiegroep (interval 19 in Figuur 6.2). Immers, voor alle personen in de proefgroep geldt immers dat ze een kans hebben om pas in het laatste onderscheiden tijdsinterval het werk te hervatten. Hoe eerder het interval, hoe minder personen in de proefgroep gematched kunnen worden met personen in de referentiegroep, maar in ieder geval wel alle personen die tot en met dat interval in de proefgroep zitten. Door het cumulatieve aandeel per interval te nemen in de proefgroep, krijgen personen in de proefgroep de mogelijkheid om dezelfde verdeling van werkherleving in de tijd aan te nemen als die in de referentiegroep.

Deze methode kan worden toegepast ongeacht de grootte van de proef- en referentiegroep en ongeacht de verdeling van waarnemingen over de tijd in beide groepen. De methode leidt alleen wel tot een inperking van de referentiegroep, omdat alleen uit de laatste periode alle personen uit de referentiegroep worden gekozen, in de perioden daarvoor alleen een aandeel. Het komt erop neer dat de controlegroep relatief klein wordt in het geval de referentiegroep niet veel groter is dan de proefgroep of wanneer de interventie relatief laat wordt ingezet.

Omdat matching op cross-sectie gegevens nauwkeuriger is dan op duurvariabelen, wordt aangeraden om de matching op duurvariabelen plaats te laten vinden voorafgaand aan de matching op cross-sectie gegevens.

6.3 Duurmodellen

In deze paragraaf komen enkele specificaties van duurmodellen aan de orde. Het betreft hier de specificatie van duurzaamheidsduur, de keuze tussen en specificatie van single hazards versus competing risks, de keuze tussen de modellering van eenmalige duurperioden en meerdere duurperioden per persoon, het al dan niet rekening houden met niet-waargenomen heterogeniteit en ten slotte de specificatie van niet-waargenomen heterogeniteit.

6.3.1 Duurafhankelijkheid

Duurafhankelijkheid is een verzamelnaam voor de tijdsvariërende invloed die allerlei factoren hebben op de duur als uitkomstmaat, zoals de werkloosheidsduur, baanduur of duur tot aan de start van een interventie. Dit betreft zowel invloeden die sterker of zwakker worden in de tijd, zoals het slijten van kennis en vaardigheden door werkloosheid, als invloeden die alleen op een bepaald moment gelden, zoals een tijdelijke sanctie of de overgang van een loongerelateerde uitkering naar een vervolguitkering bij WW-gerechtigden.

Neem bijvoorbeeld de kans op werkherhervatting vanuit werkloosheid. Iemand die pas werkloos is zal niet onmiddellijk weer aan de slag kunnen. Het vinden van een geschikte baan heeft tijd nodig. In het begin van de werkloosheid geldt dat hoe langer deze persoon zoekt, hoe groter de kans om een baan te vinden. De kans op werkherhervatting stijgt dus in eerste instantie naarmate de werkloosheid langer duurt. Tegelijkertijd geldt echter dat een te lange werkloosheidsduur de kansen op werkherhervatting doet afnemen. Immers, kennis en vaardigheden van werkzoekende verouderen en werklozen verliezen ook steeds meer het contact met de arbeidsmarkt, inclusief alle daarmee samenhangende sociale netwerken. Bovendien kan een langere werkloosheidsduur demotiverend en zelfs stigmatiserend werken, zodat werkzoekenden minder hard naar werk zoeken en werkgevers minder geneigd zijn om werkzoekenden aan te nemen. Op termijn neemt de kans op werkherhervatting dus af naarmate de werkloosheid langer duurt. Daardoor hebben twee identieke personen die op verschillende momenten een re-integratietraject ontvangen bij aanvang van het re-integratietraject niet dezelfde verwachte werkloosheidsduur, los van het effect van het re-integratietraject zelf.

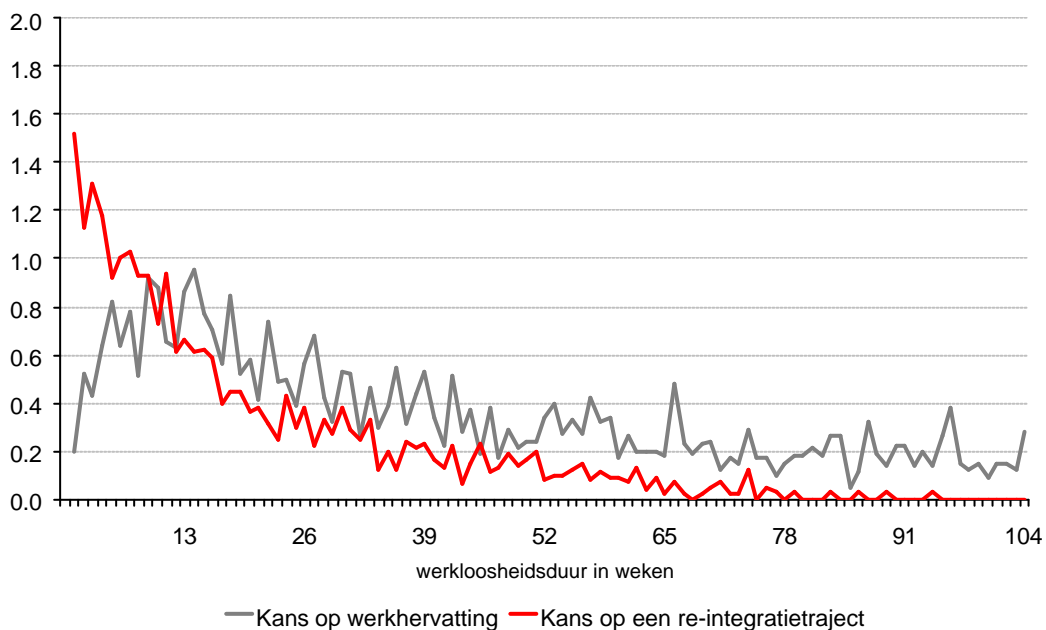
De manier waarop met die verschillen in duurzaamheidsduur rekening kan worden gehouden, is het onderwerp van deze paragraaf. Voor de specificatie van duurzaamheidsduur in een duurmodel is het goed om deze eerst in beeld te brengen, bijvoorbeeld met behulp van een Kaplan-Meier schatter (zie Figuur 6.3). De Kaplan-Meier schatter berekent voor elk tijdstip t (uitgedrukt in bijvoorbeeld een dag, week, maand of jaar) hoeveel werkherhervattingen er op dat tijdstip plaatsvinden ten opzichte van het aantal personen dat tot en met het tijdstip t nog de mogelijkheid hadden om het werk te hervatten (werkloos en op zoek naar werk). Hierbij wordt (nog) niet gecorrigeerd voor de invloed van andere verklarende kenmerken. De Kaplan-Meier schatter kan worden uitgedrukt in een conditionele kans per tijdstip. In Figuur 6.3 staan twee voorbeelden van zulke conditionele kansen voor een willekeurige groep

werklozen. In het voorbeeld stijgt de kans op werkherhvatting gedurende de eerste weken van de werkloosheid om vervolgens langzaam af te nemen. De kans dat deze groep werklozen een re-integratietraject krijgen is in het begin vrij groot, om daarna snel af te nemen.

Op basis van Figuur 6.3 kan een specificatie worden gekozen voor de duurzaamheidsafhankelijkheid in het te schatten duurmodel. Daarvoor zijn de volgende mogelijkheden:

- *Monotoon stijgende of dalende duurzaamheidsafhankelijkheid*: bijvoorbeeld exponentieel, inverse of dubbel exponentieel / Weibull. De Weibull specificatie vormt in Figuur 6.4 een goede benadering van de duurzaamheidsafhankelijkheid in de kans op een re-integratietraject. Toepassing van een monotoon stijgende of dalende duurzaamheidsafhankelijkheid voor de kans op werkherhvatting zou echter leiden tot een overschatting van werkherhvatting in de eerste weken en een (compenserende) onderschatting daarna.
- *Piecewise constant of niet-parametrische duurzaamheidsafhankelijkheid*: de duurperiode wordt opgedeeld in korte intervallen, elk waarvoor een aparte (constante) kans wordt geschat. Dit is het geval in Figuur 6.5 voor de kans op werkherhvatting. Hier zijn vaste intervallen genomen van telkens vier weken. Op die manier kan ook niet-monotone duurzaamheidsafhankelijkheid goed worden geschat. De optimale indeling in intervallen kan empirisch worden bepaald. Een indicatie van een goede indeling kan visueel worden afgeleid uit de Kaplan-Meier schatter.

Figuur 6.3: Voorbeeld van duurzaamheidsafhankelijkheid op basis van een Kaplan-Meier schatter



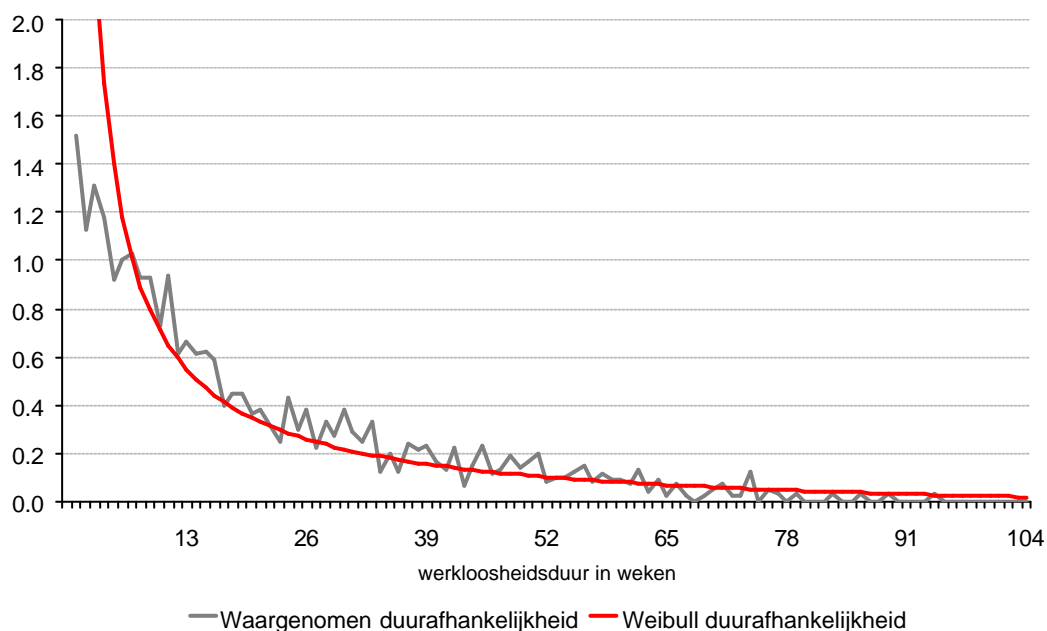
Voordeel van de piecewise constant specificatie van duurzaamheidsafhankelijkheid is dat alle variatie in duurzaamheidsafhankelijkheid kan worden opgepikt door het duurmodel, dus ook incidentele veranderingen als gevolg van instituties (bijvoorbeeld de opschoning van bestanden aan het eind van het jaar, waardoor personen die inmiddels niet meer op zoek zijn naar werk ook niet langer als werkloos worden geregistreerd). Nadeel is dat het de analyse complexer maakt ten opzichte van parametrische functies van duurzaamheidsafhankelijkheid. Ook ligt er geen duidelijk theoretische basis ten grondslag aan een niet-parametrische specificatie, waardoor de keuze van intervallen min of meer willekeurig is en de interpretatie van gevonden duurzaamheidsafhankelijkheid minder eenduidig.

Voor de meting van de netto-effectiviteit van re-integratietrajecten bij WW-gerechtigden gebruiken Heul, Berendsen, van der Eijken en Vlek (2003) een Weibull specificatie voor de duurzaamheid, terwijl Heyma (2005) in een soortgelijke studie uitgaat van een piecewise constant specificatie. Heyma (2005) laat daarbij zien wat de gevolgen zijn van het verschil in de specificatie van duurzaamheid tussen deze beide studies.

