



Centraal Planbureau

CPB Notitie | 15 juli 2016

Haalbaarheid (quasi-) experimentele evaluatie ESF-activiteiten

*Op verzoek van het Agentschap
van het ministerie van Sociale
Zaken en Werkgelegenheid*



CPB Notitie

Aan: Agentschap SZW

Datum: 15-7-2016

Betreft: Notitie Haalbaarheid (quasi-)experimentele evaluatie ESF-activiteiten

English summary (Engelse samenvatting)

The European Union (EU) distributes ESF subsidies among all its membership countries. Between 2014 and 2020 the Netherlands is allowed to spend an amount of 507 million euros. ESF subsidies are implemented by Agentschap SZW, which is part of the Ministry of Social Affairs and Employment.

The Netherlands spends the subsidy received on the following themes within the ESF program:

- **Active Inclusion:** to give those who have drifted far away from the labour market the chance to get back into gainful employment. Low-qualified jobseekers, people from immigrant backgrounds, jobseekers with disabilities, mothers returning to work after long breaks, ex-prisoners and young people without qualifications are examples of the groups being targeted by 'active inclusion' measures.
- **Supporting working lives:** against a background of an ageing population, The Netherlands wants to give people opportunities to work for longer. ESF projects are helping old and young workers resolve 'work vs health' issues and offering training opportunities in new skills to improve their job prospects. Employers are playing an important role by adapting working practices to suit the needs of workers of all ages.

The European Commission requires Member States to examine the effectiveness of ESF funded activities. One of the preferred methods is the method of counterfactual impact evaluation, an evaluation type which is based on the principle of a control and treatment group. This requires a (quasi-)experimental design of the study.

The CPB Netherlands Bureau for Economic Policy Analysis has been asked by Agentschap SZW to investigate whether it is feasible to conduct a (quasi-)experimental evaluation of ESF activities. In particular it has been asked whether it is feasible to conduct a (quasi-)experimental evaluation of:

1. The effectiveness of the policy instruments that are financed or co-financed by the European Social Fund (ESF) in the Netherlands and
2. The net impact of the total amount of ESF resources in the Netherlands.

This report documents on these two questions. It addresses these questions in two ways. Firstly, we have asked the relevant authorities who use ESF subsidies whether there were opportunities to create randomized experiments in the future for evaluation of policies that are paid by the ESF. Secondly, we have examined whether there were opportunities for quasi-experimental evaluations of existing ESF policies. The latter has been done on the basis of two important conditions:

- The presence of exogenous variation in the policy instrument. This means that there is no selection or self-selection of participants in the policy based on unobservable characteristics. Hence, it is possible to create valid control group.
- A sufficient sample size. On the basis of a power analysis for the active labour market policies assessed in this report, we have applied a minimum of 870 participants.

The conclusion to the first question is that (quasi-)experimental evaluation only seems feasible for a part of the policy instruments funded by the ESF. The Employee Insurance Agency (UWV) endorses the importance of experimental evaluation and is prepared to examine the possibilities for experimental research. It is important that a sufficiently large sample size will be chosen, and that the experiment will be conducted carefully. The group that is targeted by ESF funds is a vulnerable group, who receives benefits from the Disablement Assistance Act for Handicapped Young Persons (Wajong).

Within the other authorities/bodies who obtain ESF subsidies, i.e. the Labour Market Regions and the Ministry of Security and Justice, (quasi-)experimental research seems unfeasible. At present, there appear to be no opportunities for conducting randomized experiments within these authorities/bodies. Also, quasi-experiments with existing policies seem unpromising. The main reasons are the lack or absence of exogenous variation in the policy instruments and the sample size. Although some policies in the Labour Market Regions seemed to have exogenous variation in the assignment of participants to the policies, the sample size turned out to be too small for impact assessment. The number of participants in these instruments was below the required minimum thresholds set in this report.

The conclusion to the second question – the net impact of EFS resources – is inherently related to that of the first. As most of the individual ESF funded policy instruments cannot be evaluated by a (quasi-)experimental design, it is not possible to evaluate the total amount of the ESF in a (quasi-)experimental way. After all, the total amount consists of a large number of components that cannot be assessed by a (quasi-)experimental design.

The fact that this is not possible does not mean that there is no information regarding the policies that are co-funded by the ESF. Meta-analysis studies of active labour market policies such as Card et al. (2015) and Effective Labour Market Policies (in Dutch: Kansrijk Arbeidsmarktbeleid, deel 2 (CPB, 2016)) may give indications about the effectiveness of the policies co-funded by the ESF. In addition, policies can also be evaluated by non-experimental techniques and research methods. This report was only concerned with the feasibility of *(quasi-)experimental* evaluation.

Nederlandse samenvatting

Het Agentschap SZW heeft het CPB gevraagd te onderzoeken of het haalbaar is om met een (quasi-)experimentele methode

1. de effectiviteit van afzonderlijke beleidsinstrumenten die door het Europese Sociale Fonds (ESF) ge(co-)financierd worden, te evalueren.
2. de netto inzet van de ESF-middelen als geheel te evalueren.

Dit is op twee manieren gedaan. Ten eerste is bij de betrokken instanties (het Agentschap, de arbeidsmarktregio's, UWV en ministerie van Veiligheid en Justitie) nagegaan of er mogelijkheden zijn om *in de toekomst* gerandomiseerde experimenten op te zetten voor evaluatie van met ESF-geld betaalde beleidsinstrumenten. Ten tweede is onderzocht of er mogelijkheden zijn voor een quasi-experimentele evaluatie *achteraf* van reeds bestaande ESF-beleidsinstrumenten. Dat laatste is gebeurd aan de hand van twee belangrijke voorwaarden:

- De aanwezigheid van exogene variatie in het beleidsinstrument. Dit wil zeggen dat er geen selectie van deelnemers is in het beleidsinstrument op basis van niet-observeerbare kenmerken en dat er dus een valide controlegroep gevormd worden.
- Een voldoende grootte van de steekproefomvang. Daarbij is op basis van de in deze notitie gedane poweranalyses een minimum van 870 deelnemers gehanteerd.

Het antwoord op de eerste deelvraag is dat (quasi-)experimentele evaluatie slechts voor een deel van de met ESF-geld gefinancierde beleidsinstrumenten haalbaar lijkt.

Het UWV onderschrijft het belang van experimentele evaluatie en is bereid de mogelijkheden voor experimenteel onderzoek te onderzoeken. Daarbij is het van belang dat een voldoende grote steekproefomvang wordt gekozen en dat het experiment zorgvuldig wordt opgezet. Het gaat immers om een kwetsbare doelgroep (Wajongers) die vaak een relatief grote afstand tot de arbeidsmarkt kent. Het onthouden van beleidsinstrumenten in het kader van een experiment kan om die reden gevoelig liggen. Een manier om dit op te lossen is om in plaats van onthouding, een kandidaat altijd wel een (gepast) instrument aan te bieden, maar door loting te laten bepalen welk (gepast) instrument dat moet zijn. In dat geval kan de effectiviteit van de verschillende instrumenten tegen elkaar worden afgezet.

Bij de overige instanties lijkt (quasi-)experimenteel onderzoek niet haalbaar.

Bij de arbeidsmarktregio's, het ministerie van Veiligheid en Justitie en op het terrein van Duurzame Inzetbaarheid lijken er op dit moment geen mogelijkheden te zijn om in de nabije toekomst gerandomiseerde experimenten uit te voeren. Quasi-experimenten met bestaand beleid lijken ook niet kansrijk. De belangrijkste reden hiervoor is het gebrek aan of de afwezigheid van exogene variatie in de beleidsinstrumenten. De andere reden is de steekproefomvang. Bij sommige beleidsinstrumenten bij de arbeidsmarktregio's lijkt er potentieel te zijn voor een quasi-experimentele evaluatie, maar blijkt de steekproefomvang te klein voor een betrouwbare effectevaluatie. Het aantal deelnemers bij deze instrumenten ligt onder het in deze notitie gehanteerde minimum. Dit laat overigens onverlet dat buiten het onderzochte ESF-terrein effectiviteitsonderzoek bij arbeidsmarktregio's wel mogelijk kan zijn.

Het antwoord op de tweede deelvraag houdt verband met die op de eerste. Aangezien maar een deel van de met ESF-geld gefinancierde beleidsinstrumenten (quasi-) experimenteel te evalueren lijkt (namelijk bij UWV), is het niet mogelijk de netto inzet van ESF-middelen *in zijn geheel* op een *quasi-experimentele* manier te evalueren. De totale som van ESF-middelen bestaat immers uit een groot aantal onderdelen. Het evalueren van de middelen als geheel is daarom niet mogelijk.

Het feit dat dit niet mogelijk is wil overigens niet zeggen dat er geen kennis voorhanden is over de met ESF-subsidie gefinancierde beleidsinstrumenten. Evenmin betekent dit dat evaluatie met andere technieken niet mogelijk is. Metastudies over actief arbeidsmarktbeleid, zoals die van Card et al. (2015) en van het CPB (Kansrijk Arbeidsmarktbeleid, deel 2, 2016) geven indicaties over welke instrumenten voor welke doelgroep effectief zijn. Daarnaast kan ook met niet-(quasi-)experimentele technieken en methoden onderzoek worden gedaan. Deze notitie betreft immers alleen een haalbaarheidsadvies over het doen van *(quasi-)experimenteel* onderzoek.

1 Inleiding

Agentschap SZW heeft het CPB gevraagd te onderzoeken of het haalbaar is de (beleids)instrumenten die uit het Europees Sociaal Fonds (ESF) gefinancierd worden, op een (quasi-)experimentele manier te evalueren op hun effectiviteit. Daarnaast heeft het Agentschap SZW aan het CPB gevraagd of het haalbaar is om met een dergelijke methode te onderzoeken wat het netto-effect is van de inzet van de totale ESF-middelen.¹

Het Europees Sociaal Fonds (ESF) is het belangrijkste Europese subsidie-instrument voor arbeidsmarktbeleid. In de periode 2014-2020 is 507 miljoen euro beschikbaar voor Nederland. De twee belangrijkste prioriteiten (van in totaal vier) die het Nederlandse kabinet heeft geformuleerd voor besteding van dit geld zijn re-integratie van mensen met afstand tot de arbeidsmarkt ('Actieve Inclusie', 71% van budget) en het bevorderen van de Duurzame Inzetbaarheid van mensen (20% van budget).