Aanbeveling

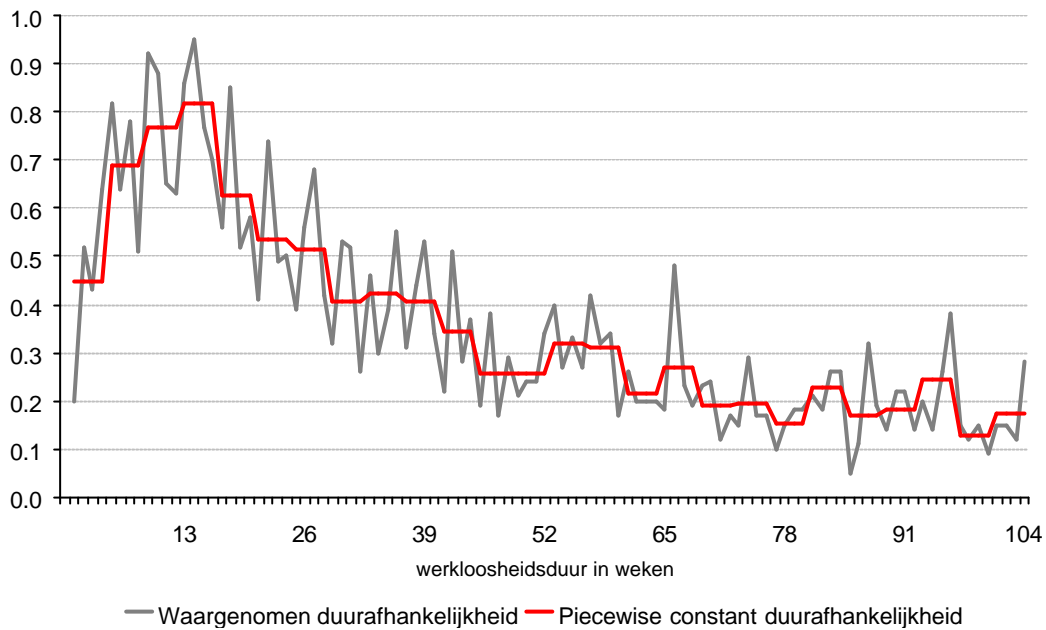
Welke specificatie van duurzaamheid kan in welke situatie het beste worden gekozen? Als de duurzaamheid zoals geïllustreerd door de Kaplan-Meier schatter redelijkerwijze kan worden benaderd door een parametrische functie, dan is zo'n parametrische functie gezien de eenvoud te prefereren. Parametrische duurzaamheid maakt vaak ook deel uit van standaard statistische software. Figuur 6.4 laat een situatie zien waarin een parametrische Weibull functie een goede benadering is van de werkelijke duurzaamheid. Werkhervatting van werklozen met een geringe afstand tot arbeid heeft een verdeling over tijd zoals in Figuur 6.4. Meteen vanaf de start van de werkloosheidsperiode is de kans op een nieuwe baan voor deze groep groot, maar die kans neemt af naarmate men langer werkloos is.

Figuur 6.4: Duurzaamheid benaderd door een Weibull specificatie



Bij werklozen met een relatief grote afstand tot arbeid is het veel moeilijker om snel een baan te vinden. Voor deze groep is de kans op werkherhvatting in het begin van de werkloosheid gering, maar stijgt als men langer de kans heeft gekregen om zich op de arbeidsmarkt te oriënteren. Als het vinden van een baan binnen een zekere periode niet is gelukt, daalt vervolgens ook in deze groep de kans om alsnog een baan te vinden. Deze situatie wordt getoond in Figuur 6.5. Varieert de duurzaamheid sterk over tijd (niet monotoon stijgend of dalend), dan wordt een piecewise constant benadering sterk aanbevolen.

Figuur 6.5: Duurafhankelijkheid benaderd door een Piecewise Constant specificatie



6.3.2 Single hazard of competing risks

Vaak is de uitkomstmaat slechts een deel van het volledige proces waar personen mee te maken hebben. Zo kunnen werklozen het werk hervatten, maar de werkloosheid ook beëindigen door zich terug te trekken van de arbeidsmarkt. In een duurmodel dat alleen de werkloosheidsduur tot aan werkherhvatting modelleert (single hazard), wordt terugtrekking van de arbeidsmarkt gezien als een afgekapte (gecensureerde) waarneming, waarbij een persoon niet langer wordt waargenomen. Terugtrekking van de arbeidsmarkt is echter niet hetzelfde als niet langer als werkloos te worden waargenomen. In een single hazard duurmodel wordt namelijk verondersteld dat wanneer een afgekapte waarneming nog wel zou zijn waargenomen, deze persoon een kans op werkherhvatting zou hebben die gelijk is aan soortgelijke werklozen in de populatie. Maar als deze persoon zich in werkelijkheid heeft teruggetrokken van de arbeidsmarkt, dan is die kans op werkherhvatting uiteraard een stuk kleiner. Als hier geen rekening mee wordt gehouden in het model, leidt dat tot een onzuivere schatting van de kans op werkherhvatting in zowel de proef- als controlegroep. Als de kans op terugtrekking van de arbeidsmarkt bovendien wordt beïnvloed door het al dan niet ontvangen van een beleidsinterventie, wordt bij zo'n modellering bovendien de effectiviteit van de interventie onzuiver ingeschat. Die onzuiverheid is klein wanneer slechts een zeer beperkt gedeelte van de populatie zich terugtrekt van de arbeidsmarkt, zoals bijvoorbeeld bij de WW-populatie. Maar wanneer dat gedeelte groter is, zoals bij de WWB-populatie, dan is die onzuiverheid aanzienlijk, zie bijvoorbeeld Graaf-Zijl, Heyma en de Hoop (2006).

Om rekening te houden met de mogelijkheid dat mensen de werkloosheidsduur behalve via werkherhvatting ook kunnen beëindigen via een andere route (zoals terugtrekking van de arbeidsmarkt, vrijwilligerswerk of gesubsidieerde arbeid in plaats van reguliere arbeid), kan gebruik worden gemaakt van een zogenaamd competing risks duurmodel. Daarin wordt de gelijktijdige kans op uitstroom via verschillende routes gemodelleerd. Basisidee is dat een persoon die op tijdstip t op manier A de werkloosheidsduur heeft beëindigd, in ieder geval tot tijdstip t de werkloosheidsduur niet op manier B of C heeft beëindigd. Voor deze persoon is de kans op route A dus relatief groot en de kans op route B en C kleiner. In elk van de uitstroomkansen kan rekening worden gehouden met een effect van beleidsinterventies. Dat effect kan bovendien verschillen per route.

Aanbeveling

Gebruik van een competing risks duurmodel levert meer complexiteit op dan het gebruik van een single hazard duurmodel. Een competing risks duurmodel dient daarom alleen te worden gebruikt wanneer de situatie daarom vraagt. Daarvoor moet voldaan zijn aan de volgende twee voorwaarden:

- Er moet sprake zijn van concurrerende uitkomsten ten opzichte van de te meten uitkomstmaat. Als de uitkomstmaat werkhervatting is, dan is terugtrekking van de arbeidsmarkt een concurrerend risico. Wanneer de uitkomstmaat beëindiging van de werkloosheidsuitkering is, dan is er geen concurrerend risico;
- Concurrerende uitkomsten dienen in voldoende mate voor te komen in de populatie. Als slechts 1 procent van de populatie zich binnen de waarnemingsperiode terugtrekt van de arbeidsmarkt, terwijl 80 procent in die tijd het werk hervat, dan is er nauwelijks sprake van een concurrerend risico. Behalve de geringe vertekening van het te meten effect, zullen er in dat geval (relatief) weinig waarnemingen zijn om het concurrerende risico goed in te schatten.

6.3.3 Eenmalige duurperioden versus meervoudige duurperioden

Wanneer op meerdere momenten in de tijd de arbeidsmarktsituatie van personen wordt gemeten, dan zal blijken dat sommige personen in de tussentijd niet of nauwelijks van baan zijn veranderd en andere personen verscheidene keren. Binnen een afgebakende onderzoeksperiode is het dus mogelijk om bijvoorbeeld meerdere werkloosheidsperioden per persoon tegen te komen. Die werkloosheidsperioden zijn gemiddeld korter dan de eenmalige werkloosheidsperiode die we voor een ander deel van de onderzoekspopulatie in dezelfde periode waarnemen. Zo zullen personen met een korte afstand tot arbeid minder vaak werkloos zijn dan personen met een wat grotere afstand tot arbeid of seizoensarbeiders. De vraag is nu of en wanneer die meerdere perioden per persoon binnen de onderzoeksperiode moeten worden gebruikt in effectiviteitsanalyse.

Voordeel van het gebruik van meerdere duurperioden per persoon is dat de invloed kan worden gemeten van tijdsinvariante kenmerken van personen, die niet noodzakelijk hoeven te worden waargenomen. Zie daarom ook de volgende paragraaf over niet-waargenomen heterogeniteit. Er wordt in zo'n geval optimaal gebruik gemaakt van de aanwezige informatie over personen.

Nadeel van het gebruik van meerdere duurperioden per persoon is dat niet alle personen in de onderzoekspopulatie een even grote bijdrage leveren aan de verklaring van de uitkomstmaat. Sommige personen kunnen met veel duurperioden een sterke invloed uitoefenen op deze verklaring. Dat is vooral nadelig als deze personen in een bepaald opzicht afwijken van het gemiddelde in de populatie. Nu is het zo dat gegeven de gelimiteerde duur van de onderzoeksperiode de kans groter is dat personen met meerdere duurperioden ook kortere duurperioden hebben. Dus bijvoorbeeld werklozen met korte werkloosheidsduren (o.a. seizoenswerklozen) en werklozen met een goede positie op de arbeidsmarkt. Dat leidt tot een selectie in de onderzoekspopulatie, waarbij het effect van een interventie sterker zou worden bepaald door personen met een korte werkloosheidsduur. Langdurig werklozen zouden bij zo'n opzet een minder dan evenredige invloed uitoefenen op de analyseresultaten, omdat ze binnen de afgebakende onderzoeksperiode minder kans hebben om met meerdere duurperioden te worden waargenomen. Dit geldt vooral als de onderzoeksperiode relatief kort is ten opzichte van de werkloosheidsduur.

Aanbeveling

Aanbevolen wordt om de duuranalyse zo veel mogelijk te baseren op één (willekeurig gekozen) duurperiode per persoon (single spell). Alleen wanneer via meerdere duurperioden in het model wordt gecorrigeerd voor de invloed van niet-waargenomen heterogeniteit, kunnen meerdere

duurperiodes worden gebruikt, mits de onderzoeksperiode lang is in vergelijking tot de gemiddelde duur als uitkomstmaat.

6.3.4 Correctie voor niet-waargenomen heterogeniteit

Een belangrijke functie van evaluatiemethoden is dat ze op één of andere manier rekening houden met verschillen in uitkomsten die het gevolg zijn van samenstellingseffecten en selectie-effecten. Maar invloeden op de uitkomst die niet kunnen worden waargenomen, kunnen ook niet in het model worden opgenomen. Daardoor blijft soms het idee bestaan dat ondanks de toepassing van de juiste evaluatiemethode er nog niet voldoende is gecorrigeerd voor verschillen tussen de proef- en controlegroep. Als voorbeeld kunnen we de motivatie van werklozen noemen om een baan te vinden. Wanneer die groter is bestaat de kans dat ook deelname aan een re-integratietraject groter is. Als we vinden dat deelnemers aan re-integratietrajecten eerder werk vinden dan werklozen die niet deelnemen aan een traject, dan is het de vraag of dat aan de werking van de interventie ligt of aan het verschil in motivatie tussen deelnemers en niet-deelnemers. Probleem is vaak dat het verschil in motivatie niet wordt of kan worden gemeten. Iets soortgelijks geldt voor tal van andere factoren, zoals de inzet van interventies in het verleden of het hebben van ‘zachte’ competenties, zoals sociale vaardigheden of specifieke kennis die niet wordt geregistreerd in gegevensbestanden. Deze niet-waargenomen kenmerken leiden tot zogenaamde heterogeniteit in de uitkomstmaat waarvan de oorzaak niet wordt waargenomen. Als die niet-waargenomen heterogeniteit verschilt tussen de proef- en controlegroep, dan is sprake van een selectieve inzet van interventies op basis van niet-waargenomen invloedsfactoren. Wanneer hier niet op een of ander manier voor wordt gecorrigeerd, dan wordt het verschil tussen de proef- en controlegroep doorgegeven in het daardoor onzuiver gemeten effect van de interventie.

Niet-waargenomen heterogeniteit kan worden geïdentificeerd zolang deze niet varieert in de tijd. Voor die identificatie is juist variatie in de tijd nodig om tijdsvariërende invloedsfactoren daarvan te kunnen onderscheiden. Dat komt neer op de noodzaak van longitudinale gegevens. Omdat duurmodellen slechts kunnen worden toegepast als er longitudinale gegevens bekend zijn, kunnen duurmodellen vaak ook corrigeren voor niet-waargenomen heterogeniteit. Er zijn in principe twee manieren waarop dat kan:

- Wanneer er meerdere werkloosheidsperiodes voor dezelfde personen bekend zijn, dan kan er een tijdsconstante factor op persoonsniveau worden geïdentificeerd. Verschil in die tijdsconstante factor tussen personen is de gezochte niet-waargenomen heterogeniteit, waarvan de oorzaak overigens onbekend blijft, omdat die niet wordt waargenomen;
- Wanneer niet-waargenomen heterogeniteit zorgt voor verschillen in de uitkomstmaat én bovendien verschilt tussen de proefgroep en de controlegroep (selectie-effecten), dan bestaat er correlatie tussen de uitkomstmaat en de kans op een interventie. Wanneer de uitkomst simultaan wordt gemodelleerd met de kans op een interventie, dan kan er rekening worden gehouden met die correlatie. In de timing-of-events benadering kan bijvoorbeeld de werkloosheidsduur en de duur tot de interventie simultaan worden verklaard uit invloedsfactoren, waarbij de constante in beide vergelijkingen aan elkaar worden gerelateerd. Door de identificatie van die correlatie (ofwel de selectieve inzet van interventies op basis van niet-waargenomen invloedsfactoren), wordt het effect van de interventie op de werkloosheidsduur zuiver(der) vastgesteld.

Niet-waargenomen heterogeniteit is voor niet-onderzoekers een abstract en moeilijk begrip. Het kunnen identificeren van de invloed van niet-waargenomen heterogeniteit vertelt nog niets over de oorzaak ervan. Belangrijk voor de evaluatie van de netto-effectiviteit van interventies is dat niet-waargenomen factoren die samenhangen met de inzet van een interventie het gemeten effect aanzienlijk kunnen beïnvloeden, waardoor onterecht de verkeerde werking wordt toegeschreven aan het beleid. Om dat te voorkomen is het belangrijk om rekening te houden met de

mogelijkheid van niet-waargenomen heterogeniteit. Dat wordt belangrijker naarmate er minder invloedsfactoren worden waargenomen die een verklaring vormen voor de waargenomen heterogeniteit in de uitkomstmaat.

Aanbeveling

Als er een vermoeden is dat niet waargenomen factoren de uitkomstmaat beïnvloeden en deze factoren niet al voldoende gedekt worden door wel waargenomen kenmerken van de onderzoekspopulatie of daarmee voldoende zijn gecorreleerd, dan wordt aanbevolen om in de analyse te corrigeren voor niet-waargenomen heterogeniteit. Dekking door wel waargenomen kenmerken is groter, wanneer er meer verschillende kenmerken worden waargenomen en naarmate er meer subjectieve kenmerken worden waargenomen, zoals motivatie en houding.

6.3.5 Specificatie van niet-waargenomen heterogeniteit

Aan de identificatie van niet-waargenomen heterogeniteit in een duurmodel ligt de veronderstelling ten grondslag dat de invloed ervan is verdeeld over de onderzoekspopulatie volgens een onbekende functie die in de analyse moet worden benaderd. Wanneer niet-waargenomen heterogeniteit wordt geïdentificeerd op basis van meerdere duurperioden per persoon, wordt vaak gebruik gemaakt van een parametrische verdelingsfunctie die goed past bij de specificatie van het duurmodel. Voorbeelden hiervan zijn de Exponentiële verdeling en de Gamma verdeling. Identificatie vindt dan grotendeels plaats op basis van de functionele vorm van de veronderstelde verdeling van niet-waargenomen heterogeniteit, waardoor de modelschatting wordt uitgebreid met enkele aanvullende modelparameters. Dit is een vrij eenvoudige benadering die in de meeste statistische softwarepakketten wordt ondersteund, maar biedt weinig flexibiliteit voor een juiste identificatie.

Bij identificatie van niet-waargenomen heterogeniteit op basis van correlatie tussen de duur tot aan de inzet van een interventie en de duur als uitkomstmaat, wordt meestal gebruik gemaakt van een niet-parametrische verdeling van die correlatie. De bivariate verdeling van niet-waargenomen heterogeniteit in beide duurprocessen wordt dan benaderd door een beperkt aantal massapunten, waarvan de locatie en de kans worden vastgesteld door middel van het duurmodel. Toepassingen hiervan kunnen worden gevonden in Heyma (2005) en Heyma en Zijl (2005).

Aanbeveling

Omdat de enige reden van de identificatie van niet-waargenomen heterogeniteit een zuivere schatting van het effect van een beleidsinterventie is, dient bij voorkeur te worden vastgesteld of er een systematische correlatie bestaat tussen de inzet van een interventie en de uitkomstmaat na correctie voor waargenomen kenmerken. Dit kan door het simultaan schatten van een duurmodel en de kans op een interventie, waarbij de correlatie tussen beide processen wordt geïdentificeerd op basis van een beperkt aantal massapunten. Dit is vooral een goede aanpak wanneer er veel variatie bestaat in het tijdstip waarop een interventie plaatsvindt. Wanneer er weinig variatie zit in het tijdstip waarop interventies plaatsvinden en er bovendien veel meervoudige duurperioden per persoon worden waargenomen, dan kan worden gekozen voor een parametrische benadering van de verdeling van niet-waargenomen heterogeniteit. Deze laatste benadering is eenvoudiger uit te voeren dan de eerste en wordt vaker ondersteund door standaard statistische software.

7 Gegevens, bewerking en presentatie

In dit hoofdstuk wordt een overzicht gegeven van gegevens die nodig zijn voor het uitvoeren van effectiviteitsanalyses, de manier waarop deze gegevens kunnen worden bewerkt en manieren waarop de resultaten van een effectiviteitsanalyse kunnen worden weergegeven.

7.1 Benodigde gegevens voor effectiviteitsanalyse

Om de verschillende evaluatiemethoden toe te kunnen passen, zijn gegevens nodig over de uitkomstmaat, de beleidsinterventie waarvan het effect dient te worden vastgesteld en overige factoren, voorzover die verschillen tussen proef- en controlegroep en tevens de uitkomstmaat kunnen beïnvloeden. Dit zijn bijvoorbeeld kenmerken van personen of groepen die het betreft, kenmerken van de baan of uitkering, arbeidsmarktomstandigheden, economische situatie, regionale kenmerken en institutionele factoren, zoals wet- en regelgeving.

Tabel 7.1 geeft een overzicht van uitkomstmaten die volgen uit de doelstellingen van het activerend arbeidsmarktbeleid zoals beschreven in Hoofdstuk 3. Voor elke uitkomstmaat geeft de tabel globaal aan welke gegevens nodig zijn om het niveau ervan te kunnen meten. Meer specifiek kan de start van deelname aan reguliere arbeid vanuit een uitkering worden gedefinieerd bij een gelijktijdige beëindiging van de uitkering of al bij een verlaging van de uitkering. Andersom kan de uitkomstmaat ook worden gedefinieerd afhankelijk van de beschikbare gegevens. Zo kan werkhervatting worden gedefinieerd als uitstroom uit de uitkering als gevolg van het aanvaarden van werk, als het moment waarop een baan wordt geaccepteerd bij een werkgever of het moment waarop het arbeidscontract wordt ondertekend. Het is voor een effectiviteitsanalyse van groot belang altijd zeer helder te zijn over de definitie van de uitkomstmaat.

Bij een meting van een uitkomstmaat op één bepaald moment, of bij de beschikbaarheid van gegevens over een uitkomstmaat op één bepaald moment, is vooral de absolute waarde van de uitkomstmaat van belang. Hoe hoog is het loon, werkt iemand of niet, hoe groot is het aantal vervulde vacatures? Wanneer een toegepaste evaluatiemethode gebruik maakt van variatie in de tijd en dus van verschillende momenten waarop de uitkomstmaat wordt gemeten, dan is meestal ook de verandering in de uitkomstmaat van belang. Dus op welke momenten in de tijd wordt de hoogte van het loon vastgesteld, wanneer is iemand gaan werken, hoeveel vacatures zijn er binnen een bepaalde periode vervuld?

Bij een aantal uitkomstmaten zit de tijdsdimensie al in de maat zelf, zoals bij de werkloosheidsduur, baanduur en uitkeringsduur. Voor het meten van een duur is het nodig om zowel de begin- als einddatum van de duurperiode waar te nemen. Daarbij is de definitie van de duurperiode van groot belang. Voor het berekenen van een uitkeringsduur tot aan werkhervatting is het nodig om te weten wanneer de uitkering is aangevangen en wanneer er een werkhervatting heeft plaatsgevonden. De einddatum van de uitkering is in zo'n geval minder van belang. Behalve natuurlijk als men wilt meten wat de uitkeringsduur is van personen die de uitkering beëindigen door werkhervatting. Het is dus belangrijk om de uitkomstmaat zo precies mogelijk te definiëren.

Soms zijn meerdere gegevensbronnen nodig om de uitkomstmaat vast te stellen. Wanneer in plaats van de werkloosheidsduur of deelname aan reguliere arbeid de baanzoekduur tot aan werkhervatting de relevante uitkomstmaat is, dan is het van belang om zowel het beginpunt van de baanzoekduur als het eindpunt van de baanzoekduur zorgvuldig te formuleren. Het begin van de baanzoekduur hoeft niet noodzakelijk samen te vallen met het begin van de uitkeringsduur of werkloosheidsduur. Arbeidsgehandicapten zullen bijvoorbeeld pas naar werk gaan zoeken als duidelijk is dat ze (een deel van) het recht op de arbeidsongeschiktheidsuitkering gaan verliezen en dus (deels) werkloos worden. Start van de baanzoekduur is dan bijvoorbeeld de start van een

WW-uitkering voor deze groep. Het vinden van een baan kan dan als einde van de baanzoekduur worden beschouwd, of pas wanneer deze baan de volledige verdien capaciteit gebruikt, dan wel duurzaam is (langer duurt dan bijvoorbeeld een half jaar). De precieze definitie zal behalve van de beschikbaarheid van gegevens ook afhangen van beleidsdoelstellingen.

Tabel 7.1: Globaal overzicht van benodigde gegevens voor het vaststellen van uitkomstmaten

Uitkomstmaat	Benodigde gegevens
(1) (Duurzame) deelname reguliere arbeid	Datum of indicator start arbeidsdeelname; Duurzame arbeidsdeelname vaak gedefinieerd als arbeidsdeelname van minimaal een half jaar
(2) Vervulling van vacatures	Datum of indicator vervulde vacature(s) of aantal vervulde vacatures, eventueel als aandeel van aantal openstaande vacatures
(3) Stapje op re-integratieladder	Trede op de re-integratieladder, indicatiestelling o.i.d.
(4) Arbeidsaanbod in personen of fte	Aantal werkzoekenden of indicatie arbeidsparticipatie
(5) Arbeidsaanbod naar competenties	Aantal werkzoekenden of indicatie arbeidsparticipatie per competentieniveau
(6) Werkloosheidsduur	Datum start en datum einde werkloosheid; Als alternatief geldt de baanzoekduur, waarvan de start kan afhangen van uitkeringsafhankelijkheid (bijvoorbeeld start WW bij arbeidsgehandicapten) en het einde van de start arbeidsdeelname
(7) Baanduur	Datum start en datum einde van baan
(8) (Hernieuwde) instroom in SUWI-keten	Datum of indicator van instroom in SUWI keten
(9) Uitkeringsduur	Datum start en datum einde uitkering; Datum einde uitkering eventueel vervangen door datum start baan
(10) Uitkeringslasten	Zie uitkeringsduur, plus uitkeringsniveau
(11) Deelname maatschappelijke activiteiten	Datum of indicator start relevante activiteiten
(12) Kwaliteit van de baan	Kenmerken van de baan, zoals beroepsniveau, soort contract, aantal uren, werktijden, etc.
(13) Beloning	Bruto of netto loon per maand of per uur
(14) Arbeidsproductiviteit	Toegevoegde waarde per werknemer of loonkosten
(15) Kwaliteit van de match	Kenmerken van de baan in combinatie met kenmerken van de werknemer, zoals gevraagde en aangeboden opleidingsniveau, gevraagde en aangeboden aantal arbeidsuren, etc.