Binnen 'Actieve Inclusie' kunnen centrumgemeenten van de 35 arbeidsmarktregio's (namens gemeenten en (ex-)leerlingen van VSO/PrO-scholen), UWV en het ministerie van Veiligheid en Justitie (V&J) financiering aanvragen voor projecten om mensen aan het werk te helpen en houden. Binnen 'Duurzame Inzetbaarheid' kunnen O&O-fondsen, sociale partners, bedrijven en (overheids)instellingen subsidie aanvragen voor het ontwikkelen en implementeren van *active ageing*-beleid. Een uitgebreidere beschrijving van het ESF is te vinden in bijlage A.

De opdracht die het Centraal Planbureau heeft gekregen van Agentschap SZW voor het onderzoeken van de haalbaarheid van quasi-experimenteel onderzoek is met name gericht op de eerste investeringsprioriteit, Actieve Inclusie, en daarnaast op de tweede investeringsprioriteit, Duurzame Inzetbaarheid.

De Europese Commissie verplicht lidstaten om de effectiviteit van met ESF-geld gefinancierde activiteiten te onderzoeken. Daarbij gaat voor een kwantitatieve evaluatie de voorkeur uit naar de methode van *counterfactual impact evaluation*, een type evaluatieonderzoek dat gebaseerd is op het principe van een controle- en behandelgroep. Dit vereist een (quasi-)experimentele opzet van het onderzoek. Deze notitie richt zich op de haalbaarheid van een dergelijk (quasi-)experimenteel opgezet onderzoek.

Met (netto-)effectiviteit wordt – afhankelijk van het beoogde (beleids)doel van het instrument – de verhoging van de arbeidsmarktparticipatie, verlaging van de

¹ Brief van Agentschap SZW van 20 november 2015, ref. 2015-0000278040

uitkeringsafhankelijkheid of bevordering van duurzame arbeidsinzet van werkenden bedoeld die (causaal) kan worden toegeschreven aan de inzet van het instrument.

In eerste instantie wordt de vraag beantwoord of het haalbaar is om de (beleids)instrumenten die gefinancierd worden vanuit het ESF (quasi-)experimenteel te evalueren. Bij experimenteel onderzoek worden individuen door de onderzoekers willekeurig aan de behandel – of controlegroep toegewezen om de effectiviteit van een instrument te toetsen. Quasi-experimenteel onderzoek maakt ook gebruik van een controle- en behandelgroep om de effectiviteit van een instrument te meten, maar benut daarvoor een op natuurlijke wijze ontstane (valide) controle- en behandelgroep. Daarna zal de vraag worden beantwoord of het haalbaar is om te onderzoeken wat het netto-effect is van de inzet van de ESF-middelen.

Als eerste stap om deze vragen te kunnen beantwoorden gaan we nader in op de uitdagingen bij het meten van effectiviteit, welke rol quasi-experimenteel onderzoek kan vervullen en wat de voorwaarden zijn om de effectiviteit van (beleids)instrumenten quasi-experimenteel te kunnen onderzoeken. Vervolgens bekijken we welke aanvragen voor de periode 2014-2020 voldoen aan deze voorwaarden. Derde stap in het proces is een toets op beschikbaarheid van alle benodigde gegevens om het effectiviteitsonderzoek, inclusief ondersteunende analyses,² uit te kunnen voeren.

2 Het meten van effectiviteit

Selectiviteit vormt het belangrijkste obstakel bij het onderzoeken van de netto-effectiviteit^{3,4} van een programma. Selectiviteit houdt in dat de deelnemers aan het programma niet willekeurig zijn aangewezen, maar dat zij deelnemen omdat zijzelf en/of de klantmanager van de gemeente of UWV verwachten dat de deelnemer hier baat bij zal hebben. Hierdoor gaat een eenvoudige vergelijking van de uitkomsten van deelnemers en niet-deelnemers mank: gevonden verschillen in uitkomst kunnen het resultaat zijn van het programma, van het verschil in kenmerken tussen de twee groepen of een combinatie hiervan.

Met econometrische methoden kan geprobeerd worden zo goed mogelijk te corrigeren voor verschillen in kenmerken tussen de behandel- en controlegroep. Er kan echter alleen gecontroleerd worden voor kenmerken die geobserveerd en

² Ondersteunende analyses zijn bijvoorbeeld tests of voldaan wordt aan in het (econometrische) model gemaakte aannames.

³ Met netto-effectiviteit van een programma wordt het verschil in uitkomst (hier: uitstroom naar regulier, betaald werk) bedoeld dat resulteert als gevolg van deelname aan het programma.

⁴ In de rest van deze notitie zal simpelweg 'effectiviteit' gebruikt worden in plaats van 'netto-effectiviteit'. Overal in deze notitie waar effectiviteit staat, kan ook 'netto-effectiviteit' gelezen worden.

geregistreerd worden. Als een of meerdere niet-geobserveerde (of geregistreerde) kenmerken (mede-)bepalend zijn voor deelname aan een programma en ook invloed hebben op de kans op uitstroom naar regulier werk, dan wordt de netto-effectiviteit niet zuiver gemeten. Het gemeten effect kan zowel een onderschatting als een overschatting van het werkelijke effect zijn.⁵

Een veel gebruikt voorbeeld van een niet-geobserveerd kenmerk is motivatie. Stel dat motivatie een bepalende factor is voor deelname aan een programma. Klantmanagers selecteren bijvoorbeeld gemotiveerde deelnemers, of werkzoekenden melden zichzelf aan voor het programma. Bij het vergelijken van de uitkomsten van deelnemers en niet-deelnemers kunnen de gevonden verschillen het gevolg zijn van het programma, maar ook worden veroorzaakt door het verschil in motivatie. Het gevonden verschil tussen de groepen kan dus niet geïnterpreteerd worden als het *causale* effect van het programma.

Idealiter zou men voor een effectevaluatie daarom middels willekeurige toewijzing willen bepalen wie er deelnemen aan een programma. Zo ontstaan er een behandelgroep en controlegroep die (gemiddeld) niet van elkaar verschillen, anders dan in deelname aan het programma. Alleen dan kunnen we uitsluiten (bij een voldoende grote steekproef) dat verschillen in uitkomst veroorzaakt worden door verschillen in kenmerken tussen de groepen en kunnen verschillen in uitkomst geheel worden toegeschreven aan het programma. Er wordt dus een *causaal* effect gemeten. In navolging van de medische wetenschap wordt een dergelijke evaluatie door willekeurige toewijzing toe te passen ook wel een experiment genoemd.

In de praktijk ligt een dergelijke willekeurige toewijzing politiek vaak gevoelig en is het ook organisatorisch (bijvoorbeeld door wisselingen onder klantmanagers) niet altijd makkelijk uitvoerbaar. Ook kan het resultaat van een evaluatie sneller gewenst zijn dan de tijd die het kost een experiment op te zetten, uit te voeren en te analyseren. Een mogelijke alternatieve wijze om toch de netto-effectiviteit van een programma te meten is door te zoeken naar een behandelgroep en controlegroep die op natuurlijke wijze ontstaan zijn, maar toch voldoen aan de eis dat ze gemiddeld niet of nauwelijks van elkaar verschillen op observeerbare en niet-observeerbare kenmerken. Dit worden ook wel 'natuurlijke experimenten' of 'quasi-experimenten' genoemd.

⁵ Of er sprake is van een overschatting of een onderschatting is afhankelijk van de wijze waarop de niet-geobserveerde kenmerken gecorreleerd zijn met deelname aan het programma en de kans op uitstroom naar werk.

3 Quasi-experimenteel onderzoek

3.1 Wat is quasi-experimenteel onderzoek?

Quasi-experimenteel onderzoek benut op natuurlijke wijze ontstane controle- en behandelgroepen, waarvan aannemelijk kan worden gemaakt dat ze voor aanvang van het programma (nagenoeg) niet van elkaar verschiden op (observeerbare en niet-observeerbare) kenmerken die van invloed zijn op de uitkomst waarin we geïnteresseerd zijn. Controle- en behandelgroepen kunnen op natuurlijke wijze ontstaan wanneer er sprake is van *exogene variatie* in de toewijzing van een instrument: een factor waar het individu of de klantmanager geen directe invloed op heeft (een *exogene* factor) is dan bepalend voor toewijzing aan een programma. Door deze exogene factor wordt een instrument soms wel en soms niet gebruikt voor individuen die (nagenoeg) gelijk zijn aan elkaar. Denk hierbij bijvoorbeeld aan een loting, de invoering van een nieuw programma, een wachtlijst voor deelname of toepassing van een strikte leeftijdsgrens voor deelname aan een programma.

Omdat quasi-experimenteel onderzoek gebruik maakt van op natuurlijke wijze ontstane controle- en behandelgroepen, is het ook mogelijk reeds ingevoerd beleid en beleid uit het verleden te evalueren. Bijkomend voordeel is dat hierdoor veel sneller antwoorden kunnen worden verkregen dan de doorlooptijd die een experimenteel opgezet onderzoek vergt.

Quasi-experimenteel onderzoek is een begrip waar een behoorlijke diversiteit aan onderzoeksstrategieën onder valt. Daarom bespreken we hieronder een aantal voorbeelden. Het is geen uitputtende lijst, maar geeft wel een idee van de diverse strategieën die mogelijk zijn. Bekende designs zijn *Regression-Discontinuity* en *Differences-in-Differences* designs.⁶

Een eerste strategie voor een quasi-experiment is gebruik maken van de situatie dat er meer mensen willen deelnemen aan een programma dan er plaatsen zijn, waarna er geloot wordt wie deel mag gaan nemen en wie niet. Door de uitkomsten van mensen die zijn 'ingeloot' te vergelijken met de uitkomsten van mensen die zijn 'uitgeloot' kan het effect van het programma worden gemeten. Het vormt hierbij geen probleem als de kans om ingeloot te worden afhangt van een of meer geobserveerde kenmerken, hiervoor kan gecorrigeerd worden. De loting of het lotnummer in de loting wordt gebruikt in de 'Instrumentele Variabele' analyse, waarbij de loting als 'instrument' fungeert.⁷

⁶ Zie bijvoorbeeld CPB-notitie 'Ontwerpen voor effectevaluatie' (2014).
<http://www.cpb.nl/sites/default/files/publicaties/download/cpb-notitie-2apr2014-ontwerpen-voor-effectevaluatie.pdf>

⁷ De term 'instrument' duidt hier op het econometrische begrip instrument. Dit staat volkomen los van de term (beleids)instrument dat in voorgaande paragrafen voorkwam.