Om effecten van beleidsinterventies vast te kunnen stellen is behalve over de uitkomstmaat ook informatie nodig over de inzet van een beleidsinterventie waarvan het effect op de uitkomstmaat dient te worden geëvalueerd. In Hoofdstuk 3 is een overzicht gegeven van de verschillende instrumenten waarop dit handboek zich richt. Tabel 7.2 geeft per instrument een overzicht van de informatie die nodig is om vast te kunnen stellen of deze beleidsinterventies zijn ingezet. Bij elke effectiviteitsanalyse is het van belang om te weten of deze instrumenten (in het recente verleden) zijn ingezet en dus invloed hebben kunnen uitoefenen op de gemeten uitkomstmaat. Als de effectiviteitsanalyse mede wordt gebaseerd op variatie in de tijd, dan is het nodig te bepalen of binnen de onderzoeksperiode (op individueel niveau) een interventie heeft plaatsgevonden. Bij voorkeur dient het moment waarop de interventie plaatsvindt zo exact mogelijk bekend te zijn. Dit is met name bij een effectiviteitsanalyse met duurmodellen van belang.

Tabel 7.2: Overzicht van benodigde gegevens voor het vaststellen van de inzet van verschillende soorten beleidsinstrumenten

Beleidsinstrument	Benodigde gegevens
Vaardigheidstraining voor het zoekproces	Datum (of indicator) start vaardigheidstraining, eventueel datum afronding vaardigheidstraining
Scholing	Datum (of indicator) start scholing, eventueel datum afronding scholing of behaalde diploma's
Financiële prikkels (zoals sancties, uitstroompremies en loonkostensubsidies)	Datum start én datum einde (of indicator) van financiële prikkel, eventueel met omvang van financiële prikkel
Directe werkgelegenheidscreatie	Datum (of indicator) start werkgelegenheid, eventueel met omvang van het aantal plaatsen
Subsidies voor starten eigen bedrijf	Datum (of indicator) van inzet subsidie of start eigen bedrijf, eventueel met omvang subsidie
Beschikbaar stellen van arbeidsmarktinformatie	Datum (of indicator) start nieuwe dienstverlening
Actieve bemiddeling	Datum (of indicator) start actieve bemiddeling
(Intensieve) begeleiding	Datum (of indicator) start intensieve begeleiding, eventueel met datums (of indicatoren) start individuele componenten

In Tabel 7.2 worden vrijwel geen einddatums van instrumenten genoemd. In principe is het genoeg om te weten of een beleidsinstrument is ingezet. Echter, in een aantal gevallen kan zo'n einddatum wel van belang zijn. Bijvoorbeeld wanneer scholing pas effectief wordt door afronding met behulp van een diploma. Afhankelijk van hoe het effect van een interventie wordt gespecificeerd (zie Paragraaf 7.2.4) dient het begin, eind en verloop van de interventie bekend te zijn. Bij een effectiviteitsanalyse moet in elk geval duidelijk zijn welke veronderstellingen er worden gemaakt over het begin, verloop en einde van de effectiviteit van een beleidsinstrument. Verder kunnen gegevens over het soort beleidsinstrument en de manier waarop het instrument wordt ingezet (welk re-integratiebedrijf, welk contract, type contract, etc.) nuttige informatie zijn voor het verklaren van verschillen in effectiviteit.

Factoren die doorgaans van invloed zijn op de uitkomstmaten in Tabel 7.1 kunnen worden opgedeeld in persoonskenmerken, kenmerken van de baan, kenmerken van de uitkering en omgevingskenmerken. Tabel 7.3 geeft voor elk van deze groepen een *niet* uitputtende lijst met factoren die relevant en in meer of minder mate beschikbaar zijn voor een netto-effectiviteitsanalyse van maatregelen van activerend arbeidsmarktbeleid. In zo'n analyse kunnen kenmerken van de huidige baan of uitkering van belang zijn, maar ook die van de laatste baan of uitkering. Voor een analyse van de baanduur zijn bijvoorbeeld vooral kenmerken van de huidige baan van belang, bij een analyse van de werkloosheidsduur vooral kenmerken van de laatste baan.

Er zijn diverse bronnen beschikbaar waaruit de benodigde gegevens voor een effectiviteitsanalyse kunnen worden betrokken. Allereerst kan onderscheid worden gemaakt tussen administratieve bestanden en enquêtegegevens. Het belangrijkste voordeel van enquêtegegevens ten opzichte van administratieve bestanden is dat ze vaak specifiek gericht zijn op het onderwerp van onderzoek. Daardoor kunnen zeer veel relevante gegevens voor een effectevaluatie worden verzameld, waaronder zachte factoren als de motivatie van werkzoekenden om een baan te vinden. Belangrijkste nadeel is echter dat enquêtegegevens ten opzichte van administratieve bestanden vaak een zeer beperkt deel van de onderzoekspopulatie betreffen, waarbij de respons niet altijd representatief is voor de gehele onderzoekspopulatie. Daardoor is de betrouwbaarheid en detaillering van de onderzoeksresultaten vaak beperkt. Ook bevatten enquêtes meer dan

administratieve bestanden opinies en percepties van individuen, waardoor wederom de betrouwbaarheid en onderlinge vergelijkbaarheid van de gegevens in het geding kunnen zijn.

Voorbeelden van bekende enquêtes voor arbeidsmarktonderzoek zijn:

- OSA-Arbeidsaanbodpanel (www.uvt.nl/osa/)
- Het Sociaal-economisch Panelonderzoek (www.cbs.nl)

Tabel 7.3: Overzicht van factoren die van invloed kunnen zijn op de uitkomstmaten in Tabel 7.1

Persoonskenmerken	Baankenmerken (huidige of laatste baan)
<ul style="list-style-type: none"> • Opleidingsniveau • Opleidingsrichting • Leeftijd of geboortedatum • Geslacht • Arbeidservaring (arbeidsjaren, werkloosheidsperiodes, aantal vervulde banen, etc.) • Overige (verworven) competenties • Uitkeringsverleden (aantal malen in uitkering, type uitkering, duur uitkering, etc.) • Mate van beschikbaarheid (o.a. soort dienst, gewenste aantal arbeidsuren) • Arbeidsmotivatie • Perceptie eigen arbeidsmogelijkheden • Gezondheid of gezondheidsklachten • Gezondheidsbeleving • Arbeidsongeschiktheidspercentage • Afstand tot arbeid (fase-indeling of soortgelijke indicator) • Etniciteit of land van herkomst (van ouders) • Burgerlijke staat • Huishoudsituatie • Aantal kinderen • Leeftijd van kinderen • Regio 	<ul style="list-style-type: none"> • Loon en looncomponenten • Arbeidsduur • Arbeidstijden • Type arbeidscontract (tijdelijk of vast, flexibele arbeid) • Beroep(sniveau) • Bedrijfstak • Bedrijfs grootte • Indicator seizoensarbeid • Gevraagde opleidingsniveau • Andere vereiste of gewenste competenties <hr/> <p>Kenmerken van (huidige of laatste) uitkering</p> <hr/> <ul style="list-style-type: none"> • Uitkeringsniveau • Uitkeringsomvang (als percentage van volledige uitkering) • Uitkeringssoort • Potentiële uitkeringsduur (recht op uitkering, bijvoorbeeld bij aanvang van uitkeringsperiode)
<hr/> <p>Omgevingskenmerken</p> <hr/> <ul style="list-style-type: none"> • Moment van evaluatie (jaar, kwartaal) • Conjuncturele omstandigheden naar tijd en regio. Voorbeelden van gangbare indicatoren voor de economische situatie die relevant zijn voor de regionale arbeidsmarkt zijn: <ul style="list-style-type: none"> ○ Omvang werkloosheid gerelateerd aan omvang werkgelegenheid, bij voorkeur voor verschillende segmenten van de arbeidsmarkt (b.v. opleidingsniveau) ○ Aantal vacatures gerelateerd aan omvang werkgelegenheid, bij voorkeur voor verschillende segmenten van de arbeidsmarkt (b.v. beroepsniveau) ○ Verhouding omvang werkloosheid en aantal vacatures (U/V-ratio), bij voorkeur voor verschillende segmenten van de arbeidsmarkt ○ Economische groei, uitgedrukt in toename toegevoegde waarde van productie ○ Arbeidsproductiviteit als de omvang van de toegevoegde waarde per arbeidsplaats <hr/>	

In de praktijk wordt veelal gebruik gemaakt van administratieve bestanden met gegevens over werkzoekenden en beleidsinterventies. De belangrijkste administratieve bestanden voor het meten van de (netto-)effectiviteit van maatregelen van activerend arbeidsmarktbeleid zijn aanwezig bij de volgende organisaties:

- UWV: met name de WW- en AO-uitkeringsadministratie, registratie van Verzekerde Personen en Dienstverbanden en registratie van de re-integratiedienstverlening aan WW-gerechtigden en arbeidsgehandicapten. De meeste van deze bestanden worden beschikbaar gesteld via het CBS
- CWI: inschrijvingsgegevens van niet-werkende werkzoekenden en vacaturegegevens
- Gemeenten: bijstandsadministratie en gegevens re-integratiedienstverlening bijstandsgerechtigden en niet-uitkeringsgerechtigden
- CBS: voor een overzicht van beschikbare bestanden met administratieve gegevens en grootschalige enquêtegegevens, zie de catalogus microbestanden op www.cbs.nl

7.2 **Bewerking van gegevens**

Beschikbare gegevens zijn niet altijd volledig of precies, zodat ze niet onmiddellijk bruikbaar zijn voor een effectiviteitsanalyse. Veel gegevens dienen eerst te worden bewerkt, zodat ze betekenisvol kunnen worden ingezet. In deze paragraaf wordt een overzicht gegeven van bewerkingen die kunnen worden uitgevoerd op gegevens om ze beter te laten aansluiten bij het doel van de effectiviteitsanalyse.

Selectie van de onderzoekspopulatie

Een eerste belangrijke stap in de bewerking van gegevens voor een effectiviteitsanalyse is de selectie van de relevante onderzoekspopulatie. Daarbij gaat het in de eerste plaats om de doelgroep waarvoor de effectiviteit van een beleidsinterventie wordt berekend. Is dit bijvoorbeeld de gehele populatie werklozen of alleen de werklozen met een grote afstand tot de arbeidsmarkt? Of specifiek oudere werklozen? Selectie van de onderzoekspopulatie kan geschieden op basis van kenmerken van personen (bijvoorbeeld naar afstand tot arbeid of leeftijd), kenmerken van banen (bijvoorbeeld werklozen afkomstig uit voltijdbanen of met uitzondering van seizoensarbeid), of kenmerken van de uitkering (bijvoorbeeld alleen werklozen met een loongebonden WW-uitkering).

Na bepaling van de onderzoekspopulatie (als deel van de gehele populatie) kan vervolgens worden gecontroleerd of personen in de proef- en controlegroep een gelijke basis voor vergelijking hebben. Als in de proefgroep geen personen zitten die ouder zijn dan 55 jaar, dan ligt het voor de hand om ook geen personen ouder dan 55 jaar in de controlegroep op te nemen. Andersom geldt dan uiteraard hetzelfde. Hiermee wordt de onderzoekspopulatie verder ingeperkt. Bij toepassing van een Propensity Score bij de matching methode kan de gelijke basis voor vergelijking worden beperkt tot hetzelfde domein van de Propensity Score in de proef- en controlegroep.

Belangrijk is om te realiseren dat effecten van beleidsinterventies die zijn vastgesteld met behulp van een effectiviteitsanalyse op een afgebakende onderzoekspopulatie ook alleen geldig zijn voor die afgebakende onderzoekspopulatie. Zie voor de specificatie van effecten voor deelpopulaties ook Paragraaf 7.2.5.

Bepaling van begin en eind van duurperioden

Voor de bepaling van duurperioden, zoals werkloosheidsduren en baanduren, zijn goede datums van het begin en eind van de toestand waarvoor een duur wordt bepaald van groot belang. Deze datums komen vaak uit administratieve bestanden, waarin deze gegevens niet altijd exact worden

geregistreerd. Daarbij kan onderscheid worden gemaakt naar het detailniveau waarop datums worden geregistreerd en welke datums het beste worden geregistreerd.

Vaak zijn datums bekend per dag, soms alleen per maand. Datums per dag geven een zeer gedetailleerd idee van het begin en eind van toestanden. Vaak is die nauwkeurigheid schijn. Veel datums worden pas achteraf ingevuld, waardoor er bijvoorbeeld veel datums aan het begin of het eind van een maand worden geregistreerd. Om schijnnauwkeurigheid te voorkomen, kan het beste gebruik worden gemaakt van maandgegevens. Analyses kunnen dan niet meer worden uitgedrukt in het aantal dagen, maar daar staat minder verstoring van de resultaten door administratieve onnauwkeurigheden tegenover.

Verder bestaat in de praktijk verschil tussen het gebruik van begindatums en einddatums. Begindatums zijn over het algemeen beter geregistreerd dan einddatums. Als er kan worden gekozen, dan wordt aanbevolen om alleen begindatums te gebruiken in een effectiviteitsanalyse. Zo kan bijvoorbeeld het einde van een werkloosheidsperiode in veel gevallen ook worden geïdentificeerd aan de hand van het begin van een werkperiode.

Voor een baanzoekperiode geldt dat deze niet noodzakelijkerwijs hoeft samen te vallen met een uitkeringsperiode. Personen kunnen al op zoek zijn naar een baan voordat ze een uitkering ontvangen. Iemand die zonder werk komt te zitten hoeft niet onmiddellijk een bijstandsuitkering te ontvangen, bijvoorbeeld omdat de partner de eerste tijd nog voldoende verdient of sprake is van voldoende eigen vermogen. Pas als ook de partner onder een zekere inkomensdrempel zakt, of het eigen vermogen uitgeput raakt, kan men in aanmerking komen voor een bijstandsuitkering. De baanzoekduur is dan langer dan de uitkeringsperiode. Aan de andere kant kan het voorkomen dat bij werkhervatting (een deel van) de uitkering wordt voortgezet, terwijl men in de praktijk niet verder zoekt naar een baan. In dat geval kan de zoekperiode korter zijn dan de uitkeringsperiode.

Als de uitkomstmaat de baanduur is, dan hoeft die niet gelijk te zijn aan een werkperiode. Iemand kan na re-integratie duurzaam werk vinden door na 4 maanden over te stappen op een andere baan met een vast dienstverband. De eerste baanperiode kan op basis van het gangbare criterium van minimaal 6 maanden niet duurzaam worden genoemd. De werkperiode is dat wel. Het is dus telkens van belang de uitkomstmaat goed te definiëren en duidelijk te maken op basis van welke grootheden de uitkomstmaat wordt vastgesteld.

Bepaling van het moment waarop een uitkomstmaat geldt

Wanneer is iemand werkzaam? Wanneer start iemand zijn uitkering? Ook hier is het van belang duidelijk te zijn over de uitkomstmaat. Iemand is officieel werkzaam vanaf het moment dat zijn of haar contract ingaat. Maar wanneer iemand afspreekt om het contract over een maand in te laten gaan, dan is hij of zij niet langer op zoek naar werk. Het eerste moment (start contract) is van belang indien het effect van een beleidsinterventie op de uitkeringsduur tot aan werkhervatting wordt onderzocht. Het tweede moment (ondertekening contract) is van belang in het geval van een effectmeting op de baanzoekduur.

Bij de start van een uitkering valt het moment waarop iemand een uitkering aanvraagt nooit samen met het moment waarop een uitkering wordt toegewezen. Deze beide momenten staan weer los van het moment waarop een uitkering wordt ontvangen, soms als voorschot, soms als definitieve beschikking. Het is dan niet eenvoudig het moment te specificeren waarop een uitkering start. In de praktijk wordt vaak gekozen voor het datumveld waarbij voor de meeste mensen in de onderzoekspopulatie de meest betrouwbare gegevens zijn of kunnen worden verzameld. Zeker wanneer de variatie binnen dat datumveld onafhankelijk is van de inzet van een beleidsinterventie, zorgt die keuze nauwelijks voor vertekening van de effectmeting.

Bepaling van begin en einde van een interventie

Wat nu precies wordt bedoeld met de beleidsinterventie waarvan het effect in een effectiviteitsanalyse wordt gemeten, dient altijd zeer helder te worden gemaakt in de evaluatie. Bestaat de interventie bijvoorbeeld uit de inzet van een re-integratietraject of uit de inzet van een re-integratie-instrument waarvan er meerdere kunnen worden ingezet binnen een re-integratietraject? Is het effect van een re-integratietraject de som van het effect van de afzonderlijke instrumenten, of wordt het effect bepaald door de unieke combinatie van afzonderlijke instrumenten? De manier waarop het effect het beste kan worden gespecificeerd in een effectiviteitsanalyse wordt nader toegelicht in Paragraaf 7.2.5.

Ook het moment waarop een interventie begint en eindigt dient duidelijk te worden bepaald. Daar hangt immers de werkingsduur van de interventie op de uitkomstmaat van af. Start een re-integratietraject bijvoorbeeld met het eerste instrument dat binnen dat traject wordt ingezet, met de intake voor het traject of met de brief waarin deelname aan het re-integratietraject wordt aangekondigd? Belangrijk voor de effectiviteitsanalyse is de vraag wanneer men denkt dat welk deel van het traject effectief wordt. Wil men evalueren wat het effect is van de aankondiging van een traject, eventueel naast het effect van de aparte instrumenten binnen het ingezette traject, dan is de brief waarin deelname aan het re-integratietraject wordt aangekondigd de start van de interventie. De voordracht voor een re-integratietraject is in dat geval onderdeel van de beleidsinterventie, waardoor alle personen met een voordracht tot de proefgroep behoren. Wil men evalueren wat het effect is van instrumenten binnen een traject, dan is de start van de inzet van dat instrument de start van de interventie. In het laatste geval behoren personen die alleen een brief of intake hebben gehad tot de controlegroep.

Bij het einde van een interventie speelt een soortgelijke kwestie. Neem als voorbeeld scholing. Wordt de interventie beëindigd met de laatste cursusdag of met het uitreiken van het diploma? Is de interventie effectief tot de datum van beëindiging of juist vanaf die datum? Hoelang blijft de inzet van de interventie invloed hebben op de uitkomstmaat? Een jaar na afronding of voor altijd? In de praktijk mag men al blij zijn met betrouwbare informatie over één van deze momenten. De empirische specificatie van beleidsinterventies in effectiviteitsanalyses is daarom altijd een afweging tussen de gewenste definitie en de beschikbaarheid van gegevens. Voor de interpretatie van de analysesresultaten is het daarom van groot belang om keuzes in de uiteindelijke empirische specificatie te verantwoorden.

Specificatie van effecten van de beleidsinterventie

Effecten van beleidsinterventies zijn niet voor iedereen en onder alle omstandigheden hetzelfde. Scholing zal bij de ene werkloze beter aanslaan dan bij de andere. Intensieve bemiddeling lijkt bijvoorbeeld voor groepen met een korte afstand tot arbeid een te zware maatregel, maar voor groepen die ver afstaan van de arbeidsmarkt is bemiddeling wellicht nog een brug te ver.

Wanneer men geïnteresseerd is in het gemiddelde effect van een interventie zoals die is ingezet in een populatie, dan kan men volstaan met het berekenen van een *overall* effect in die populatie. De informatieve waarde van een gemiddeld effect van ingezette instrumenten in de gehele populatie is echter beperkt, omdat die afhangt van de specifieke inzet van bepaalde instrumenten bij bepaalde (groepen van) personen. Een andere samenstelling van re-integratietrajecten tussen groepen kan leiden tot een ander overall effect.

Wanneer men geïnteresseerd is in welke beleidsinterventie er nu precies voor wie werkt, dan is het van belang om het effect van de beleidsinterventie uit te splitsen naar doelgroepen en ingezette instrumenten. De uitsplitsing naar doelgroepen kan op twee manieren. Op de eerste plaats kunnen aparte deelpopulaties worden geselecteerd, waarbinnen het effect van een interventie apart wordt bekeken (zie ook Paragraaf 7.2.1). Het berekende effect geldt dan uitsluitend en specifiek voor die deelpopulatie. Deelpopulaties kunnen echter te klein zijn voor

betrouwbare effectiviteitsanalyses. In dat geval kan men ook een effect schatten op de gehele populatie, waarin het netto-effect wordt uitgesplitst naar deelgroepen in de populatie. Formeel kan dit voor twee deelpopulaties worden weergegeven door

$$UM = (X \times \mathbf{b}) + (BI_1 \times \mathbf{g}_1) + (BI_2 \times \mathbf{g}_2) + \mathbf{e} \quad (7.1)$$

De indicator BI_1 geeft aan dat de interventie is ingezet bij iemand in deelpopulatie 1. De parameter \mathbf{g}_1 geeft dan het effect van de interventie in deelpopulatie 1. Evenzo voor de interventie in deelpopulatie 2. Samen met de invloedsfactoren X en een onzekerheidsfactor \mathbf{e} wordt op die manier de uitkomstmaat UM verklaard. Nadeel van deze benadering is dat het effect minder specifiek (nauwkeurig) wordt gemeten dan in een aparte analyse van de betreffende deelpopulatie, omdat de correctie voor samenstellingseffecten ($X \times \mathbf{b}$) plaatsvindt op het niveau van de gehele populatie in plaats van de deelpopulatie. Voordeel is echter dat het effect geldig is voor de gehele populatie waarop de effectiviteitsanalyse heeft plaatsgevonden.

Effecten van beleidsinterventies kunnen op verschillende momenten ingaan, bijvoorbeeld bij de start van een interventie, na afloop ervan of gedurende de interventie. Ook kunnen effecten na verloop van tijd minder sterk worden. Dit dient empirisch te worden vastgesteld door in de modellering rekening te houden met verschillende effecten gedurende en na afloop van de looptijd van een interventie. Als voorbeeld kan de inzet van scholing op de kans om een baan te vinden worden gebruikt. Als men start met scholing heeft men er vermoedelijk niet onmiddellijk profijt van. Pas wanneer die scholing vrijwel is afgerond of een diploma is behaald zal het naar verwachting effect sorteren. Na verloop van tijd verwatert de kennis opgedaan door scholing echter weer, waardoor de effectiviteit afneemt. Het is dus voor sommige instrumenten van belang in de effectiviteitsanalyse rekening te houden met mogelijke verschillen in effectiviteit naar de duur van de inzet van het instrument. Dit kan door het effect van een interventie uit te splitsen naar de periode waarin iemand met de interventie te maken heeft. Dat kan op dezelfde wijze als in bovenstaande vergelijking van deelpopulaties, alleen gelden de indicatoren BI_1 en BI_2 niet voor deelpopulaties, maar voor het moment waarop personen in de interventie worden waargenomen. Zo kan BI_1 gelden voor een inzet van de interventie gedurende de eerste maand en BI_2 vanaf de tweede maand. Het effect van de interventie gedurende de eerste maand wordt dan gemeten door \mathbf{g}_1 . Het totale effect vanaf de tweede maand is dan gelijk aan $(\mathbf{g}_1 + \mathbf{g}_2)$. De meest eenvoudige benadering is de veronderstelling dat effecten vanaf de start van een interventie gelden en constant blijven in de tijd. Dat is een redelijke benadering als de effectiviteit wordt vastgesteld kort na afloop van de interventie.