Een tweede strategie voor een quasi-experiment is de situatie waarin een strikte grens/criterium is gesteld voor deelname aan een programma of regeling. Bijvoorbeeld een regeling waarvoor alleen mensen van 45 jaar of ouder in aanmerking komen. Door de uitkomsten van 44-jarigen (die niet in aanmerking komen) en 45-jarigen (die wel in aanmerking komen) te vergelijken kan het effect van de regeling worden gemeten. Hoewel leeftijd een effect kan hebben op de uitkomst, is de veronderstelling hier dat het leeftijdsverschil tussen een 44-jarige en een 45-jarige dusdanig klein is dat dit effect verwaarloosd kan worden. Hoe kleiner de gebruikte range van leeftijden, hoe aannemelijker deze veronderstelling is. Deze strategie werkt niet alleen met een leeftijds criterium, maar ook met een geografisch criterium, positie op een ranglijst als criterium, etc. Dit wordt ook wel een 'Regression Discontinuity Design' genoemd.

Een derde strategie is gebruik maken van het bestaan van een wachtlijst of wachttijd voor deelname aan het programma. Een wachtlijst kan ontstaan wanneer er meer aanmeldingen zijn voor een programma dan het huidige beschikbaar aantal plaatsen. Door een wachtlijst ontstaat een situatie waarin een deel van de mensen die geselecteerd zijn om deel te gaan nemen aan een programma daar niet direct mee kan aanvangen. Door mensen met eenzelfde werkloosheidsduur op moment van selectie voor het programma, maar met verschillende duur tot instroom in het programma te vergelijken, kan gemeten worden wat het effect van het programma is.

Een wachttijd kan ontstaan wanneer instroom in een programma slechts enkele malen per jaar mogelijk is. Sommige potentiële deelnemers worden kort voor de start aangemeld en kunnen vrijwel meteen beginnen; andere potentiële deelnemers worden aangemeld als een programma net gestart is en moeten daardoor wachten tot de start van een nieuwe cyclus van het programma. Door deze groepen te vergelijken gedurende de periode dat de een wel deelneemt en de ander niet, kan gemeten worden wat het effect van het programma is.

Een vierde strategie voor een quasi-experiment is gebruik maken van de introductie van een programma voor een subgroep. In een meting voorafgaand aan de introductie van het programma wordt voor zowel de controle als behandelgroep de uitkomstvariabele gemeten. Nadat vervolgens voor de behandelgroep een nieuw programma beschikbaar is gekomen (maar niet voor de controlegroep), wordt opnieuw de uitkomstvariabele gemeten voor de behandel- en controlegroep. Door individuen te vergelijken voor wie een programma niet beschikbaar is gekomen met individuen voor wie het programma wel beschikbaar is gekomen, kan het effect gemeten worden van de introductie van het programma. Door de voor- en nameting wordt gecorrigeerd voor algemene veranderingen over tijd (bijvoorbeeld conjunctuur). Het is daarom belangrijk dat de controle- en behandelgroep vóór de introductie van het nieuwe programma op dezelfde manier reageerden op algemene veranderingen (gezamenlijke trend veronderstelling). De econometrische methode die hierbij gebruikt wordt is de 'Difference-in-Difference' techniek.

3.2 Kanttekeningen bij en beperkingen van quasi-experimenteel onderzoek

Quasi-experimenteel onderzoek kan uitkomst bieden als experimenteel onderzoek moeilijk te realiseren is of wanneer er geen tijd is experimenteel onderzoek op te zetten. Maar quasi-experimenteel onderzoek kent ook een aantal beperkingen.

Een quasi-experiment op basis van een strikte grens/criterium levert als resultaat het effect op voor mensen die zich dicht bij die grens begeven. In het voorbeeld met een strikte leeftijdsgrens van 45 zoals hierboven, is het gevonden resultaat van de analyse het effect van het programma op mensen rond de 45 jaar. Welk effect het programma heeft op 25-jarigen of juist op 50-plussers kan op basis van deze ene leeftijdsgrens niet gemeten worden. De gevonden resultaten zijn in dat geval alleen van toepassing op een specifieke (sub)groep.⁸

Een quasi-experiment op basis van een wachtlijst of wachttijd gebruikt de periode waarin een deel van de mensen (door de wachtlijst/wachttijd) nog wachten op deelname aan een programma, terwijl anderen met een vergelijkbare werkloosheidsduur al wel begonnen zijn aan het programma om het effect van het programma te schatten. Hiermee kunnen we alleen de effectiviteit op de korte termijn schatten, omdat iedereen in de controlegroep immers na verloop van tijd aan het programma gaat deelnemen (of werk vindt). De maximale wachttijd bepaalt dus meteen de maximale termijn waarop we het effect nog kunnen meten.

Als de invoering van een nieuw programma voor een subgroep wordt gebruikt, is het belangrijk dat er in de periode tussen de voor- en nameting voor de controle of behandelgroep geen andere beleidswijzigingen zijn geweest. Stel dat er zich voor de behandelgroep tussen de voor- en nameting ook een andere beleidsverandering heeft voorgedaan, dan is niet duidelijk of het gemeten effect veroorzaakt wordt door de invoering van het programma, door de beleidswijziging of door beide en welke deel dan toe te schrijven is aan de invoering van het programma.

In een situatie waarin er niet expliciet sprake is van een experiment, en gegevens achteraf worden gebruikt in een quasi-experimenteel onderzoek, vinden er geen inspanningen plaats om de controlegroep 'schoon' te houden. Zo kan bij het bestaan van een wachtlijst voor een programma aan een werkzoekende een andere interventie aangeboden worden. Dit speelt nog sterker als de introductie van een programma benut wordt als quasi-experiment. Voordat een programma beschikbaar was, werden werkzoekenden mogelijk op andere manieren geactiveerd of werd hun een ander programma aangeboden. Het gemeten effect in het quasi-experiment is

⁸ Soms is er sprake van meerdere leeftijdsgrenzen die benut kunnen worden, waardoor een beeld voor de gehele (of in ieder geval een bredere) groep verkregen kan worden.

hierdoor niet het effect van deelname aan het programma ten opzichte van niet-deelname aan het programma, maar het effect van deelname aan het nieuw geïntroduceerde programma versus deelname aan een van de andere programma's. Er kan dus nog steeds een effectiviteitsmeting worden gedaan, maar de interpretatie verandert, omdat de uitgangssituatie waarmee je vergelijkt anders is.

Een gerelateerde kanttekening bij quasi-experimenteel onderzoek is dat er mogelijke spillovers hebben plaatsgevonden naar de controlegroep. Spillovers treden op als door de behandeling van de behandelgroep er ook veranderingen optreden in wat er met de controlegroep gebeurt. Een voorbeeld hiervan is wanneer de 'treatment' van de behandelgroep meer inzet van klantmanagers vraagt, waardoor zij minder tijd hebben om aan hun overige klanten (waaronder te controlegroep) te besteden. Het effect van de treatment zal hierdoor overschat worden. Maar dit effect kan ook andersom werken, als klantmanagers door de treatment (bijvoorbeeld deelname aan een nieuw re-integratieprogramma) meer tijd overhouden om aan hun overige klanten te besteden. In dat geval zal het effect van het nieuwe programma onderschat worden.

Ten slotte: een quasi-experimentele opzet is zo sterk als de exogene variatie is.

3.3 Hoe beoordelen we de haalbaarheid voor quasi-experimenteel onderzoek?

In deze notitie beoordelen we of het haalbaar is om met een (quasi-)experimentele methode

- De afzonderlijke beleidsinstrumenten die door het ESF-gecofinancierd worden te evalueren.
- De netto inzet van de ESF-middelen als geheel te evalueren.

Dit doen we op twee manieren. Ten eerste onderzoeken we of er bij de betrokken instanties mogelijkheden zijn *om in de toekomst* gerandomiseerde experimenten op te zetten. Ten twee onderzoeken we of er mogelijkheden liggen voor een quasi-experimentele evaluatie *achteraf*. We doen dat aan de hand van twee belangrijke voorwaarden die in de vorige paragrafen zijn besproken:

1. Exogene variatie in het beleidsinstrument. Dit wil zeggen dat er geen selectie van deelnemers is in het beleidsinstrument op basis van niet-observeerbare kenmerken. Er kan dus een valide controlegroep gevormd worden (zie paragraaf 3.1)
2. Voldoende grootte van de steekproefomvang.

Ad 1) Als ESF-geld of het beleidsinstrument selectief wordt ingezet waardoor er geen vergelijkbare controlegroep gevormd kan worden, dan valt het beleidsinstrument af voor een evaluatie met quasi-experimenteel design.

Ad 2) Als het beleidsinstrument zodanig wordt ingezet dat er wel een mogelijkheid is voor quasi-experimentele evaluatie maar de steekproefomvang te klein, dan valt het instrument ook af voor een evaluatie met quasi-experimenteel design. Met te weinig data is er te weinig statistische kracht (*power*) om betrouwbare uitspraken te kunnen doen. Bij de beoordeling van de haalbaarheid voor (quasi-)experimenteel onderzoek hanteren we in deze notitie 870 deelnemers als minimum (zie onder voor toelichting).

Power

De power is de kans dat je met de empirische analyse-effecten vindt van de beleidsinterventie als effecten aanwezig zijn. Strikt genomen is het de kans dat je de nulhypothese van geen effect verwerpt als er wel een effect is. Daarbij spelen twee zaken een rol: het werkelijke (niet waargenomen) effect van het beleidsinstrument en de steekproefomvang. Hoe groter het werkelijke effect van het beleidsinstrument, hoe hoger de power. En: hoe groter de steekproefomvang, hoe hoger de power. Bij een powerberekening wordt uitgerekend hoe groot de steekproefomvang moet zijn om bij een gegeven effect van het beleidsinstrument een bepaalde power te bereiken (bijvoorbeeld een power van 80%). Het maken van een dergelijke berekening is belangrijk, omdat zonder kennis hiervan men het gevaar loopt met een te kleine steekproef het (quasi-)experiment te starten. In een dergelijk geval kan geen effect van de interventie aangetoond worden, terwijl het er wel is. Om te beoordelen hoe groot de steekproefomvang moet zijn voor de in deze notitie aangehaalde projecten voeren we daarom een powerberekening uit.

Powerberekening

Voor onze berekening houden wij een power van 80% aan. Dit percentage wordt in het algemeen als uitgangspunt genomen voor wetenschappelijk onderzoek. Voor de grootte van het effect gaan we uit van de effectgroottes die in de wetenschappelijke literatuur gevonden worden van actief arbeidsmarktbeleid (zoals re-integratietrajecten, jobcoaching etc.). We gebruiken daarbij de overzichtsstudie van Card et al. (2015) als leidraad. Deze studie laat zien dat kortetermijneffecten (<1 jaar) van actief arbeidsmarktbeleid relatief klein zijn en dat middellange- (1> jaar <2) en langetermijneffecten (>2 jaar) doorgaans wat groter zijn. Het gemiddelde kortetermijneffect is 0.04 standaarddeviatie (sd), en het middellange- en langetermijneffect zijn 0.12 en 0.19 sd. Ter indicatie van het begrip standaarddeviatie: stel dat van de bijstandsgerechtigden 50% na 2 jaar een baan vindt zonder re-integratietraject, en dit percentage wordt verhoogd naar 60% door een ESF gefinancierd re-integratietraject, dan is het (langetermijn)effect van dit traject 10%-punt (=60%-50%). Dit komt overeen met ongeveer 0.20 sd. Uitgaande van een power van 80% en de effectgroottes 0.04, 0.12 en 0.19 komen we op een

minimale steekproefomvang van 19600, 2200 en 870 respectievelijk. Voor het vinden van een relatief groot (langetermijn)effect zijn dus op zijn minst 870 deelnemers nodig van wie 435 in de behandelgroep en 435 in de controlegroep zitten. Voor het vinden van kleinere effecten zijn dus nog veel meer deelnemers nodig. Bij de beoordeling van de haalbaarheid voor (quasi-)experimenteel onderzoek hanteren we in deze notitie 870 daarom als minimum. Dit is waarschijnlijk een conservatieve schatting. In de powerberekeningen zijn we uitgegaan van een gerandomiseerd experiment. Quasi-experimenten vergen doorgaans grotere aantallen, omdat ze bijvoorbeeld aanvullende controlevariabelen nodig hebben die de power verlagen (zoals de 'forcing variable' bij een *Regression-Discontinuity* (RD-) design).

3.4 Wat als er geen (quasi-)experimentele methode gebruikt kan worden?

In deze paragraaf bespreken we een aantal mogelijke analyses en de interpretatie van resultaten die verkregen worden met niet-(quasi-)experimentele methoden als ze ingezet worden voor een effect-evaluatie. Wederom is de lijst niet uitputtend, maar beoogt zij de belangrijkste methoden op een rij te zetten.

We beginnen met een methode waarbij geen sprake is van een bepaalde mate van willekeurigheid in de inzet van een (beleids)instrument waardoor er een controle- en behandelgroep ontstaat, maar waarbij sprake is van willekeurigheid in het *moment* waarop het instrument wordt ingezet (Abbring en Heckman, 2007). Dit is het geval wanneer individuen iedere periode *een kans hebben* dat het instrument wordt ingezet. Hierdoor kunnen individuen die (vrijwel) gelijk aan elkaar zijn, toch erg verschillen in het moment waarop het instrument ingezet wordt (zelfs met eenzelfde kans op het instrument in elke periode kan het instrument bij het ene individu al in periode 1 ingezet worden en bij het andere individu pas in periode 100). In deze specifieke gevallen is het mogelijk de 'timing of events'-methode te gebruiken, die wél rekening houdt met selectie, ook als dat gebeurt op voor de onderzoeker niet-geobserveerde kenmerken. Bij 'timing of events' worden gelijktijdig twee duurmodellen geschat, één voor de duur tot uitstroom naar werk en één voor de duur tot het starten van de 'behandeling', waarbij de duur tot uitstroom afhangt van de duur tot start van de behandeling. Door ze gelijktijdig te schatten wordt er rekening mee gehouden dat ook niet-geobserveerde kenmerken ervoor kunnen zorgen dat deze duren gerelateerd zijn. Een belangrijke voorwaarde om deze methode te kunnen gebruiken is dat de start van de behandeling niet (of heel kort) vooraf bekend is bij het individu. Deze voorwaarde wordt ook wel de 'non-anticipatie veronderstelling' genoemd. Alleen als geloofwaardig beargumenteerd kan worden dat hieraan voldaan is, geeft deze methode correcte schattingen.

Als er geen bron van exogene variatie is waardoor er een natuurlijk experiment ontstaat en timing-of-events ook geen optie is, is het nog steeds mogelijk

econometrische analyses uit te voeren. De resultaten kunnen dan echter vaak niet causaal geïnterpreteerd worden omdat ze vertekend worden door selectie-effecten. Het gemeten effect geeft dan een correlatie weer. Met deze beperking in het achterhoofd zijn de resultaten van de analyses bruikbaar in een beschrijvende studie.

Met (propensity score) matching kunnen selectie-effecten geschoond worden van selectie-effecten veroorzaakt door voor de onderzoeker observeerbare kenmerken. Voor selectie-effecten veroorzaakt door voor de onderzoeker niet-observeerbare kenmerken kan hiermee echter niet gecorrigeerd worden. Uit de literatuur is bekend dat juist niet-observeerbare kenmerken vaak een belangrijke rol spelen, zowel bij de toewijzing aan een programma als bij uitstroom naar werk. Veelal is er wel degelijk een reden dat een klantmanager de ene werkzoekende wel en de andere werkzoekende juist niet aan een programma heeft toegewezen, zelfs als deze werkzoekenden dezelfde leeftijd, opleidingsniveau, ervaring, etc. hebben. Een voorbeeld van een vaak niet-geobserveerd kenmerk dat erg belangrijk kan zijn bij zowel toewijzing aan een programma als uitstroom naar werk is motivatie. Resultaten van matching kunnen niet causaal geïnterpreteerd worden als matching wordt gebruikt zonder dat er een bron van exogene variatie is, of er niet overtuigend beargumenteerd kan worden dat alle variabelen geobserveerd worden die een rol kunnen spelen bij (zelf-)selectie in de groep bij wie het instrument ingezet wordt.⁹

Wat weten we uit eerder onderzoek?

In diverse landen is onderzoek gedaan naar de netto-effectiviteit van actief arbeidsmarktbeleid. De afgelopen jaren zijn ook een aantal metastudies gedaan die een interessant totaalbeeld geven van de effectiviteitsstudies gedaan in diverse landen (Card et al., 2010; Kluve, 2010; Card et al., 2015). Deze metastudies selecteerden onderzoeken zorgvuldig op de manier waarop zij omgaan met selectiviteit. Opvallend is dat de resultaten uit verschillende landen vaak tot vergelijkbare inzichten leiden. De gevonden effecten (op uitstroom naar werk) zijn over het algemeen klein, waarbij er wel een verschil bestaat naar welke termijn gekeken wordt. Veelal zijn de effecten op de langere termijn groter dan op korte termijn. Daarnaast zijn er ook verschillen tussen typen instrumenten. Begeleiding bij het zoeken naar werk en handhaving op voldoende zoekinspanningen hebben daarbij relatief grote effecten op korte termijn die daarna nauwelijks groter worden, terwijl training en scholing weinig (en soms zelfs een negatief) effect hebben op korte termijn, dat juist groter (en positief) wordt op de langere termijn. Ook zijn er verschillen tussen groepen: begeleiding en handhaving zijn het meest effectief voor mensen met een korte afstand tot de arbeidsmarkt, terwijl loonkostensubsidies het meest effectief zijn voor mensen met een grote afstand tot de arbeidsmarkt.

Met regressietechnieken als multivariate analyse en logistische regressie kan in de afwezigheid van exogene variatie geen causaal effect gemeten worden. De gevonden effecten zijn vertekend door selectiviteit, zoals beschreven in hoofdstuk 2: er is niet langer te onderscheiden welk deel van het geobserveerde verschil tussen de groep met en zonder instrument veroorzaakt wordt door het instrument en welk deel het

⁹ Zie bijvoorbeeld DiNardo en Lee (2011) voor een discussie over de interpretatie van resultaten van (propensity score) matching bij effect-evaluatie.

gevolg is van verschillen in kenmerken tussen de twee groepen. De effecten moeten daarom geïnterpreteerd worden als correlaties.

4 Inventarisatiemethode van ESF-activiteiten

Om instrumenten te kunnen beoordelen op haalbaarheid voor een quasi-experimentele methode moeten we eerst een inventarisatie maken van de beleidsinstrumenten die door het ESF worden gefinancierd en hoe zij zijn geïmplementeerd. We inventariseren de ESF-activiteiten die zich richten op het thema Actieve Inclusie en Duurzame Inzetbaarheid. Op het gebied van Actieve inclusie inventariseren we de ESF-activiteiten voor de arbeidsmarktregio's (gemeentelijke en VSO/PrO-projecten), het UWV en V&J afzonderlijk. Op het gebied van duurzame inzetbaarheid inventariseren we de ESF-subsidies die worden verstrekt aan Opleiding en Ontwikkelings (O&O-)fondsen en aan bedrijven.

Voor de inventarisatie bij de arbeidsmarktregio's hebben we gebruik gemaakt van een enquête van Regioplan/Bureau Bartels waarbij elke arbeidsmarktregio is bevraagd over hun ESF-activiteiten. Naast de vragen van Regioplan/Bureau Bartels, heeft het CPB twee vragen toegevoegd aan deze enquête. De eerste vraag betrof een vraag over wel/niet aanwezigheid van een wachtlijst bij (sub)projecten in hun regio. De tweede vraag betrof een vraag over de wel/niet aanwezigheid van een leeftijdsgrens bij (sub)projecten in hun regio.¹⁰ Op deze manier zou een mogelijkheid kunnen ontstaan voor een quasi-experimentele evaluatie (zie hoofdstuk 3). Naast deze enquête zijn er diverse gesprekken gevoerd met betrokkenen binnen de arbeidsmarktregio's om aanvullende informatie te krijgen over bepaalde activiteiten.¹¹ Daarbij zijn aanvullende vragen gesteld over de selectiecriteria van de ingezette beleidsinstrumenten en het aantal beoogde deelnemers.

Voor de inventarisatie van de activiteiten bij UWV en V&J zijn gesprekken gevoerd met diverse betrokkenen.¹² In deze gesprekken zijn vragen gesteld over de mogelijkheden tot opzetten van gerandomiseerde experimenten en de invoering van verschillende beleidsinstrumenten. Bij V&J is gekeken of de invoering zodanig was dat er vergelijkbare controlegroepen gevormd kon worden, zoals besproken in

¹⁰ Wij danken Regioplan/Bureau Bartels voor het toevoegen van deze vragen en het ter inzage geven van hun gegevens.