Effecten van een beleidsinterventie kunnen apart worden vastgesteld voor verschillende omstandigheden waaronder de beleidsinterventie plaatsvindt. Zo kan het effect van bemiddeling afhangen van het moment waarop het wordt ingezet (vroeg of laat gedurende de werkloosheidsperiode). Bij een te snelle inzet is een werkzoekende wellicht nog niet klaar om de arbeidsmarkt op te gaan of kan uit zichzelf ook prima een baan vinden. In het laatste geval betekent de inzet van bemiddeling een zogenaamde ‘deadweight loss’. Bij een late inzet van bemiddeling kan het effect worden gedrukt doordat de afstand tot de arbeidsmarkt (te) groot is geworden. In een effectiviteitsanalyse kan het effect naar het moment van inzet worden vastgesteld door bijvoorbeeld onderscheid te maken tussen een vroege inzet van bemiddeling, weergegeven door BI_1 in vergelijking (7.1), en een late inzet van bemiddeling, weergegeven door BI_2 in vergelijking (7.1). Als het effect \mathbf{g}_1 groter is dan het effect \mathbf{g}_2 , dan kan bemiddeling beter vroeg worden ingezet. Als \mathbf{g}_1 kleiner is dan \mathbf{g}_2 , dan is late bemiddeling beter. Overigens kunnen de effecten \mathbf{g}_1 en \mathbf{g}_2 ook worden vastgesteld door middel van aparte analyses in deelpopulaties met een vroege en late inzet van bemiddeling. Daarbij dient wel speciale aandacht te worden besteed aan de samenstelling van de controlegroep.

Een ander voorbeeld van variatie in het effect van een beleidsinterventie naar de omstandigheden waaronder die inzet plaatsvindt, is variatie in het effect naar conjuncturele omstandigheden. Zo is het effect van scholing anders tijdens hoogconjunctuur dan tijdens laagconjunctuur. Ook hier kan een aparte effectiviteitsanalyse worden uitgevoerd voor verschillende deelpopulaties, afhankelijk van de stand van de conjunctuur. En kan het effect voor die deelpopulaties worden gemeten in een effectanalyse voor de gehele populatie, waarbij onderscheid wordt gemaakt tussen BI_1 en BI_2 .

Specificatie van variabelen

In een econometrische vergelijking worden vaak tal van variabelen opgenomen om te corrigeren voor samenstellingseffecten tussen de proef- en controlegroep. Deze variabelen verklaren de invloed van verschillende factoren, zoals benoemd in Paragraaf 7.1. Die invloed geldt altijd ten opzichte van een gekozen referentiecategorie. Zo kan de invloed van geslacht worden vastgesteld door de uitkomstmaat van vrouwen te vergelijken met die van mannen. Mannen vormen dan bijvoorbeeld de referentiecategorie ten opzichte waarvan het effect van het vrouw zijn wordt bepaald. Zo dienen voor alle kenmerken referentiecategorieën te worden gekozen. Aanbevolen wordt om zoveel mogelijk referentiegroepen te kiezen die de grootste omvang in de onderzoekspopulatie hebben (het vaakst voorkomen). Dus het effect van seizoenswerkloosheid ten opzichte van de referentiecategorie niet-seizoenswerkloosheid en niet andersom. Bij continue variabelen zoals leeftijd kan gebruik worden gemaakt van categorieën (leeftijdsklassen) of voor een functionele vorm (bijv. een kwadratische relatie door leeftijd en leeftijd in het kwadraat op te nemen). Daarbij kan worden gekozen om alles in afwijking van 0 te specificeren of ten opzichte van de gemiddelde leeftijd. In dat laatste geval dient van de werkelijke leeftijd de gemiddelde leeftijd in de onderzoekspopulatie te worden afgetrokken.

7.3 Presentatie van effecten

Vastgestelde effecten kunnen op verschillende manieren worden gepresenteerd. Het ligt voor de hand om effecten uit te drukken in een verandering van de uitkomstmaat. Maar zowel de uitkomstmaat als effecten kunnen sterk variëren afhankelijk van de situatie waarin ze zich voordoen. Het is daarom van belang heel precies te zijn in de presentatie van effecten. Voor welke samenstelling van de populatie worden effecten gepresenteerd? Voor welke periode of voor welk moment? Voor welke combinatie van beleidsinstrumenten? Bij inzet op welk tijdstip? Moeten de resultaten worden uitgedrukt in absolute effecten of relatieve effecten? Deze paragraaf probeert een overzicht te geven van manieren waarop resultaten kunnen worden gepresenteerd en aanbevelingen te doen voor een uniforme presentatie van resultaten.

Effecten binnen of buiten de onderzoekspopulatie

Belangrijk is om te benadrukken dat vastgestelde effecten van beleidsinterventies uitsluitend gelden voor de specifieke onderzoekspopulatie en onderzoeksperiode waarop de effectmeting heeft plaatsgevonden. Effecten kunnen worden gepresenteerd als het verschil in de uitkomstmaat tussen personen of groepen mét interventie en personen of groepen zonder interventie. Effecten kunnen echter ook worden berekend voor personen die buiten de onderzoekspopulatie vallen. Berekeningen vallen buiten de onderzoekspopulatie als ze een doelgroep betreffen die niet in de onderzoekspopulatie zit (bijvoorbeeld ouderen), maar waarnaar op basis van kenmerken in de onderzoekspopulatie wel kan worden geëxtrapoleerd (leeftijd), of als ze een voorspelling voor de toekomst inhouden. Zo'n extrapolatie van effecten wordt uitgevoerd onder de veronderstelling dat de effectiviteit van de beleidsinterventie niet wezenlijk verschilt binnen en buiten de onderzochte onderzoekspopulatie en -periode. Die veronderstelling kan echter niet worden getoetst. Als er al effectberekeningen plaatsvinden buiten de onderzoekspopulatie en -periode, dienen deze daarom goed te worden verantwoord.

Een speciaal geval is het uitdrukken van effecten op basis van duurmodellen. Effecten kunnen worden uitgedrukt in kansen, bijvoorbeeld de kans op werkhervatting of de kans op uitstroom uit een uitkering binnen een bepaalde periode, of in duren, bijvoorbeeld het aantal maanden verkorting van de verwachte uitkeringsduur. De berekening van de verwachte uitkeringsduur is gebaseerd op kansen die verder in de toekomst liggen dan de kans op werkhervatting binnen een bepaalde periode. De verwachte uitkeringsduur maakt dan ook veel meer gebruik van berekeningen die buiten de onderzoeksperiode vallen dan de kans op werkhervatting. Het verdient daarom de voorkeur om effecten op basis van duurmodellen uit te drukken in kansen in plaats van duren.

Absolute verschillen versus relatieve verschillen

Omdat effecten van beleidsinterventies kunnen worden uitgedrukt in het verschil in de uitkomstmaat tussen personen of groepen mét een interventie en personen of groepen zonder interventie, is het mogelijk om effecten zowel uit te drukken in een absoluut verschil als in een relatief verschil. Een absoluut verschil in een uitkomstmaat is bijvoorbeeld het verschil in uurloon dat mensen met en zonder interventie verdienen of de verandering van de kans op werkhervatting, uitgedrukt in procentpunten. Daaraan gerelateerde relatieve verschillen zijn het percentage inkomen dat iemand met een interventie meer verdient dan iemand zonder interventie en het percentage verandering in de kans op werkhervatting.

Aanbevolen wordt om effecten bij een min of meer continue uitkomstmaat (zoals uitkeringsduur of inkomen) uit te drukken in een relatieve verandering van de uitkomstmaat (procenten) en effecten bij een zogenaamde discrete of categorale uitkomstmaat (zoals werkhervattingkans of aandeel moeilijk vervulbare vacatures) uit te drukken in een absolute verandering (procentpunten).

Evaluatiemoment

Behalve tussen personen in de onderzoekspopulatie kunnen uitkomstmaten en effecten van beleidsinterventies ook sterk variëren in de tijd. Een effect kan een maand na de beleidsinterventie anders zijn dan een jaar na de beleidsinterventie, afhankelijk van de variatie in de uitkomstmaat en de toe- of afname van de effectiviteit in de loop van de tijd. Het kan daarom van belang zijn om een eenduidig tijdstip te kiezen waarvoor het effect van een beleidsinterventie wordt gepresenteerd. Een afweging daarbij is dat bij een latere evaluatie alle effecten van de interventie tot uitdrukking kunnen komen, maar bepaalde effecten ook al weer aan erosie onderhevig zullen zijn. Bij een vroege evaluatie komen de resultaten van de effectiviteitsanalyse eerder beschikbaar voor beleidsmakers.

Aanbevolen wordt om dat tijdstip kort te leggen na de volledige implementatie van de interventie. Bij re-integratietrajecten moet bijvoorbeeld rekening worden gehouden met de duur tot aan de start van een traject en de doorlooptijd van het traject zelf. In de praktijk betekent dat een evaluatie van de werkloosheidsduur of de kans op werkhervatting het beste ongeveer anderhalf jaar na de start van een traject kan plaatsvinden, of rekening houdend met het feit dat de meeste trajecten binnen een jaar na aanvang van werkloosheid worden ingezet, tussen twee en twee en een half jaar na aanvang van de werkloosheid.

Moment waarop interventie effectief wordt

Ook bij een tijdsvariabele inzet van beleidsinterventies treden er in de praktijk verschillen op in uitkomstmaten en effecten tussen personen in de populatie. Voor de presentatie van een gemiddeld effect dient een ‘gemiddeld’ moment van inzet te worden gekozen om te abstraheren van de tijdsvariabele inzet van beleidsinterventies en zo individuele personen vergelijkbaar te maken. Dit ‘gemiddelde’ moment van inzet dient voor elke persoon in de populatie gelijk te zijn, omdat anders geen zuivere vergelijking plaatsvindt tussen personen. Dit geldt in het bijzonder

wanneer het werkelijke moment van inzet bij personen selectief is, bijvoorbeeld op grond van leeftijd of afstand tot arbeid. In de berekening en presentatie van effecten dient voor die selectiviteit te worden gecorrigeerd. Het gekozen 'gemiddelde' moment van inzet dient bij voorkeur dicht bij het gemiddelde moment te liggen waarop interventies in werkelijkheid worden ingezet. Bij de inzet van re-integratietrajecten voor werklozen dus ergens tussen een half jaar en een jaar na aanvang van de werkloosheid.

Populatie waarvoor effecten worden berekend

Omdat effecten vaak afhangen van de individuele situatie van personen, kan ook het beste een 'gemiddeld' effect worden gepresenteerd. Belangrijk daarbij is de samenstelling van de populatie waarover dat gemiddelde wordt berekend. Niet alleen omdat effecten af kunnen hangen van kenmerken van personen, maar ook omdat effecten anders kunnen uitpakken bij een ander niveau van de uitkomstmaat. Bij werklozen met een grote kans op werkhervatting zal die kans niet veel meer kunnen toenemen door de inzet van een re-integratietraject. Bij werklozen met een kleine kans op werkhervatting kan die kans in principe nog flink toenemen, maar is het bijvoorbeeld mogelijk dat re-integratietrajecten een veel kleiner effect hebben. Voor een gemiddeld effect van de beleidsinterventie kan daarom in veel gevallen niet worden volstaan met het berekenen van een effect voor een referentiepersoon met gemiddelde kenmerken.

Het gemiddelde effect over de gehele onderzoekspopulatie kan het beste worden berekend door eerst het verschil te nemen in de uitkomstmaat mét en zonder beleidsinterventie voor alle personen in de onderzoekspopulatie en daarover vervolgens het gemiddelde te berekenen. Het zo verkregen effect is specifiek voor de samenstelling van de onderzoekspopulatie en kan dus niet goed worden vergeleken met de inzet van dezelfde beleidsinterventie in een andere onderzoekspopulatie. Om de inzet van bijvoorbeeld bemiddeling te vergelijken tussen de WW-populatie en de bijstandspopulatie, kan ervoor worden gekozen om uit te gaan van een referentiepersoon met typische karakteristieken die in beide populaties voorkomt (bijvoorbeeld oudere mannen met een grote afstand tot de arbeidsmarkt). Op die manier kan worden achterhaald wat het verschil in effectiviteit van bemiddeling is voor vergelijkbare personen in de WW- en bijstandspopulatie. Daarbij moet wel worden bedacht dat een referentiepersoon nooit volledig vergelijkbaar kan worden gemaakt tussen beide populaties, omdat onvergelijkbare kenmerken (zoals het arbeidsverleden) ervoor zorgen dat een persoon zich juist in de ene of de andere populatie bevindt.

Om effecten van allerlei kenmerken op de uitkomstmaat te berekenen, kan bij de berekening van gemiddelde effecten steeds één van de kenmerken constant worden gehouden. Het gemiddelde effect voor mannen en vrouwen wordt bijvoorbeeld berekend door te veronderstellen dat de gehele populatie respectievelijk man en vrouw is en vervolgens het gemiddelde te nemen over de gehele populatie. Dat voorkomt dat kenmerken die samenhangen met geslacht het verschil in effectiviteit tussen mannen en vrouwen verstoren. Vaak versterken bepaalde effecten elkaar echter, zodat het inzichtelijk is om ook een combinatie van verschillende kenmerken constant te houden bij de berekening van effecten. Er kunnen bijvoorbeeld een beperkt aantal typische personen worden gedefinieerd aan de hand van een combinatie van kenmerken, waarvoor effecten van de interventie worden berekend.

Gevoeligheidsanalyses

Voor het vaststellen van de netto-effectiviteit van maatregelen van activerend arbeidsmarktbeleid kan gedegen empirisch onderzoek een belangrijke rol vervullen. In empirisch onderzoek wordt echter noodzakelijkerwijs een aantal keuzes gemaakt die gebaseerd zijn op subjectieve inzichten van de onderzoeker. Dit betreft keuzes ten aanzien van de te gebruiken onderzoeksgegevens, de te hanteren onderzoeksmethode, bewerking van de onderzoeksgegevens, specificaties van de onderzoeksmethode en de presentatie van de onderzoeksresultaten. Elk van deze keuzes heeft invloed op de uitkomsten van het onderzoek. Vaak blijft echter onduidelijk in welke richting de

uitkomsten worden beïnvloeden en in welke mate. De beste manier om erachter te komen welke aanpak 'optimaal' is, is om deze af te zetten tegen alternatieven en de verschillen in uitkomsten te vergelijken op basis van plausibiliteit, verklaringskracht, reproduceerbaarheid en complexiteit. Dit gebeurt in zogenaamde gevoeligheidsanalyses.

Een voorbeeld van zulke gevoeligheidsanalyses wordt gegeven in Heyma (2005). Geconstateerd werd dat verschillende Nederlandse studies naar de netto-effectiviteit van re-integratietrajecten voor WW-gerechtigden tot verschillende uitkomsten hadden geleid, ook al waren ze grotendeels gebaseerd op dezelfde gegevensbronnen. Door middel van gevoeligheidsanalyses probeerde Heyma (2005) erachter te komen door welke factoren deze verschillen konden worden verklaard. Daarbij werd niet alleen gekeken naar verschillende specificaties van de gehanteerde onderzoeksmethode, maar ook naar verschillen in de specificatie van de ingezette instrumenten en naar verschillen in de afbakening van de gekozen onderzoekspopulatie. Het aanvullen van de uiteindelijke resultaten van een effectmeting met de resultaten van gevoeligheidsanalyses die laten zien wat de consequenties zijn van gemaakte veronderstellingen en keuzes in de effectiviteitsanalyse, zorgt voor een betere verantwoording van de analyseresultaten en een groter vertrouwen in de robuustheid van de resultaten. Gevoeligheidsanalyses zijn daarom zeer wenselijk bij netto-effectiviteitsanalyses.

Bijlage 1: Literatuur

- Angrist, J. en B. Evans (1996), *Schooling and Labor Market Consequences of the 1970 State Abortion Reforms*, NBER working paper 5406, January 1996.
- Angrist, J. en B. Evans (1998), "Children and Their Parents Labor Supply: Evidence from Exogenous Variation in Family Size", *American Economic Review*, June 1998, 88, 450-477.
- Angrist, J. en A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?", *Quarterly Journal of Economics*, November 1991, 106, 979-1014.
- Angrist, J. en V. Lavy (2002), *The Effect of High School Matriculation Awards: Evidence from Randomized Trials*, NBER working paper 9389, December 2002.
- Ashenfelter, O. (1978), "Estimating the Effect of Training Programs on Earnings", *The Review of Economics and Statistics*, Vol. 60, No. 1. (Feb., 1978), pp. 47-57.
- Ashenfelter, O. En A. Krueger (1994), "Estimates of the Economic Return to Schooling From A New Sample of Twins", *American Economic Review*, December 1994, 84, 1157-1173.
- Beek, H. ter, Brukman, M. & Gent. M. van (2008). *Sturingsinformatie lokaal re-integratiebeleid*. Den Haag: Raad voor Werk en Inkomen.
- Berg, G.J. van den, Homburg, G.H., Hospers, L.A. & Klaauw, B. van der (2003). Activering en controle weinig effectief. *ESB* 88 nr. 4397 pp. 140-142.
- Berkhout, P.H.G., N. Niessen, W. Eddy, R. van de Lustgraaf en A. op het Veld (2008), *Evaluatie verdeelmodel Wet werk en bijstand*, SEO-rapport 2008-13.
- Bolhaar, J., Ph. de Jong, B. van der Klaauw en M. Lindeboom (2004), *UWV als poortwachter; Effecten van intensieve toetsing van het reïntegratieverslag: resultaten van een experiment*, APE-rapport 01-125, Amsterdam / Den Haag.
- Brouwer, P., Zwinkels, W.S. en Genabeek, J. van (2006). *Maatschappelijke kosten en baten van gesubsidieerde arbeid en sociale activering*. (TNO-rapport 22516/25420). Hoofddorp, TNO.
- Cameron, A.C. en P.K. Trivedi (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.
- Centraal Planbureau (2000), *Arbeidsbemiddeling en –reïntegratie van werklozen. Welke rol heeft de overheid te spelen?*, CPB-Werkdocument no. 118, Den Haag: CPB.
- CPB (2007). *De maatschappelijke kosten en baten van re-integratie*. (CPB notitie 9 juli 2007). Den Haag: CPB
- Doodeman, Th.J.J., Lobregt, J.H.S., Montfort, K. van & Zwinkels, W.S. (1996). Effecten van reïntegratieactiviteiten op de arbeidsmarkt voor verschillende opleidingsniveaus. *Maandschrift Economie* 60 pp. 500-513.
- Eijgenraam, C.J.J.M., Koopmans, C.C., Tang, P.J.G. & Verster, P.J.G. (2000). *Evaluatie van infrastructuurprojecten, leidraad voor kosten-batenanalyse*. (CPB bijzondere publicatie 22). Den Haag/Rotterdam: CPB/NEI.
- Gorter, C. & Kalb, G.R. (1996). Estimating the effect of counseling and monitoring the unemployed using a job search model. *Journal of Human Resources* 3 pp. 590-610.
- Graaf-Zijl, M. de, en D. Hollanders (2007), *Tweedekans re-integratie in cijfers*, SEO rapport 2007-46, Amsterdam: SEO Economisch Onderzoek.
- Graaf-Zijl, M. de, A. Heyma en T. de Hoop (2006), *Van bijstand naar werk in Amsterdam*, SEOrapport 903, Amsterdam: SEO Economisch Onderzoek.
- Graaf-Zijl, M. de, C.M.E. Groot en J.P. Hop (2006), *De weg naar werk*, SEO rapport 887, Amsterdam: SEO Economisch Onderzoek.
- Graaf-Zijl, M. de, C.M.E. Groot, A. Heyma, J.P. Hop en L.R. Janssens (2005), *Marktwerving in de reïntegratie*, SEO rapport 760, Amsterdam: SEO Economisch Onderzoek.
- Groot, C., Kok, L. & Koopmans, C. (2003). *Kosten en baten van reïntegratiediensten*. (SEO-rapport 677). Amsterdam: SEO Economisch Onderzoek.

- Groot, I., Graaf-Zijl, M. de, Hop, P., Kok, L.M., Fermin, B. Ooms, D. & Zwinkels, W. (2007). *De lange weg naar werk. Beleid voor langdurig uitkeringsgerechtigden in de WW en de WWB*. Den Haag: Raad voor Werk en Inkomen.
- Groot, C.M.E., M. de Graaf-Zijl en J.P. Hop (2007), *Re-integratie van niet-uitkeringsgerechtigde werkzoekenden*, SEO rapport 2007-62, Amsterdam: SEO Economisch Onderzoek.
- Heckman, J. en R. Robb (1985a), "Alternative Methods for Evaluating the Impact of Interventions." In *Longitudinal Analysis of Labor Market Data*, J. Heckman and B. Singer (eds.), Cambridge, MA: Cambridge University Press, 156–245.
- Heckman, J. en R. Robb (1985b), "The Value of Longitudinal Data for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes." In *Panel Surveys*, D. Kasprzyk, G. Duncan, G. Kalton, and M. Singh (eds.), New York: John Wiley & Sons, 512–538.
- Heckman, J.J., R.J. Lalonde R. J. en J. A. Smith (1999), "The economics and econometrics of active labor market programs", in *Handbook of labor economics*, Vol. 3, Ashenfelter, A. en D. Card, eds., Elsevier Science, Amsterdam.
- Hekelaar A., Zwinkels, W.S., & Braat, A. (2006). *De juiste klant op het juiste traject, een onderzoek naar de netto-effectiviteit van het Rotterdamse reïntegratiebeleid voor het ontwikkelen van klantprofielen*. Rotterdam/Hoofddorp: SWA/TNO Kwaliteit van Leven | Arbeid.
- Heul, H. van der, E.M. Berendsen, B.C.J.L. van der Eijken en R.J.J. Vlek (2003), *Op weg naar werk: Een evaluatie van sluitende aanpak in de WW*, UWV, Afdeling Beleidsinformatievoorziening.
- Heyma, A. (2002), "De effectiviteit van reïntegratie-instrumenten voor arbeidsgehandicapten", in E.L. de Vos (redactie), *Daadwerkelijk effectief; Prestatiemeting van reïntegratie en activering*, Hoofddorp: TNO Arbeid.
- Heyma, A. (2005), "Reïntegratie instrumenten voor WW-gerechtigden: behoefte, inzet en effecten", in E.L. de Vos (redactie), *Verraderlijk effectief. Prestatiemeting van reïntegratie en activering deel 2*, Hoofddorp: TNO Arbeid.
- Heyma, A. en M. Zijl (2005), *De weg terug: Van arbeidsongeschiktheid naar werk*, SEO-rapport 844, Amsterdam: SEO Economisch Onderzoek.
- Heyma, A., W. Zwinkels en J. van Seters (2003), *Doelgroepindeling aanbesteding reïntegratie; derelatie tussen persoonskenmerken, kosten en effectiviteit van reïntegratietrajecten*, SEO-rapport 718B, Amsterdam: SEO Economisch Onderzoek.
- Holland (1986), "Statistics and Causal inference (with discussion)", *Journal of the American Statistical Association*, vol. 81, 945-970
- Interdepartementaal Overlegorgaan Financieel Economische Zaken (IOFEZ) (2004). *Eindrapport VBTB-evaluatie - Lessen uit de praktijk*.
- Jansen, H. (2001), *De arbeidsmarktpositie van de ex-deelnemers aan de EAU*, Regioplan, Den Haag: Ministerie van Sociale Zaken en Werkgelegenheid.
- Jong, Ph. de, M. Lindeboom en B. van der Klaauw (2005), *Stricter screening of disability insurance applications*, mimeo, Amsterdam: Tinbergen Instituut.
- Lauwerys, L., Matheus, N. & Nicaise, I. (2000). *Sociale tewerkstelling in Vlaanderen: doelgroepbereik, kwaliteit en doelmatigheid*. Leuven: HIVA.
- Klaveren, C. van, en H. Maassen van den Brink (2007), "Intra-household work time synchronization togetherness or material benefits?", *Social Indicators Research*, 84(1), 39-52.
- Kok, L., Hollanders, D. & Hop, J. (2006). *Kosten en baten van reïntegratie*. (SEO-rapport, 928). Amsterdam: SEO Economisch Onderzoek.
- Koning, J. de, Gelderblom, A., Zandvliet K. & Boom, K. van den (2006). *Effectiviteit van re-integratie. De stand van Zaken. Literatuuronderzoek, in opdracht van het ministerie van Sociale Zaken en Werkgelegenheid*. (SEOR werkdocument nr. 339). Rotterdam: SEOR.