¹¹ Wij danken de betrokkenen bij de arbeidsmarktregio's Breda, Helmond, Zwolle, Heerlen, Haaglanden en Rotterdam.

¹² Het gesprek bij het UWV heeft plaatsgevonden op 21 juni 2016 bij het UWV. Het gesprek met V&J (Dienst Justitiële Inrichtingen) heeft plaatsgevonden op 26 april 2016 bij het Agentschap SZW. Voortkomend uit dat gesprek is het visiedocument ESF 2014-2020 aan ons verstrekt waarin informatie staat over de inzet van ESF-trajecten bij (ex-)gedetineerden.

paragraaf 3.1. Daarbij zijn vooral vragen gesteld over de selectiecriteria (zoals wel/geen leeftijdsgrens) van het beleidsinstrument en de aanwezigheid van wachtlijsten om voor het beleidsinstrument in aanmerking te komen.

Voor de inventarisatie van de ESF-subsidies naar de sectorfondsen en bedrijven in het kader van duurzame inzetbaarheid hebben we ons beroepen op gesprekken met het Agentschap SZW en documentatie van ESF over de subsidie. Daarbij zijn vooral vragen gesteld over aantal aanvragen van bedrijven/sectoren, en capaciteitsrestricties in het budget (dat wil zeggen onder- of overvraging).

5 Analyse

5.1 Thema Actieve Inclusie

5.1.1 Arbeidsmarktregio's

In Nederland krijgen 35 arbeidsmarktregio's geld van het ESF. Dit geld gaat naar beleidsinstrumenten van (centrum)gemeentes en VSO/PrO-scholen die als doel hebben actieve inclusie te bevorderen. We onderzoeken twee routes voor een quasi-experimenteel design.

Een eerste route tot (quasi-)experimentele evaluatie van deze gelden zou zijn als er een bron van exogene variatie zit in de manier waarop het geld over de verschillende arbeidsmarktregio's wordt verdeeld. Dit wil zeggen dat de ene regio dan door toeval meer geld krijgt dan de ander. In dat geval zou het effect van de ESF-middelen bepaald kunnen worden door de uitstroom naar werk in regio's die meer ESF-middelen krijgen te vergelijken met de regio' die geen of minder ESF-middelen krijgen. Deze opzet lijkt echter lastig. In de praktijk wordt het geld volgens een verdeelsleutel over de arbeidsmarktregio's verdeeld. Daarbij zijn keuzes gemaakt waarom bepaalde regio's meer moeten krijgen dan anderen. Dit betekent dat de regio's die meer ESF-middelen krijgen onvergelijkbaar zijn met regio's die minder van dergelijke middelen krijgen. Daarnaast is er nog een ander probleem. Er kan namelijk substitutie/vervanging van middelen plaatsvinden. Als een regio minder ESF-middelen krijgt dan een andere regio, dan kan deze regio het gemis aan ESF-middelen compenseren met middelen uit een andere financieringsbron dan het ESF. Per saldo kan er dan evenveel aan actieve inclusie worden gedaan in de verschillende regio's waardoor het effect van de inzet van ESF-middelen niet goed bepaald kan worden.

Een tweede route is te kijken naar de mogelijkheid tot (quasi-)experimentele evaluatie van (deel)projecten die binnen de arbeidsmarktregio's met ESF-middelen worden ge(co)financierd.

Zoals gezegd, heeft Regioplan/Bureau Bartels een enquête uitgezet bij (de accounthouders van) deze regio's om informatie te krijgen over de door de ESF-gefinancierde activiteiten binnen de regio's. Het CPB heeft inzicht gekregen in deze enquête om te onderzoeken of er mogelijkheden lagen voor een quasi-experimentele evaluatie achteraf. Tot op heden blijkt er weinig draagvlak te zijn bij de arbeidsmarktregio's om in de toekomst gerandomiseerde experimenten op te zetten voor de trajecten die door het ESF worden ge(co)financierd.¹³

Uit de enquête blijkt dat de variëteit van de ingezette beleidsinstrumenten zeer groot is, zowel binnen als tussen de regio's. Verspreid over de 35 arbeidsmarktregio's zijn er ongeveer 178 type beleidsinstrumenten die worden ingezet om de actieve inclusie te bevorderen (bron: Regioplan). Wij hebben in tabel B.1 in de appendix per arbeidsmarktregio aangegeven welke type instrumenten worden ingezet. Grofweg kan gesteld worden dat het overgrote deel van de met ESF-geld gefinancierde activiteiten binnen de gemeentes bestaan uit jobcoaching, sollicitatietrainingen, re-integratietrajecten en loonkostensubsidies, en dat het overgrote deel van de met ESF-geld gefinancierde activiteiten bij VSO/PrO-scholen bestaan uit leer-werktrajecten, branchecursussen en stages. In 10 arbeidsmarktregio's wordt bij (sub)projecten gewerkt met een wachtlijst en/of een leeftijdsgrens. Voor de overige 25 arbeidsmarktregio's geldt dat dit niet het geval is.

Dit betekent dat het voor de (sub)projecten in deze 25 arbeidsmarktregio's het waarschijnlijk erg lastig wordt een quasi-experimenteel design op te stellen. De hoofdreden hiervoor is dat klantmanagers van de (centrum)gemeenten deze trajecten vooral inzetten bij mensen/scholen waarvan hij/zij verwacht dat ze hier baat bij zullen hebben. Dit zorgt voor selectieve inzet van beleidsinstrumenten, waardoor het moeilijk wordt een vergelijkbare controlegroep te vormen voor een quasi-experimentele evaluatie (zie hoofdstuk 2). Op basis van de aan ons verstrekte gegevens hebben we geen bron van exogene variatie in deze 25 arbeidsmarktregio's kunnen vinden.¹⁴

In tabel 1 zoomen we verder in op de projecten waarbij wordt gewerkt met een leeftijdsgrens of wachtlijst die potentieel een bron van exogene variatie kunnen zijn.

¹³ Dit bleek o.a. uit gesprekken met Agentschap SZW en betrokkenen bij de arbeidsmarktregio's Haaglanden en Rotterdam.

¹⁴ We baseren ons op de aan ons verstrekte gegevens uit de enquête en welke zijn vrijgekomen in de gesprekken. We kunnen niet volledig uitsluiten dat er enkele trajecten binnen bepaalde AMR's zijn die weliswaar geen leeftijdsgrens of wachtlijst kennen, maar waarvan de vormgeving toch zodanig is dat er mogelijkheden liggen tot quasi-experimentele evaluatie.

Tabel 5.1 Inzoomen op projecten die een wachtlijst/leeftijdsgrens kennen

AMR	Wachtlijst	Project	Beoogd aantal deelnemers
Doetinchem*	Ja	Ervaar je baan, stagewerkplek leren	30
Zwolle	Ja	Zwolle Werkt Aanpak (ZWA), casemanagement	200
Helmond	Ja	Peelland anderstaligen, re-integratietrajecten voor anderstaligen	96-125
Roermond	Ja	Westrom re-integratietrajecten	Minder dan 790
AMR	Leeftijdsgrens	Project	Beoogd aantal deelnemers
Heerlen	Ja	Project 45+ Maastricht, re-integratietrajecten voor 45-plussers	150
Breda	Ja	Plaatsingssubsidies voor jongeren onder de 27, ouderen boven de 50	100
Nijmegen	Ja	Jongerenloket tot 27 jaar	Onbekend
Groningen	Ja	Jongerenvouchers voor jongeren tot 27 jaar	Onbekend
Tiel	ja	Re-integratietrajecten 50-plussers	100
Gorinchem	Ja	Jongerenloket tot 27 jaar	300

* Heeft ook leeftijdsgrens.

Er zijn vier arbeidsmarktregio's (AMR's) die projecten kennen waarbij gewerkt wordt met een wachtlijst: Doetinchem, Zwolle, Helmond en Roermond. Er zijn zeven AMR's die projecten kennen met een leeftijdsgrens: Heerlen, Breda, Nijmegen, Groningen, Tiel, Gorinchem en Doetinchem. Bij de andere AMR's wordt niet gewerkt met wachtlijsten of leeftijdsgrenzen, afgaande op de informatie die aan ons is verstrekt.

Op basis van een wachtlijst of leeftijdsgrens bieden de projecten in tabel 1 kansen voor een evaluatie met quasi-experimenteel design (zie paragraaf 3.1). Toch vallen alle projecten af voor een quasi-experimentele evaluatie. Daar zijn twee redenen voor: de eerste betreft de leeftijdsgrens voor 27 jaar voor bijstandsgerechtigden en de tweede de steekproefomvang.

In het eerste geval gaat het om beleidsinstrumenten waarvoor bijstandsgerechtigden in aanmerking komen die 27 jaar of ouder zijn. Het idee van een leeftijdsgrens is dat mensen die iets ouder zijn dan die leeftijdsgrens vergeleken kunnen worden met mensen die iets jonger zijn dan die grens. Echter, de regels om in aanmerking te komen voor de bijstand verschillen tussen mensen die jonger of ouder dan 27 jaar zijn. Dat wil zeggen dat de bijstandsgerechtigden van jonger dan 27 jaar een andere groep bijstandsgerechtigden is dan de groep ouder dan 27 jaar. Deze groepen zijn dus op de grens niet vergelijkbaar, waardoor aan het eerste criterium van valide controlegroepen niet wordt voldaan. Hierdoor vallen de projecten in Nijmegen, Groningen en Gorinchem af.

De tweede groep projecten die afvallen zijn de projecten die op het eerste gezicht een quasi-experimenteel design lijken te hebben, maar een te laag aantal deelnemers kennen. Een te kleine steekproefomvang zorgt ervoor dat je zelfs relatief grote effecten van een beleidsinstrument niet kan uitsluiten met een quasi-experimentele

evaluatie. Poweranalyses hierboven wijzen uit dat je op zijn minst 870 deelnemers nodig hebt voor het vinden van grote effecten. Hierdoor vallen Doetinchem (30 deelnemers), Zwolle (200 deelnemers), Helmond (96-125 deelnemers), Tiel (100 deelnemers), Heerlen (150 deelnemers) en Breda (100 deelnemers) af.¹⁵ Het aantal deelnemers bij deze projecten komt niet in de buurt van het in deze notitie gehanteerde minimum van 870. Roermond (<790 deelnemers) valt ook af, niet alleen op basis van het aantal deelnemers, maar ook op basis van het feit dat de wachttijd erg kort is (<3 weken) waardoor er *de facto* geen wachttijd is.

Een (quasi-)experimentele evaluatie lijkt dus op dit moment niet kansrijk bij de voor nu geïnventariseerde ESF-trajecten bij de arbeidsmarktregio's.

5.1.2 UWV

Het UWV in Nederland krijgt ook geld van het ESF. Dit geld gaat voornamelijk naar voorzieningen voor mensen met een Wajong-uitkering. Activiteiten die door het ESF worden gefinancierd zijn de voorziening van een Jobcoach en Re-integratietrajecten Wajong. Op dit moment worden kandidaten door klantmanagers van het UWV aan ESF- trajecten/beleidsinstrumenten toegewezen die het best bij hen past, waardoor een quasi-experimentele evaluatie achteraf lastig is (zie hoofdstuk 2). Echter, uit het gesprek met het UWV is naar voren gekomen dat UWV het belang van experimentele evaluatie onderschrijft en bereid is de mogelijkheden voor zo'n onderzoek te onderzoeken. Hiermee lijkt er draagvlak te zijn om in de toekomst gerandomiseerde experimenten op te zetten voor de evaluatie van met ESF-geld gefinancierde Wajong-instrumenten. Zoals gezegd, bevinden de met ESF-geld gefinancierde Wajong-instrumenten zich op het terrein van jobcoaching, en re-integratietrajecten. De kwaliteit van een evaluatie met een gerandomiseerd experiment zal zitten in de details en zorgvuldigheid waarmee het wordt opgezet. Het gaat immers om een kwetsbare doelgroep die een grote afstand kent tot de arbeidsmarkt. Het onthouden van beleidsinstrumenten in het kader van een experiment kan daarom gevoelig liggen. Een manier om dit op te lossen is om in plaats van onthouding een kandidaat altijd wel een instrument aan te bieden, maar door loting te laten bepalen welk instrument dat moet zijn. In dat geval kan de effectiviteit van de verschillende instrumenten tegen elkaar worden afgezet. Om te voorkomen dat door willekeur ongepaste instrumenten voor een kandidaat worden ingezet, kan men daarbij besluiten alleen te loten tussen een set instrumenten die de klantmanager met zorgvuldigheid heeft uitgekozen voor de Wajonger in kwestie.

Het voordeel van het doen van gerandomiseerde experimenten bij het UWV is dat in principe voldoende data voorhanden zijn. De minimumdrempel van 870 waarnemingen zal waarschijnlijk worden gehaald, omdat UWV een nationale

¹⁵ Daarmee is nog niet gezegd dat bij meer deelnemers dit goede quasi-experimentele designs zijn. Bij deze quick scan hanteren we de twee criteria die genoemd zijn in paragraaf 3.3. Als aan een van deze criteria niet wordt voldaan, valt het design af voor een quasi-experimentele evaluatie.

overheidsinstantie is waar alle mensen met een Wajong-uitkering geregistreerd staan.¹⁶ Bijkomend voordeel is dat de deelnemers een bsn-nummer hebben en daarmee zonder al te veel moeite geanonimiseerd gekoppeld kunnen worden aan (CBS-)bestanden met sociaaleconomische uitkomsten van deelnemers.

Een (quasi-)experimentele evaluatie van door het ESF gefinancierde trajecten lijkt dus kansrijk bij UWV.

5.1.3 Veiligheid en Justitie

Op het terrein van veiligheid en justitie gaan middelen uit het ESF naar re-integratie van deelnemers uit penitentiaire inrichtingen, jeugdinstituten en instellingen voor forensische zorg. Deelnemers hebben doorgaans een grote afstand tot de arbeidsmarkt. Deze trajecten of subsidies worden over het algemeen selectief ingezet. Dit valt ook te lezen in het visie-document ESF 2014-2020. Voor gedetineerden zijn er de volgende criteria:

- de gedetineerde is in staat en gemotiveerd om arbeid te verrichten;
- er is een uitgebreide intake gedaan;
- de doelstelling is dat de gedetineerde na zijn straf zal uitstromen naar een vorm van (begeleid) werk;
- op basis van de individuele wensen en mogelijkheden van de gedetineerde, waar nodig ondersteund door de beroepskeuze- en competentietest, is een (voorlopige) keuze gemaakt voor een beroepsrichting, waarin reële mogelijkheden voor arbeid voorhanden zijn;
- beschikt over een detentie- en re-integratieplan waarin de doelen en aandachtspunten met betrekking tot werk en/of scholing staan beschreven;
- is bereid tot deelname aan het ESF-project en is akkoord met de administratieve handelingen die aan het project verbonden zijn.

Forensische patiënten die geselecteerd worden moeten voldoen aan de volgende eisen:

- de verwachting is dat de patiënt na zijn behandeling kan uitstromen naar een vorm van (begeleid) werk;
- er is een uitgebreide intake gedaan, waar nodig aangevuld met een capaciteitentest en/of een assessment;
- de patiënt is in staat en gemotiveerd arbeid te verrichten;
- op basis van de individuele wensen en mogelijkheden van de patiënt is een (voorlopige) keuze gemaakt voor een van de geselecteerde beroepsrichtingen;

¹⁶ Ongeveer 17400 kandidaten stroomde in de Wajong in 2014 (bron: UWV Monitor Arbeidsparticipatie 2015)

- de patiënt beschikt over een behandelplan en een trajectplan waarin de doelen en aandachtspunten met betrekking tot werk en/of scholing staan beschreven;
- de patiënt is bereid tot deelname aan het ESF-project en is akkoord met de administratieve handelingen die aan het project verbonden zijn.

Deze selectiecriteria zorgen ervoor dat een quasi-experimentele evaluatie niet haalbaar is. Er is geen exogene variatie in het beleidsinstrument, aangezien gedetineerden en forensische patiënten zelf de hand hebben in selectie in ESF-trajecten (ze moeten ze gemotiveerd zijn). Er kan dus geen goede controlegroep van gedetineerden gevormd worden die net zo gemotiveerd is als de gedetineerden in de behandelgroep, maar door 'toeval' niet aan een ESF-traject mag deelnemen (zie ook hoofdstuk 2).¹⁷ Verder is uit het interview met V&J gebleken dat het ingewikkeld is om gerandomiseerde experimenten op te zetten ter evaluatie van ESF-trajecten (uit gesprek met V&J op 26 april 2016).

Een (quasi-)experimentele evaluatie van door het ESF gefinancierde trajecten lijkt op dit moment dus weinig kansrijk op het terrein van V&J.

5.2 Thema Duurzame Inzetbaarheid

Er gaat ongeveer 100 miljoen euro naar het thema Duurzame Inzetbaarheid. Bedrijven, samenwerkingsverbanden van sociale partners, overheidsinstellingen en O&O-fondsen kunnen aanvragen indienen bij Agentschap SZW voor projecten die duurzame inzetbaarheid bevorderen. De activiteiten die worden gefinancierd zijn gericht op ondersteuning van bedrijven en sectoren bij het ontwikkelen en implementeren van een *active ageing*-beleid in hun bedrijf en/of sector (*active ageing*-beleid is beleid dat ervoor moet zorgen dat mensen duurzaam inzetbaar blijven op de arbeidsmarkt).

Het is lastig de inzet van deze middelen (quasi-)experimenteel te evalueren. Dit komt omdat het geld selectief wordt ingezet: alleen die instellingen krijgen geld die een aanvraag doen en wier aanvraag gehonoreerd wordt. In principe zijn er twee routes mogelijk voor een (quasi-)experimentele evaluatie van deze gelden.

¹⁷ In het gesprek dat we met V&J hebben gevoerd op 26 april 2016 leek er in eerste instantie een mogelijkheid te bestaan voor een quasi-experimentele evaluatie. Er zou een 3 maanden straftermijn gelden: pas bij straffen vanaf 3 maanden zou een gedetineerde in aanmerking komen voor een ESF traject. In zo'n geval zou een vergelijking mogelijk zijn geweest tussen degenen die een straf hebben gekregen van meer dan 3 maanden en degenen die een straf hebben gekregen van minder dan 3 maanden. Bij navraag bleek dit echter niet zo te zijn 'De 3-maanden termijn [...] staat er bij nader inzien toch niet in. Er staat dat er sprake moet zijn van de verwachting dat het strafrestant (mede op basis van de uitspraak van de rechter in eerste aanleg) lang genoeg is om het traject succesvol te doorlopen, of over te dragen is aan de gemeente op basis van een deelresultaat.' Hierdoor vervalt dit als mogelijkheid voor een quasi-experimentele evaluatie.

Een eerste route is via een gerandomiseerd experiment. In dat geval zou een loting bepalen welke instellingen de subsidie krijgen en welke niet. Het effect van de subsidie zou dan kunnen worden bepaald door de (mensen in de) instellingen die in de behandelgroep zitten - en dus de subsidie krijgen - te vergelijken met de (mensen in de) instellingen die een aanvraag hebben gedaan, maar de subsidie niet hebben gekregen (de controlegroep). Loting zou bijvoorbeeld kunnen worden ingesteld als er 'overuitputting' van middelen is, dat wil zeggen dat er meer aanvragen zijn dan budget om alle instellingen te helpen. Een loting kan een effectieve manier zijn om met de schaarste aan middelen om te gaan.

Een tweede mogelijkheid is, als loting niet mogelijk is, gebruik te maken van de 'afvallers' bij overuitputting van middelen. In dat geval zijn er meer aanvragers dan budget, maar worden de 'afvallers' niet door loting bepaald, maar bijvoorbeeld door het feit dat ze 'te laat' waren bij de aanvraag (first come, first serve principe). In dat geval kan een vergelijking worden gemaakt van de gehonoreerde aanvragen met de afgevallene aanvragen. Daarbij kan dan gecontroleerd worden voor de aanmelddatum van de (gehonoreerde) aanvraag.

Zijn deze mogelijkheden haalbaar voor het thema duurzame inzetbaarheid? Dit lijkt niet het geval. Ten eerste blijkt uit gesprekken met het Agentschap dat loting van ESF-middelen binnen dit thema niet mogelijk is.¹⁸ Ten tweede blijkt dat er in de praktijk geen sprake is van 'overuitputting' van middelen, maar eerder van 'onderuitputting': er wordt genoeg geld beschikbaar gesteld (vaak achteraf) om alle aanvragen te honoreren. Dit betekent dat er geen controlegroep gevormd kan worden van instellingen die vergelijkbaar zijn.