- Leuven, E. en H. Oosterbeek (2004), "Evaluating the Effect of Tax Deductions on Training", *Journal of Labor Economics*, University of Chicago Press, vol. 22(2), pages 461-488.
- Leuven, E., H. Oosterbeek en B. van der Klaauw (2005), *The Effect of Financial Rewards on Students' Achievement: Evidence From a Randomized Experiment*, Working Paper, version March 2005.
- Martin, J.P. & Grubb, D. (2001). What works and for whom: a review of OECD countries' experiences with active labour market policies. *Swedish Economic Policy Review* 8(2) pp. 9-56.
- Ministerie van Financiën (1999). *Van beleidsbegroting tot beleidsverantwoording*, Regeringsnota 19 mei, Den Haag.
- Ministerie van Financiën (1999). *Regeling Prestatiegegevens en Evaluatieonderzoek*, persbericht 01/067, Den Haag.
- Ministerie van Financiën (2006). Regeling periodiek evaluatieonderzoek en beleidsinformatie 2006. *Staatscourant* nr.83, p. 14.
- Mooij, R. de (2006). *Reinventing the Welfare State*. (CPB bijzondere publicatie 60). Den Haag: CPB.
- Nes, P.J. van, Gravesteyn-Ligthelm, J.H., Vossen, I. & Visscher, M. (2002). *Evaluatie WW-experimenten. Onderzoek naar het gebruik van de experimenten*. Den Haag: SZW.
- Plug, E. (2004), "Estimating the Effect of Mother's Schooling on Children's Schooling using a Sample of Adoptees", *American Economic Review*, March 2004, 94, 358-368.
- Roy, A. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers Psychology*, 3, 135-146.
- Rubbrecht, I., Matheus, N., d'Addio, A. & Nicaise, I. (2005). *Sociale tewerkstelling in Vlaanderen: effecten en maatschappelijk rendement op lange termijn*. Leuven: HIVA.
- Rubin, D.B. (1974), "Estimating Causal Effects and Treatments in Randomized and Non Randomized Studies", *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D.B. (1976), "Inference and Missing Data", *Biometrika*, 63, 581-592.
- Sacerdote, B. (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*, May 2001, 116, 681-703.
- Scholten, P. (2003), *Maatschappelijk rendement gemeten, social return on investment*. Amsterdam, Uitgeverij SWP.
- Scholten, P. (red.) (2005). *Social return on investment, handleiding voor het meten van maatschappelijk rendement*. Amstelveen: Lenthe Publishers.
- (Splawa)-Neyman, J. (1923, 1990), "On the Principles of Probability Theory to Agricultural Experiments. Essays on Principles. Section 9," translated and edited by D.M. Dabrowska and T.P. Speed from *Roczniki Rolniczych Tom X* (1923), 1-51 (*Annals of Agricultural Sciences*), *Statistical Sciences*, 5, 465-471.
- SZW (2001). *Begroting Sociale Zaken en Werkgelegenheid*. In: MinFin (2001). *Rijksbegroting 2002*. Den Haag: Ministerie van Financiën.
- SZW (2004). *Begroting Sociale Zaken en Werkgelegenheid*. In: MinFin (2004). *Rijksbegroting 2005*. Den Haag: Ministerie van Financiën.
- SZW (2007). *Begroting Sociale Zaken en Werkgelegenheid*. In: MinFin (2007). *Rijksbegroting 2008*. Den Haag: Ministerie van Financiën.
- SZW (2008). *Beleidsdoorlichting Re-integratie*. Den Haag: Ministerie van Sociale Zaken en Werkgelegenheid.
- UWV (2008), *Werk*, UWV Kwartaal Verkenning, UKV 2008-I, Amsterdam: UWV.
- Veld, R. in 't, Korving, W. Hamdan, Y. & Steen, M. van den (2005). *Kosten en Baten van Voortijdig Schoolverlaten*.
- Vos, E.L. (eds.) (2004). *Measuring effects: roads to success in rehabilitation : how to measure re-employment policies in The Netherlands? : proceedings of the conference*. (TNO-rapport 18441/25042). Hoofddorp: TNO Kwaliteit van Leven | Arbeid.

- Zwinkels, W. (2005). *Prestaties van reïntegratiebedrijven vergeleken: Benchmarking van effectiviteit*. In Vos, E.L. de & van Genabeek, J. (red.), *Verraderlijk effectief: prestatie meting van reïntegratie en activering*, deel 2, pp. 61-93. Amsterdam: PlantijnCasparie.
- Zwinkels, W., A. Hekelaar en A. Braat (2006), *De juiste klant op het juiste traject: een onderzoek naar de netto-effectiviteit van het Rotterdamse reïntegratiebeleid voor het ontwikkelen van klantprofielen*, Hoofddorp: TNO.

Bijlage 2: Begrippenlijst

Afroming

Het bewust inzetten van interventies bij de groep waar de kans op het bereiken van de doelstelling het grootst is.

Beleidsinstrumenten / beleidsinterventies

Instrumenten die worden ingezet om bepaalde beoogde beleidsdoelen te realiseren.

Benchmark

Vergelijking van uitkomstmaten voor groepen van cliënten, uitvoerders of aanbesteders.

Bruto-effecten

Bruto-effecten zijn gelijk aan het (maatschappelijke) resultaat, waarop de beleidsinspanningen zijn gericht. Bruto-effecten worden ook wel aangeduid als *outcome*. Bruto-effecten zijn het resultaat van effecten van beleid en effecten van externe factoren.

Duurafhankelijkheid

De invloed van de tijd en daarmee samenhangende invloedsfactoren op een uitkomstmaat, zoals de invloed van de werkloosheidsduur op de kans om alsnog een baan te vinden.

Deadweight loss

Het verlies dat optreedt door de inzet van interventies in gevallen waar ook zonder de interventie de doelstelling(en) van het beleid zouden zijn behaald.

Duurmodel

Een duurmodel verklaart de duur van perioden (bijvoorbeeld de werkloosheidsduur of baanduur) uit verschillende invloedsfactoren, zoals kenmerken van personen, de inzet van beleidsinstrumenten (effecten), de reeds verstreken duur zelf en toevallige gebeurtenissen. Het meenemen van toevalligheden maakt een duurmodel tot een kansmodel, waarbij de kans op uitstroom uit duurperioden wordt gemodelleerd.

Effectiviteitsanalyse

Met een *effectiviteitsanalyse* wordt vastgesteld of de inzet van een interventie toegevoegde waarde heeft bij het bereiken van de gestelde doelen. Vinden mensen bijvoorbeeld vaker een baan nadat zij een re-integratietraject hebben gevolgd?

Effectenonderzoek ex-post

Het achteraf op een systematische en methodologisch verantwoorde wijze onderzoeken en beoordelen van de effecten van ingezet beleid in de maatschappij of (overheids)organisatie.

Evaluatieonderzoek ex ante

Een systematische analyse vooraf van de te verwachten maatschappelijke effecten van beleidsalternatieven in relatie tot de maatschappelijke kosten.

Externe factoren

Exogene factoren die het behalen van de beleidsdoelstelling beïnvloeden, maar die buiten het te evalueren beleid vallen.

Gevoeligheidsanalyse

Analyse van de uitkomsten van een evaluatie waarbij de verschillende modelveronderstellingen worden gevarieerd.

Hazard rate

Kans op uitstroom uit een duurperiode (bijvoorbeeld werkloosheid of een baan) die conditioneel is op de reeds verstreken duur van die periode.

Input

Input (middelen) bestaande uit apparaatskosten, waaronder materiële en personele uitgaven (inclusief inhuur) en beleidsuitgaven voor beleidsprogramma's (zoals subsidies, leningen en voorlichtingsprogramma's).

Kosten-batenanalyse

Vergelijking van input en effecten, waarbij beiden zoveel mogelijk in geld uitgedrukt worden. Hiermee kan worden nagegaan hoe de opbrengsten van een beleidsinstrument zich verhouden tot de kosten ervan.

Kosteneffectiviteitanalyse

Met een kosteneffectiviteitanalyse kan de inzet van verschillende instrumenten worden vergeleken op de verhouding tussen kosten en effectiviteit. De effecten van het beleid(sinstrument) worden in een kosteneffectiviteitanalyse afgezet tegen de gemaakte kosten van de inputs.

Matching

De koppeling van personen in de proefgroep met personen in de referentiegroep op basis van allerlei kenmerken. Met matching wordt een controlegroep samengesteld die zo veel mogelijk lijkt op de proefgroep.

Netto-effecten

De toegevoegde waarde van de inzet van beleid op het behalen van de beleidsdoelstellingen. Het netto-effect is gelijk aan de outcome (het bruto-effect) gecorrigeerd voor externe factoren.

Neveneffecten

Gevolgen in maatschappij of overheidsorganisatie die weliswaar voortvloeien uit de inzet van beleidsinstrumenten, maar die niet primair zijn beoogd.

Niet waargenomen heterogeniteit

Verschillen in de uitkomstmaat tussen de proefgroep en controlegroep die wordt veroorzaakt door de invloed van factoren die niet (kunnen) worden waargenomen.

Proefgroep, controlegroep en referentiegroep

De proefgroep is de groep die wel te maken heeft met een bepaalde interventie en de controlegroep is de groep die niet te maken heeft met die interventie en waarmee de proefgroep wordt vergeleken. Een controlegroep kan worden samengesteld uit een grotere groep die niet te maken heeft met de interventie, de zogenaamde referentiegroep.

Profiling

Het onderscheiden van groepen binnen een populatie naar de mate van de verwachte uitkomstmaat, onder de veronderstelling dat beleid een verschillende uitwerking heeft per groep.

Propensity Score

De kans op een interventie (kans om deel uit te maken van de proefgroep), verklaard uit waargenomen kenmerken.

Referentiepersoon

Voor de presentatie van effecten kan worden uitgegaan van een specifiek persoon met bepaalde kenmerken (bijvoorbeeld gemiddelde kenmerken of de meest voorkomende kenmerken), waarvoor de uitkomstmaat met en zonder beleidsinterventie wordt berekend. Deze specifieke persoon is de referentie-persoon voor de presentatie van effecten. De invloed van specifieke kenmerken op de uitkomstmaat kan worden gepresenteerd door de referentiepersoon wel en niet deze kenmerken mee te geven.

Re-integratieladder

Als tussenstap naar deelname aan reguliere arbeid of beëindigen van werkloosheid wordt vaak gewerkt met de (stapjes op de) re-integratieladder.

Samenstellingseffecten

Verschillen in de uitkomstmaat tussen de proefgroep en controlegroep die direct worden veroorzaakt door verschillen in de samenstelling van *proefgroep en controlegroep*.

Selectie-effecten

Verschillen in de uitkomstmaat die worden veroorzaakt door een niet-willekeurige inzet van interventies, ofwel een niet-willekeurige toewijzing aan de *proefgroep en controlegroep*.

Selectieve groep

Deelnemers aan het re-integratietraject (de proefgroep) vormen een groep die niet zonder meer te vergelijken is met de groep die niet deelneemt aan een re-integratietraject (de controlegroep) als gevolg van samenstellingseffecten en selectie-effecten.

Targeting

Inzet van beleidsinstrumenten bij de groep waar de kans op het bereiken van de doelstelling het grootst is.

Uitkeringsduur, werkloosheidsduur en baanzoekduur

Mensen kunnen tijdens de uitkeringsduur of werkloosheidsduur op zoek zijn naar een baan. Het begin van de baanzoekduur hoeft niet perse samen te vallen met het begin van de uitkeringsduur of werkloosheidsduur.

Uitkomstmaten

Een operationalisatie van beleidsdoelstellingen waarbij effecten van beleid uitgedrukt worden in uitkomstmaten. Deze uitkomstmaten worden gebruikt bij het vaststellen van de netto-effectiviteit.