Een vergelijking waarbij aanvragers worden vergeleken met niet-aanvragers zou mank gaan, aangezien het zeer selectief is waarom instellingen een aanvraag doen voor ESF-geld binnen dit thema. Een instelling die bijvoorbeeld al veel aan active ageing doet (bijvoorbeeld via scholing) en gemotiveerd is daar nog meer aan te gaan doen, verschilt van een niet-aanvrager die weinig met dit thema opheeft. Als bij een vergelijking wordt gevonden dat (de mensen in) het eerste type instelling vaker geschoold worden dan (de mensen in) het tweede type instelling, is niet duidelijk of dat door de ESF-subsidie komt, of omdat de eerste instelling op voorhand al meer aan scholing deed.

¹⁸ Dit is gebleken uit de gesprekken die zijn gevoerd op 13 mei 2016 en 17 juni 2016. De bezwaren komen voort uit de vereiste van 'sound financial management'. Dit houdt in dat projecten beoordeeld moeten worden op kwaliteit en dat dus niet het lot bepaalt of de subsidie wordt verstrekt.

6 Conclusie

In deze notitie hebben we beoordeeld of het haalbaar is om met een (quasi-)experimentele methode:

1. De afzonderlijke beleidsinstrumenten die door het ESF-gecofinancierd worden te evalueren.
2. De netto inzet van de ESF-middelen als geheel te evalueren.

Dit is op twee manieren gedaan. Ten eerste is bij de betrokken instanties (het Agentschap, de arbeidsmarktregio's, UWV en Veiligheid & Justitie) nagegaan of er mogelijkheden zijn om *in de toekomst* gerandomiseerde experimenten op te zetten voor evaluatie van met ESF-geld betaalde beleidsinstrumenten. Ten tweede is onderzocht of er mogelijkheden zijn voor een quasi-experimentele evaluatie *achteraf* van reeds bestaande ESF-beleidsinstrumenten. Dat laatste is gebeurd aan de hand van twee belangrijke voorwaarden:

1. De aanwezigheid van exogene variatie in het beleidsinstrument. Dit wil zeggen dat er geen selectie van deelnemers is in het beleidsinstrument op basis van niet-observeerbare kenmerken en dat er dus een valide controlegroep gevormd worden.
2. Een voldoende grootte van de steekproefomvang. Daarbij is op basis van de in deze notitie gedane poweranalyses een minimum van 870 deelnemers gehanteerd.

Het antwoord op de eerste deelvraag is dat (quasi-)experimentele evaluatie slechts voor een deel van de met ESF-geld gefinancierde beleidsinstrumenten haalbaar lijkt. Het UWV onderschrijft het belang van experimentele evaluatie en is bereid de mogelijkheden voor experimenteel onderzoek te onderzoeken. Daarbij is het van belang dat op basis van een poweranalyse een voldoende grote steekproefomvang wordt gekozen en dat het experiment zorgvuldig wordt opgezet. Het gaat immers om een kwetsbare doelgroep (Wajongers) die vaak een relatief grote afstand tot de arbeidsmarkt kent. Het onthouden van beleidsinstrumenten in het kader van een experiment kan om die reden gevoelig liggen. Een manier om dit op te lossen is om in plaats van onthouding, een kandidaat altijd wel een (gepast) instrument aan te bieden, maar door loting te laten bepalen welk (gepast) instrument dat dan moet zijn. In dat geval kan de effectiviteit van de verschillende instrumenten tegen elkaar worden afgezet.

Bij de overige instanties lijkt (quasi-)experimenteel onderzoek niet haalbaar.

Bij de arbeidsmarktregio's, het ministerie van Veiligheid en Justitie en bij het Agentschap (op het terrein van duurzame inzetbaarheid) lijken er op dit moment geen mogelijkheden te zijn om in de toekomst gerandomiseerde experimenten uit te voeren. Quasi-experimenten met bestaand beleid lijken ook niet kansrijk. De belangrijkste reden hiervoor is het gebrek aan of de afwezigheid van exogene variatie in de beleidsinstrumenten. De andere reden is de steekproefomvang. Bij sommige beleidsinstrumenten bij de arbeidsmarktregio's lijkt er potentieel te zijn voor een quasi-experimentele evaluatie, maar blijkt de steekproefomvang te klein voor een betrouwbare effectevaluatie. Het aantal deelnemers bij deze instrumenten ligt ver onder het in deze notitie gehanteerde minimum aantal deelnemers. Dit laat overigens onverlet dat buiten het onderzochte ESF-terrein effectiviteitsonderzoek bij arbeidsmarktregio's wel mogelijk kan zijn.

Het antwoord op de tweede deelvraag houdt verband met dat op de eerste. Aangezien op dit moment slechts een deel van de met ESF-geld gefinancierde beleidsinstrumenten (quasi-)experimenteel te evalueren lijkt, is het niet mogelijk de netto inzet van ESF-middelen *in zijn geheel* op een *quasi-experimentele* manier te evalueren. De totale som van de ESF-middelen bestaat immers uit een groot aantal onderdelen. Dit geheel is niet eenvoudig te evalueren.

Het feit dat dit niet mogelijk is wil overigens niet zeggen dat er geen kennis voorhanden is over de met ESF-subsidie gefinancierde beleidsinstrumenten. Evenmin betekent dit dat evaluatie met andere technieken niet mogelijk is. Metastudies over actief arbeidsmarktbeleid, zoals die van Card et al. (2015) en van het CPB (Kansrijk Arbeidsmarktbeleid, deel 2, 2016) geven indicaties over welke instrumenten voor welke doelgroep effectief zijn. Daarnaast kunnen ook met niet-(quasi-)experimentele technieken en methoden onderzoek worden gedaan. Deze notitie betreft immers alleen een haalbaarheidsadvies over het doen van (quasi-)experimenteel onderzoek. Het doen van dit type onderzoek is niet eenvoudig, aangezien het een bron van exogene variatie in het beleidsinstrument vereist.

Het feit dat we nu weinig mogelijkheden hebben gevonden voor quasi-experimenteel onderzoek, betekent niet dat men in de toekomst niet moet proberen op zoek te gaan naar nieuwe mogelijkheden. De sleutel tot succes is een bron van exogene variatie. Het ex post op zoek gaan naar een dergelijke bron om een programma te kunnen evalueren is een kwestie van creativiteit van de onderzoeker, en van 'geluk hebben', bijvoorbeeld dat bij de bron ook voldoende observaties aanwezig zijn om met een redelijke waarschijnlijkheid uitspraken te kunnen doen.

Verder verdient het in de eerste plaats voorkeur om grote AMR's en V&J te proberen te enthousiasmeren voor het opzetten van een experimenteel onderzoek naar de effectiviteit van met ESF-subsidie gefinancierde instrumenten. Door het opzetten van experimenten is men niet afhankelijk van 'geluk' om het beleid achteraf nog te kunnen onderwerpen aan een effectmeting. Bovendien kan dan eenvoudiger op het

niveau van subgroepen een empirische analyse worden uitgevoerd. Randomisatie is nu eenmaal gemakkelijker te realiseren als een instrument nieuw ingevoerd wordt, omdat op dat moment in de controlegroep niets afgenomen wordt. Voordat een nieuw instrument breed ingezet wordt, kan het dan eerst in de vorm van een pilot op een kleinere groep getest worden.

Vooraf bij de AMR's zijn de aantallen bij veel beleidsinstrumenten vaak erg klein, waardoor effecten van het beleid relatief groot moeten zijn om statistisch significant aangetoond te kunnen worden. Door AMR's hun instrumenten met elkaar te laten afstemmen, waardoor hetzelfde instrument onder dezelfde voorwaarden in meerdere AMR's ingezet wordt, kan het probleem van een te kleine steekproef verholpen worden.

Referenties

Abbring, J.H. en J.J. Heckman, 2007, Econometric evaluation of social programs, part III: distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation, in: *Handbook of Econometrics Volume 6B*, Elsevier.

Agentschap SZW, 2014, Factsheet Operationeel programma 2014-2020 Europees Sociaal Fonds

Card, D., J. Kluve en A. Weber, 2015, What works? A Meta Analysis of Recent Active Labor Market Program Evaluations, NBER Working Paper 21431.

CPB, *Kansrijk Arbeidsmarktbeleid, deel 2*.

<http://www.cpb.nl/sites/default/files/publicaties/download/cpb-boek-22-kansrijk-arbeidsmarktbeleid-deel-2.pdf>

DiNardo, J. en D.S. Lee, 2011, Program Evaluation and Research Designs, in: *Handbook of Labor Economics Volume 4a*, Elsevier.

European Commission 2015, Monitoring and Evaluation of European Cohesion Policy. European Social Fund, Guidance document

Operational Programme ESF 2014-2020

Subsidieregeling ESF 2014-2020: Regeling van de Staatssecretaris van Sociale Zaken en Werkgelegenheid van 31 maart 2014, 2014-0000040627, tot de besteding van gelden uit het Europees Sociaal Fonds 2014-2020

Bijlage A: Het Europees Sociaal Fonds (ESF)

Het Europees Sociaal Fonds (ESF) is het belangrijkste Europese subsidie-instrument voor arbeidsmarktbeleid. Uitvoering van de subsidieverstrekking is in Nederland in handen van agentschap SZW. Voor de periode 2014-2020 is 507 miljoen euro beschikbaar voor Nederland uit het Europees Sociaal Fonds. Nederland heeft drie investeringsprioriteiten geformuleerd en bepaalt welk deel van de beschikbare 507 miljoen aan elke prioriteit besteed zal worden. De investeringsprioriteiten zijn:

1. Actieve Inclusie: re-integratie van mensen met een afstand tot de arbeidsmarkt (71% van het beschikbare budget, waarvan 1% voor projecten gericht op Sociale Innovatie en transnationale samenwerking).
2. Vier grootste steden met een Geïntegreerde Territoriale Investering (GTI): toegang tot werkgelegenheid voor werkzoekenden en niet-actieven, ondersteuning arbeidsmobiliteit in vier grote steden (5%).
3. Actief en gezond ouder worden: het bevorderen van duurzame inzetbaarheid van werkenden (20% van het beschikbare budget)

De overige 4% van het budget is beschikbaar voor technische ondersteuning door Agentschap SZW. Middels deze investeringsprioriteiten wil Nederland aan het eind van de subsidieperiode (in 2020) bijgedragen hebben aan het bereiken van een tweetal doelstellingen: het verhogen van de (bruto) arbeidsparticipatie onder 20-64 jarigen naar 80% (was 76,6% in 2012) en het verminderen met 100.000 van het aantal personen tussen 0-64 jaar die leven in een huishouden met een lage werkintensiteit (ESF Operational Programme).

Actieve Inclusie

Binnen de investeringsprioriteit Actieve Inclusie kunnen de centrumgemeenten van de 35 arbeidsmarktregio's (namens gemeenten en (ex-)leerlingen van VSO/PrO-scholen), UWV en het ministerie van Veiligheid en Justitie (V&J) financiering aanvragen voor projecten die mensen in de doelgroep aan het werk helpen en houden. De doelgroep bestaat uit¹⁹:

- Iedereen die in een periode van 12 maanden meer dan 6 maanden een uitkering van gemeente of UWV ontving.
- Nuggers (niet-uitkeringsgerechtigde werkzoekenden).
- Mensen met een arbeidsbelemmering.
- 50-plussers met een uitkering van de gemeente of UWV.
- Jongeren tot 28 jaar.

¹⁹ Informatieflyer ESF voor gemeenten en informatieflyer ESF voor VSO/PrO-scholen.

- (Ex-) leerlingen van VSO/PrO-scholen van 15 jaar en ouder die niet langer dan 24 maanden geleden de school hebben verlaten.
- (Ex-) gedetineerden.
- Vreemdelingen die rechtmatig in Nederland verblijven.²⁰

Alle activiteiten binnen een door het ESF gefinancierd project moeten gericht zijn op uitstroom naar betaald werk. Kosten die vergoed kunnen worden middels een subsidie uit het ESF zijn

- Directe loonkosten met opslagen voor werkgeverslasten en overige kosten.
- Externe kosten.
- Plaatsingssubsidies.
- Kosten van extern ingekochte trajecten (alleen voor UWV en V&J).

De maximale duur van projecten die gesubsidieerd kunnen worden is 12 maanden voor VSO/PrO 24 maanden voor gemeenten, 15 maanden voor V&J en 24 maanden voor UWV.

Duurzame Inzetbaarheid

Duurzame inzetbaarheid richt zich op de vraag hoe ervoor gezorgd kan worden dat werknemers productief, gezond en gemotiveerd blijven en door kunnen blijven werken tot aan de pensioengerechtigde leeftijd. Het kunnen combineren van werk en zorgtaken is hiervan een belangrijk onderdeel. Deze investeringsprioriteit richt zich dan ook met name op oudere werknemers en vrouwen.

Binnen deze investeringsprioriteit kunnen O&O-fondsen, sociale partners, bedrijven en (overheids)instellingen subsidie aanvragen voor het ontwikkelen en implementeren van *active ageing*-beleid. Daarbij kan het zowel gaan om sectorbrede analyses en adviezen als ook adviezen aan specifieke bedrijven of (overheids-)instellingen. Er is zowel een budget voor kleine projecten (circa 5500 projecten van 10.000 euro per project), waarbij een bedrijf of instelling advies kan inwinnen over *active ageing*-beleid, als voor grote projecten (circa 480 projecten), waarbij naast advies ook een implementatietraject wordt gesubsidieerd.

De kosten die vanuit het ESF vergoed kunnen worden zijn de kosten van het opstellen van een advies en de kosten van begeleiding bij de implementatie van een advies.

Onderzoek naar effectiviteit

De Europese Commissie verplicht lidstaten om onderzoek te doen naar de 'effectiviteit, efficiency en impact' van met ESF-geld gefinancierde activiteiten

²⁰ Rechtmatig in Nederland verblijf houden in de zin van artikel 8, onderdelen a tot en met e en l, van de Vreemdelingenwet 2000, met uitzondering van de gevallen, bedoeld in artikel 24, tweede lid, van Richtlijn 2004/38/EG.

(art.54(1) CPR). Van lidstaten wordt verwacht dat zij een evaluatieplan opstellen waarin wordt vastgelegd hoe zorg wordt gedragen voor de evaluatie van met ESF-geld gefinancierde activiteiten.²¹ Lidstaten zijn daarbij vrij te kiezen welke onderzoeksmethoden zij toepassen. De EC laat echter wel haar voorkeur doorschemeren door in haar 'Monitoring and Evaluation Guidance document' te stellen dat twee typen evaluatieonderzoek breed gedragen worden: *theory-based impact evaluation* voor een kwalitatieve evaluatie van de impact, en *counterfactual impact evaluation* voor een kwantitatieve evaluatie van de impact. Dit laatste type evaluatieonderzoek is gebaseerd op het principe van een controle- en behandelgroep en vereist een (quasi-)experimentele opzet van het evaluatieonderzoek (Europese Commissie 2015).


²¹ Zie ook Evaluatieplan ESF 2014-2020: <https://www.agentschapszw.nl/documenten/publicaties/subsidies/esf-2014-2020-algemene-informatie-europees-sociaal-fonds/rapporten-over-esf-2014-2020/evaluatieplan-esf-2014-2020>

Bijlage B: overzicht AMR's

Nr.	AMR	Type projecten AMR	Type projecten VSO/PRO	Wachtlijst?*	Leeftijdsgrens?*
1	AMR Ede	Re-integratietrajecten/Loonkostensubsidies	Branchecursussen/Stages	Nee	Nee
2	AMR Enschede	Re-integratietrajecten/Arbeidsinpassing	Leerwerkbedrijf	Nee	Nee
3	AMR Doetinchem	Jobcoaching/Loonkostensubsidie/Trajectbegeleiding	Leerwerkbedrijf/Branchecursussen	Ja	Ja
4	AMR Haarlem	Re-integratietrajecten	Leerwerkbedrijf	Nee	Nee
5	AMR Amersfoort	Re-integratietrajecten/Casemanagement	Leerwerkbedrijf/Stagebegeleiding	Nee	Nee
6	AMR Zoetermeer	Re-integratietrajecten/Jobcoaching/Loonkostensubsidies/Scholing	Stages/Branchecursussen	Nee	Nee
7	AMR Den Haag	Jobcoaching/Loonkostensubsidie/Re-integratietrajecten	Leerwerkbedrijf/Stages/Branchecursussen	Nee	Nee
8	AMR Amsterdam	Jobcoaching/Loonkostensubsidie/Re-integratietrajecten/Trainingen	Leerwerkbedrijf/Stages/Branchecursussen	Nee	Nee
9	AMR Dordrecht	Re-integratietrajecten/Leerwerkbedrijf/Loonkostensubsidies	Stages/Leerwerkbedrijf	Nee	Nee
10	AMR Breda	Re-integratietrajecten/Loonkostensubsidie/Stimuleren zelfstandig ondernemerschap	Stagebegeleiding/Onderzoek vaardigheden	Nee	Ja
11	AMR Almere	Re-integratietrajecten/Loonkostensubsidie/Jongerenloket/Trainingen/Taallessen	Stages/Branchecursussen	Nee	Nee
12	AMR Den Bosch	Re-integratietrajecten	Stages/Leerwerkbedrijf	Nee	Nee
13	AMR Tilburg	Re-integratietrajecten	Stages/Branchecursussen	Nee	Nee
14	AMR Nijmegen	Re-integratietrajecten/Loonkostensubsidies/(Taal)trainingen	Stages/Branchecursussen	Nee	Ja
15	AMR Venlo	Loonkostensubsidies/Branchecursussen	Leerwerkbedrijf/Branchecursussen	Nee	Nee
16	AMR Groningen	Re-integratietrajecten	Stages/Leerwerkbedrijf/Jobcoaching/Avondschool	Nee	Ja
17	AMR Tiel	Re-integratietrajecten/Jobcoaching/Taaltrainingen	Stages/Branchecursussen/Leerwerkbedrijf	Nee	Ja
18	AMR Gorinchem	Re-integratietrajecten/Loonkostensubsidies	Stages/Branchecursussen	Nee	Ja
19	AMR Leiden	Trajectbegeleiding	Stages	Nee	Nee
20	AMR Goes	Re-integratietrajecten/Plaatsingsubsidies/Jobcoaching/Scholing	Stages/Branchecursussen	Nee	Nee
21	AMR Apeldoorn	Re-integratietrajecten/Loonkostensubsidies/Jobcoaching/Branchecursussen	Stages/Leerwerkbedrijf/Branchecursussen	Nee	Nee
22	AMR Arnhem	Re-integratietrajecten	Leerwerkbedrijf/Branchecursussen	Nee	Nee
23	AMR Leeuwarden	Casemanagement/Loonkostensubsidies/Scholing	Stages	Nee	Nee
24	AMR Zwolle	Casemanagement	Leerwerkbedrijf	Ja	Nee
25	AMR Emmen	Casemanagement/Loonkostensubsidies/Re-integratietrajecten	Stages/Leerwerkplaatsen	Nee	Nee
26	AMR Hilversum	Re-integratietrajecten	Stages/Leerwerkbedrijf/Avondschool	Nee	Nee

27	AMR Helmond	Re-integratietrajecten/Leerwerplekken	Stages/Leerwerkbedrijf/Branchecursussen	Ja	Nee
28	AMR Eindhoven	Re-integratietrajecten/Loonkostensubsidie/Scholing	Leerwerkbedrijf	Nee	Nee
29	AMR Gouda	Re-integratietrajecten/Loonkostensubsidies/Jobcoaching	Leerwerkbedrijf	Nee	Nee
30	AMR Roermond	Re-integratietrajecten/Loonkostensubsidies	Stages/Branchecursussen/Leerwerkbedrijf	Ja	Nee
31	AMR Utrecht	Re-integratietrajecten/Loonkostensubsidies/Jobcoaching/Taaltraining	Stages/Jobcoaching/Avondschoon/Branchecursussen	Nee	Nee
		Begeleiding naar werk/Loonkostensubsidies/Begeleiding bij			
32	AMR Alkmaar	proefplaatsing/Jobcoaching	Leerwerkbedrijf	Nee	Nee
33	AMR Rotterdam	Re-integratietrajecten	Jobcoaching/Leerwerkbedrijf/Stages	Nee	Nee
34	AMR Zaanstad	Re-integratietrajecten	Stages/Branchecursussen	Nee	Nee
35	AMR Heerlen	Re-integratietrajecten/Jobcoaching/Loonkostensubsidies	Stages/Leerwerkbedrijf/Branchecursussen	Nee	Ja

*De vraag die gesteld is: "Wordt er (binnen uw arbeidsmarktregio) bij (sub)projecten met een wachtlijst of leeftijdsgrens gewerkt?"



Dit is een uitgave van:

Centraal Planbureau
Van Stolkweg 14
Postbus 80510 | 2508 GM Den Haag
T (070) 3383 380

info@cpb.nl | www.cpb.nl

Juli 2016